Giuseppe Biondi-Zoccai
*Editor*

# Umbrella Reviews

## Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies

Springer

# Umbrella Reviews

Giuseppe Biondi-Zoccai
Editor

# Umbrella Reviews

## Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies

Springer

*Editor*
Giuseppe Biondi-Zoccai
Department of Medico-Surgical Sciences and Biotechnologies
Sapienza University of Rome
Latina
Italy

ISBN 978-3-319-25653-5 ISBN 978-3-319-25655-9 (eBook)
DOI 10.1007/978-3-319-25655-9

Library of Congress Control Number: 2015960744

Springer Cham Heidelberg New York Dordrecht London


Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

*To Giulia, Erica, Gina, and Marzia, the four women of my life*

# Foreword

*Considerate la vostra semenza:*
*fatti non foste a viver come bruti,*
*ma per seguir virtute e canoscenza.*

*Consider ye the seed from which ye sprang;*
*Ye were not made to live like unto brutes,*
*But for pursuit of virtue and of knowledge.*

Dante Alighieri. "The Divine Comedy – Inferno" in Canto 26:116–118.

Like Ulysses in the Greek mythology, physicians and researchers are destined to endlessly wonder in the search for the *truth*.

Sir Karl Pooper (1902–1994) in his philosophical treaty on Empirical Falsification concludes that "A theory in the empirical sciences can never be proven, but it can be falsified, meaning that it can and should be scrutinized by decisive experiments (designed to test and) contradict the theory." Given this assumption, it derives that no clinical decision can be irrefutably considered "correct."

So how is a physician to decide when faced with a clinical question or challenge? Paradoxically, if there were two therapeutic strategies that appear equally effective, the physician could risk behaving like the *Buridian's ass* which, being equally hungry and thirsty and being placed precisely midway between a stack of hay and a pail of water, will inevitably die of thirst and hunger since it cannot make any rational decision to choose one of the other. This is an obvious paradox, which Aristotle, the Greek philosopher, had highlighted in his *On the Heavens* more that 2000 years ago, and referring to the risk of becoming *paralyzed* by the uncertainties of life.

Now, are we physicians in 2015 more knowledgeable than physicians were 100 years ago? And are we capable to make better-informed decisions?

I would answer "yes" to these questions but not without having digressed some on the unconscious and conscious processes involved in the decision-making. The human brain is a formidable computing system able to integrate different sensorial inputs and historical memories into a complex decision-making. The brain has also limitations: one of it being the time required and the ability to input and elaborate a large amount of notions, numbers, and information in general.

The editor of this book, Dr. Biondi-Zoccai, had already introduced the reader to network meta-analysis as a mean for integrating the data and signals from various

clinical trials including direct and indirect comparisons. This new book is about "umbrella reviews" or "reviews of reviews." One way to see it is like an actual "umbrella" that prevents you from getting "soaking wet" under a "rain of evidence."

Taking this further, umbrella reviews are not only helpful, but they are also needed. Our learning process is indeed additive and cumulative, as such that the latest data inputted does not erase prior data. A positive trial on a treatment X can only be interpreted in the setting of all prior trials with such treatment and in comparison with other available treatments Y and Z. This is particularly important when professional societies use data to create clinical guidelines, and guidelines drive standard-of-care and possibly also cost coverage. We all need to recognize that any clinical trial, review, or meta-analysis provides us with a more or less incomplete view of the "truth." It is therefore necessary to continuously and indefinitely cumulate and analyze all available evidence, recognizing that no conclusion is ever final.

This book edited by Dr. Biondi-Zoccai constitutes the first book of this kind in which the definition, techniques, and uses of umbrella reviews are systematically presented. For those who are trained and expert in biostatistics, this book represents a way to further improve the skills and focus on the global view of the umbrella reviews. For those who are not skilled, and perhaps not interested, in the technical aspects of umbrella reviews, this book represents an essential reading to understand the principles, goals, and limitations of this approach, as these reviews will more and more guide clinical practice.

I have to commend Dr. Biondi-Zoccai for assembling such an international faculty as authors, which fits well with the view that knowledge is a global process.

I found Part I as an important introduction to the topic. Part II describes the technical aspects of the process. Part III provides specific examples and case studies that help the reader better understand the process. The final Part IV is a call to action in the translation to clinical practice and avenues for further research.

In conclusion, under a "rain" of evidence deriving from reports, trials, and reviews, the umbrella review provides the reader with a protective strategy that limits the "noise" and allows to see one or more "signals" that could otherwise be poorly distinguished or missed.

Richmond, VA, USA Antonio Abbate, MD, PhD

# Acknowledgments

*For benefits received are a delight to us as long as we think we can requite them; when that possibility is far exceeded, they are repaid with hatred instead of gratitude.*

Tacitus (AD 56–after 117) [1]

In this second book on evidence synthesis I have had the honor to edit [2], I am lucky enough to be able to continue to give back as much as I received. The recipients of my gratitude are all the worthwhile persons whose seminal support enabled the completion of my present endeavor, as well as all those of the past. First and foremost, my wife, Marzia, and our awesome children, Attilio Nicola, Giuseppe Giulio, and Giovanni Vincenzo, are behind my every heartbeat and thought. My mother Giulia and my father Gianni continue of course to be as important. This is notwithstanding my sisters, Gina and Erica, and my brother Vincenzo. Their inextinguishable love and adamant faith in their beloved husband, father, son, and brother easily overwhelms any challenge from the past, present, or future.

My true friends, entrusting their affection and esteem to me even when this could have been apparently detrimental, should not be disregarded as well. First of all Antonio Abbate, who has also agreed to provide his enlightening perspective as leading translational researcher and clinician–investigator on the topic of umbrella reviews in the preceding pages, but also his wonderfully intelligent and caring wife, Vera Di Trocchio Abbate, and their treasured daughter, Gerardina. Antonio's leading and caring mentorship toward me has only been surpassed over the years by his love and affection for me and my family. Enrico Romagnoli has proved a competent colleague but even more a trustful friend, as well as being, with his wife Marianna, a caring godparent of my junior son, Attilio Nicola. Their sons, Andrea and Elisa, are as close to mine as their parents are. Pierfrancesco Agostoni, now also proud father of Antonio Juventus, has been and will always be one of my dearest friends and smartest colleagues. Indeed, he was lead author of my very first scholarly publication [3] and senior author of my first published overview of reviews and meta-epidemiologic study [4]. Giacomo Frati, Antonino Giuseppe Maria Marullo, Mariangela Peruzzi, and Massimo Mancone, brilliant colleagues at Sapienza University of Rome, have been adamantly supportive and warmly close to me throughout these last few years, and I am happy they participated in one of our most recently published umbrella reviews [5]. This is even more peculiar considering that both Giacomo and Antonino are cardiac surgeons befriending an interventional cardiologist such as myself. Marco

Antonioli has always been a treasured friend and colleague, from medical school to the military service and later as practicing physician, despite his interest in psychiatry which has eventually brought him to work in the UK. More recently, I am very happy to have had the opportunity of befriending Arturo Giordano, whose competence and commitment to the highest standards in interventional cardiology have proved a true inspiration to me. Similarly seminal for me has been the work and support of Francesco Nudi, whose insight and skill in cardiovascular medicine at large and cardiovascular imaging in particular are unprecedented.

On top of Antonio Abbate's outstanding help and faith in me, I am honored to recognize the leading cardiologists of the Virginia Commonwealth University, in Richmond, George W. Vetrovec and Kenneth A. Ellenbogen, whose commitment to excellence in clinical practice and research have been instrumental in furthering my scholarly goals. In particular, I am very thankful for having had the opportunity to participate in the prestigious Congdon Visiting Scholar program established at the Virginia Commonwealth University Pauley Heart Center by Jack Congdon in memory of his late wife, Natalie. Heartfelt thanks also go to Springer and their competent and supportive staff. In particular, I am happy to acknowledge Roberto Garbero, Donatella Rizza, and Angela Schulze-Thomin from Springer, who had the patience and insight of letting me coordinate a veritable dream team of international leaders in evidence synthesis to realize the first book ever on umbrella reviews, overviews of reviews, and meta-epidemiologic studies.

Finally, I wish to emphasize again that this work is strenuously dedicated to the four women of my life, my mother Giulia, my wife Marzia, and my sisters Erica and Gina.

Latina, Italy Giuseppe Biondi-Zoccai, MD, MStat

## References

1. Tacitus. The annals. Book IV (trans: Church AJ, Brodribb WJ). Available at: http://classics.mit.edu/Tacitus/annals.html. Last accessed on 31 July 2015.
2. Biondi-Zoccai G. Network meta-analysis: evidence synthesis with mixed treatment comparison. Hauppauge: Nova Science Publishers; 2014.
3. Agostoni P, Biondi-Zoccai GG. Prognostic value of C-reactive protein in unstable angina. Circulation. 2000;102:E177.
4. Biondi-Zoccai GG, Lotrionte M, Abbate A, Testa L, Remigi E, Burzotta F, Valgimigli M, Romagnoli E, Crea F, Agostoni P. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: case study. BMJ. 2006;332:202–9
5. Peruzzi M, De Falco E, Abbate A, Biondi-Zoccai G, Chimenti I, Lotrionte M, Benedetto U, Delewi R, Marullo A, Frati G. State of the art on the evidence base in cardiac regenerative therapy: overview of 41 systematic reviews. Biomed Res Int. 2015;2015:613782.

# Contents

**Part I Umbrella Reviews in the Evidence Hierarchy**

**1 Introduction** . . . . . 3
Giuseppe Biondi-Zoccai

**2 Evidence Hierarchy** . . . . . 11
Colin Ng and Umberto Benedetto

**3 From Qualitative Reviews to Umbrella Reviews** . . . . . 21
Ana Ortega, Eduardo Lopez-Briz, and María Dolores Fraga-Fuentes

**4 Umbrella Reviews, Overviews of Reviews, and Meta-epidemiologic Studies: Similarities and Differences** . . . . . 43
Michail Tsagris and Konstantinos C. Fragkos

**Part II Sound Design, Conduct, and Reporting**

**5 Designing the Review** . . . . . 57
Spyridon N. Papageorgiou and Giuseppe Biondi-Zoccai

**6 Registering the Review** . . . . . 81
Alison Booth

**7 Searching Evidence** . . . . . 95
Su Golder and Kath Wright

**8 Abstracting Evidence** . . . . . 107
Giuseppe Biondi-Zoccai, Giuseppe La Torre, Leonardo Roever, and Fabrizio D'Ascenzo

**9 Appraising Evidence** . . . . . 115
William L. Baker, Meg Bennetts, Craig I. Coleman, and Joseph C. Cappelleri

**10 Statistical Models for Overviews of Reviews** . . . . . 137
Tzu-Ting Chen and Yu-Kang Tu

**11 Bayesian Approach to Evidence Synthesis** . . . . . 155
Francisco José Vázquez Polo, Miguel A. Negrín, and María Martel

**12 Appraising Between-Study Homogeneity, Small-Study Effects, Moderators, and Confounders** . . . . . 161
Areti Angeliki Veroniki, Tania B. Huedo-Medina, and Kostas N. Fountoulakis

**13 State-of-the-Art Reporting** . . . . . 189
Akira Onishi and Toshi A. Furukawa

**Part III Case Studies**

**14 Case Study of Allied Health and Complementary and Alternate Medicine** . . . . . 205
Saravana Kumar

**15 Case Study in Dentistry** . . . . . 217
Clovis Mariano Faggion Jr and Utsana Tonmukayakul

**16 Case Study in Neonatology** . . . . . 235
David Andrew Osborn

**17 Case Study in Nursing** . . . . . 273
Walter De Caro, Anna Rita Marucci, Loreto Lancia, and Julita Sansoni

**18 Case Study in Orthopedics** . . . . . 305
Maciej Płaszewski and Josette Bettany-Saltikov

**19 Case Study in Prevention** . . . . . 331
Leandro Fórnias Machado de Rezende, Juan Pablo Rey-López, Olinda do Carmo Luiz, and Jose Eluf-Neto

**20 Case Study of Public Health** . . . . . 343
Clare Bambra and Marcia Gibson

**Part IV Implementation and Further Research**

**21 Moving from Evidence to Action** . . . . . 365
Fabrizio D'Ascenzo, Claudio Moretti, Christian Templin, and Fiorenzo Gaita

**22 Avenues for Further Research** . . . . . 373
Dawid Pieper, Lun Li, and Roland Brian Büchter

**23 Conclusion** . . . . . 389
Giuseppe Biondi-Zoccai

# Part I

# Umbrella Reviews in the Evidence Hierarchy

# Introduction

1

Giuseppe Biondi-Zoccai

**Abstract**

Sound decision making requires the optimal use of available data on the topic at hand. Whenever evidence is multifaceted, abundant, or otherwise complex to face, specific tools must be envisioned for informed decision making. Historically, reviews were the first type of evidence synthesis tool. Subsequently, meta-analyses have complemented the most refined type of reviews (i.e. systematic reviews). This field of research methodology has been further expanded by the recent availability of umbrella reviews, overviews of reviews and meta-epidemiologic studies, which provide a more general framework for evidence synthesis and decision making, encompassing multiple sources of information (e.g. different systematic reviews on the same topic, or different systematic reviews on different but connected topics). This chapter serves as the introduction to our textbook devoted to this novel and fascinating topic.

*It is the habitual carriage of the umbrella that is the stamp of Respectability. The umbrella has become the acknowledged index of social position…. Crusoe was rather a moralist than a pietist, and his leaf-umbrella is as fine an example of the civilized mind striving to express itself under adverse circumstances as we have ever met with.*

Robert Louis Stevenson (1851–1894) [1]

What is evidence? Evidence can be defined as the body of facts and information available on a specific belief [2, 3]. And what is synthesis? Synthesis can be regarded as the combination of different entities to form a coherent system.

G. Biondi-Zoccai, MD, MStat
Department of Medico-Surgical Sciences and Biotechnologies,
Sapienza University of Rome, Corso della Repubblica 79, Latina 04100, Italy

Eleonora Lorillard Spencer Cenci Foundation, Rome, Italy
e-mail: giuseppe.biondizoccai@uniroma1.it



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_1

Not unexpectedly, the pursue of accurate, precise, and efficient evidence synthesis remains rather challenging. Thanks to the increased participation of multiple investigators and stakeholders, sources of evidence in clinical medicine as well as in other fields of human endeavor will continue to increase, possibly exponentially [4]. Thus, the only sensible means to navigate such information overload are flexible yet powerful tools for evidence synthesis [5, 6].

Evidence is also hierarchical (Table 1.1). From preclinical studies to primary studies (e.g., randomized controlled trials) and secondary studies (e.g., systematic reviews, pairwise meta-analyses, and network meta-analyses), there is a continuum of different study designs, yielding altogether different results, in terms of internal and external validity [3, 15]. Indeed, major developments have occurred in the field of secondary research, and in particular the introduction and now rather common application of network meta-analytic techniques have enabled powerful, robust, and elegant synthesis of apparently incoherent sets of evidence [16]. Further refinements of these approaches are expected, including multivariate meta-analytic studies, capable of providing insights on the comparative safety and efficacy of different interventions on different domains of a given condition or even on different conditions as well.

Nonetheless, we cannot consider the exponential accrual of secondary research studies (e.g., systematic reviews and meta-analyses) as solely positive. It is already very difficult to remain up to date given that so many reviews are being published on a daily basis (for instance, on average at least four network meta-analyses were being published daily during the months leading to the completion of this book in the summer of 2015). Moreover, it is not uncommon to find different meta-analyses focusing on similar topics and providing potentially different results. Finally, no systematic review per se is usually capable of providing a comprehensive yet succinct perspective on complex conditions or problems.

The idea of looking at reviews as objects of research rather than solely as a scholarly product is not new and was pioneered in the 1980 and 1990s by several leaders in evidence synthesis, such as Andrew Oxman and Gordon Guyatt, among many others [2, 17]. The success of the Cochrane Collaboration and its leadership worldwide [7, 18, 19], as well as the commitment of other leading institutions such as the Joanna Briggs Institute [8], have further supported the development of a new set of tools for evidence synthesis, operating at a higher level than systematic reviews and meta-analyses, which editors, reviewers, and readers already seem to enjoy quite remarkably [7, 8, 20, 21]. Moreover, the flexibility of this type of research design is substantial, as animal experiments and observational studies can also be included and findings may be combined formally with multivariate modeling [22–24].

The concept of this book stemmed from the successful collaborative effort we have conducted in 2013–2014 in producing the first textbook solely dedicated to network meta-analysis [3] and to the interest we have had for more than 15 years in looking at systematic reviews and meta-analyses as uniquely elegant and interesting tools, worthy of study, comparison, and synthesis. Accordingly, we have planned a comprehensive textbook, the first uniquely dedicated to umbrella reviews, overviews of reviews, and meta-epidemiologic studies. As our common ultimate goal remains evidence synthesis, this book should best be appraised together with our own opus and similar ones on mixed treatment comparisons, as there is a substantial continuum and overlap between these apparently different research designs [3, 9, 22].

**Table 1.1** Galaxy of research designs, distinguishing three main levels of research, with corresponding study designs and features [2, 3, 7–14]

| Research level | Research design | Key features |
|---|---|---|
| Primary | Case report | Single case description |
| | Case series | Description of a limited number of similar cases |
| | Case-control study | Observational study comparing a set of cases and controls |
| | Cohort study | Observational study following patients during a specified time (may include controls) |
| | Cross-sectional study | Observational study not following patients during time |
| | Qualitative study | Systematic description of subjective experiences and opinions |
| | Preclinical study | Preclinical (e.g., animal or in vitro) research report |
| | Randomized controlled trial | Experimental study based on the random allocation of different subjects to different types of interventions |
| Secondary | Meta-analysis | Statistical analysis of primary studies (typically within the context of a systematic review)[a] |
| | Mixed method systematic review | Review integrating quantitative and qualitative studies |
| | Narrative review | Review without any distinct and systematic feature |
| | Qualitative synthesis review | Review focusing solely on qualitative studies |
| | Rapid review | Succinct review aiming at informing on a given topic in a timely fashion (typically completed within a few weeks) |
| | Scoping review | Succinct review aiming at mapping the key concepts relevant to a broader topic and guiding further and more comprehensive systematic reviewing efforts |
| | Systematic review | Review based on explicit and standardized methods for design, search, selection, abstraction, appraisal, synthesis, and reporting of sources of evidence |
| Tertiary | Meta-epidemiologic study | Study typically (but not only) appraising systematic reviews, without explicitly aiming at informing on a specific clinical condition or topic |
| | Overview of reviews | Study only appraising reviews and typically (but not always) aiming at informing on a specific clinical condition or topic |
| | Umbrella review | Study typically (but not only) appraising systematic reviews and aiming at informing on a specific clinical condition or topic |

[a]Occasionally conducted within a set of different systematic reviews or in the context of an umbrella review (in such cases the most appropriate terms are network meta-analysis, mixed treatment comparison, or multivariate meta-analysis)

First and foremost, why should we use three different identifying terms in a book which focuses mainly on overviews of systematic reviews [10, 25, 26]? Actually, umbrella reviews, overviews of reviews, and meta-epidemiologic studies share much in common, but cannot be considered perfect synonyms (Fig. 1.1). Indeed,

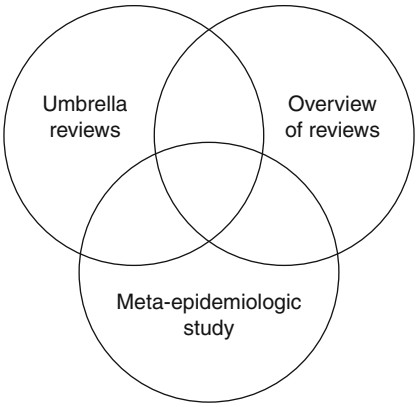



**Fig. 1.1** Venn diagram showcasing the overlap between umbrella reviews, overviews of reviews, and meta-epidemiologic studies. For instance, umbrella reviews and meta-epidemiologic studies may both include primary studies (e.g., randomized controlled trials) not included in any systematic review, whereas umbrella reviews and overviews of reviews typically focus on a specific clinical topic, at odds with meta-epidemiologic studies

they are different types of tertiary research, i.e., research mainly using as study objects systematic reviews.[1] Yet, umbrella reviews can be operatively considered exercises in evidence synthesis focusing on a specific clinical topic or condition, and including mainly systematic reviews, but with the possible inclusion of primary studies outside any prior meta-analysis [23, 28]. Overviews of systematic reviews also focus explicitly on a clinical topic or condition, but should not typically include primary studies or other non-review studies outside the realms of prior systematic reviews [29]. Finally, meta-epidemiologic studies usually disregard the goal of providing practical guidance on a specific clinical condition or topic, but usually include secondary research studies as well as primary research ones [30]. At one theoretical extreme, a meta-epidemiologic study could, for instance, include only editorials and thus disregard altogether primary or secondary research studies. Thus, while the goals or the analysis sets may differ substantially, these three types of studies typically do share more than what they do not, especially in terms of scope and methods. More pragmatically, the premises, the tools used, the skills required, and the final products are similar enough for those interested in critically reading or proficiently conducting them to justify, in our humble opinion, a common playground for their scholarly presentation [9, 31].

More explicitly, what is our purpose with the compilation of this multiauthored textbook? This book aims at providing readers a practical opportunity for comprehensive, effective, and efficient evidence synthesis, through an explicit

[1] Our proposed stratification of sources of clinical research in primary (clinical studies), secondary (systematic reviews and meta-analysis of clinical studies), and tertiary (umbrella reviews, overviews of reviews, and meta-epidemiologic studies) is divergent from the original one proposed in 1976 by Gene V. Glass, who defined primary research as original research, secondary research as re-analysis of a primary research dataset, and meta-analysis as an upper level of research, summarizing primary and secondary studies [27].

structure divided in four main sections. The first one highlights in different chapters the peculiarities of umbrella reviews, overviews of reviews, and meta-epidemiologic studies and their rightful place in the modern hierarchy of evidence. The second section contains chapters which guide the reader through the process of designing, registering, and conducting with the utmost validity and transparency an umbrella review, an overview of reviews, or a meta-epidemiologic study [11, 32]. In particular, we provide explicit details on searching, abstracting, and appraising evidence and then perform statistical analysis and appraisal of homogeneity, small study effects, moderators and confounders, as well as state-of-the-art reporting. In the third section the reader will find several authoritative case studies of tertiary research, which highlight the strengths as well as weaknesses of this type of research endeavor. Finally, the fourth and final section suggests how to move from the process of synthesizing evidence to actually acting upon it and which future areas of research and development for this field can be envisioned today. Indeed, tertiary research simply represents one of the steps in the life cycle of evidence, with a persistent continuum between the different levels of research, each informing on the following as well as the previous ones (Fig. 1.2) [5]. Accordingly, only in putting umbrella reviews in the larger context of evidence accrual can we righteously use them.

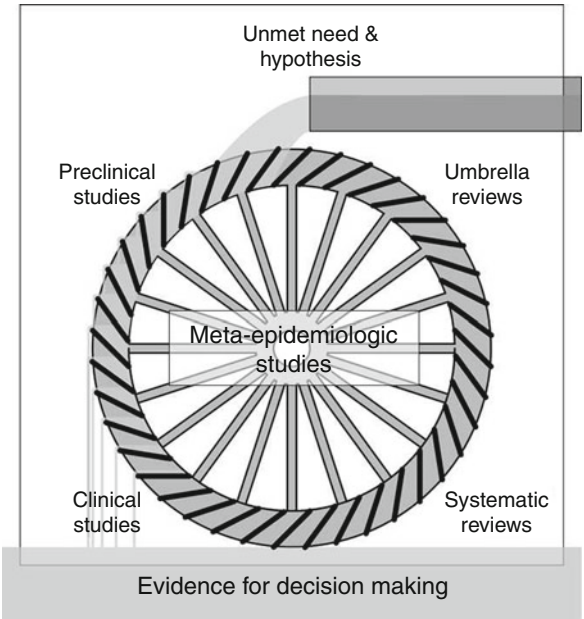



**Fig. 1.2** The evidence mill, highlighting how unmet needs and novel hypotheses fuel the conduct of preclinical studies and primary clinical research explicitly guiding decision-making. Such research products are then the object of systematic reviews and umbrella reviews, eventually informing on needs and hypotheses, as well as influencing further primary studies and decision-making. Meta-epidemiologic studies simultaneously offer an alternative way to appraise the complex relationships between these types of research designs and the potential weaknesses in the evidence base, thus also, albeit indirectly, guiding decision-making

Whereas we do not wish nor need to defend reviews in general in this introductory chapter of the book [33], we would like to make the case that umbrella reviews, overviews of reviews, and meta-epidemiologic studies do close the circle of evidence, potentially reconciling all sources of evidence, even those of lower quality or focusing on less important issues or interventions [16, 34]. Indeed, a historical critique of systematic reviews and meta-analyses is that they can easily end up mixing apples and oranges, especially if pooling heterogeneous trials [2]. Even if we concede this, looking comparatively at apples and oranges will often tell us a great deal about fruit in general. Umbrella reviews, overviews of reviews, and meta-epidemiologic studies enable us to make another step in generalization, such that we could now, metaphorically, focus on food in general, rather than on fruit only, in a strenuous yet constructive effort against reductionism.

On the other hand, on a cautionary metaphorical note, an umbrella is a useful tool, but only if it rains. In addition, it may help staying dry, but cannot save from drowning in case of a flood. Accordingly, an umbrella review, an overview of reviews, or a meta-epidemiologic study including few or faulty primary or secondary studies will most likely have a hard time providing credible and useful conclusions. In addition, while our work aimed to be comprehensive, we have not focused on other types of review, such as rapid reviews or scoping reviews, which are well and poignantly discussed elsewhere (Table 1.1) [12–14, 35, 36]. Moreover, we recommend our readers to also diligently study the Cochrane Handbook for Systematic Reviews of Interventions and the Joanna Briggs Institute Reviewers' Manual, which both provide very useful and sound guidance on how to best conduct an overview of reviews and an umbrella review [7, 8]. Other very important resources, albeit mainly focusing on secondary level research, are the Standards for Systematic Reviews issued by the US Institute of Medicine (IOM), the Methods Guide for Effectiveness and Comparative Effectiveness Reviews issued by the US Agency for Healthcare Research and Quality (AHRQ), and the Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare issued by the UK Centre for Reviews and Dissemination (CRD) [37–39].

This being a multiauthored opus with some overlap between chapters, we can surely expect some apparent discrepancies in the way the contributors have set the boundaries of umbrella reviews, overviews of reviews, and meta-epidemiologic studies or have provided specific recommendations for best practices. Frankly, we are more than happy with that, as no single recipe or formula can be considered correct per se, and the best service we can offer the reader is to help him or her navigate the complexity of evidence synthesis with tertiary level studies, but without overlooking the nuances and the constructive debates that still persist between experts. Given the novelty of the topic and our enthusiasm in leading this authoritative group of international experts, errors, inaccuracies, and typos are also unfortunately likely. We will be more than happy to receive any type of feedback in order to improve the future editions of the book.

Finally, it is paradoxically our hope that this book will have become obsolete in a few years. This would mean in fact that other and better books on the topic have become available or that this scholarly field has progressed so remarkably to

challenge most of what is available here. In the meanwhile, this being the only textbook explicitly dedicated to the fascinating topic of umbrella reviews, overview of reviews, and meta-epidemiologic studies, we humbly recommend its critical albeit constructive perusal.

## References


1. Stevenson RL. The philosophy of umbrellas. In: Madden P, editor. Quotidiana. 1894. 24 Mar 2007. Available at: http://essays.quotidiana.org/stevenson/philosophy_of_umbrellas. Last accessed on 13 Aug 2015.
2. Guyatt G, Rennie D, Meade MO, Cook DJ. Users' guide to the medical literature. In: A manual for evidence-based clinical practice. 2nd ed. New York: McGraw-Hill Professional; 2008.
3. Biondi-Zoccai G, editor. Network meta-analysis: evidence synthesis with mixed treatment comparison. Hauppauge: Nova; 2014.
4. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7:e1000326.
5. Biondi-Zoccai G, Abbate A, Peruzzi M, Frati G. Observations, trials, and meta-analyses: the life cycle of evidence-based endovascular therapy. J Endovasc Ther. 2014;21:693–6.
6. Biondi-Zoccai G, Peruzzi M, Frati G. Which do you like better…a bowl of Cheerios or a Big Mac? Pros and cons of meta-analyses in endovascular research. J Endovasc Ther. 2013;20:145–8.
7. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available from: www.cochrane-handbook.org. Last accessed on 18 Aug 2015.
8. The Joanna Briggs Institute. The Joanna Briggs Institute Reviewers' Manual: methodology for JBI umbrella reviews. Adelaide: The University of Adelaide; 2014.
9. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ. 2009;181:488–93.
10. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J. 2009;26:91–108.
11. Li L, Tian J, Tian H, Sun R, Liu Y, Yang K. Quality and transparency of overviews of systematic reviews. J Evid Based Med. 2012;5:166–73.
12. Varker T, Forbes D, Dell L, Weston A, Merlin T, Hodson S, O'Donnell M. Rapid evidence assessment: increasing the transparency of an emerging methodology. J Eval Clin Pract. 2015 [Epub ahead of print].
13. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc. 2015;13(3):141–6.
14. Hartling L, Vandermeer B, Fernandes RM. Systematic reviews, overviews of reviews and comparative effectiveness reviews: a discussion of approaches to knowledge synthesis. Evid Based Child Health. 2014;9:486–94.
15. Biondi-Zoccai G, Landoni G, Modena MG. A journey into clinical evidence: from case reports to mixed treatment comparisons. HSR Proc Intensive Care Cardiovasc Anesth. 2011;3:93–6.
16. Petticrew M. Time to rethink the systematic review catechism? Moving from 'what works' to 'what happens'. Syst Rev. 2015;4:36.
17. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44:1271–8.
18. Gülmezoglu M, de Onis M, Villar J. Effectiveness of interventions to prevent or treat impaired fetal growth. Obstet Gynecol Surv. 1997;52:139–49.
19. Farquhar C, Rishworth JR, Brown J, Nelen WL, Marjoribanks J. Assisted reproductive technology: an overview of Cochrane reviews. Cochrane Database Syst Rev. 2015;7:CD010537.

20. Conn VS, Coon Sells TG. WJNR welcomes umbrella reviews. West J Nurs Res. 2014;36:147–51.
21. Thomson D. Evidence synthesis in child health: overviews of reviews. Evid Based Child Health. 2014;9:1–2.
22. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol. 2010;63:875–82.
23. Belbasis L, Bellou V, Evangelou E, Ioannidis JP, Tzoulaki I. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. Lancet Neurol. 2015;14:263–73.
24. Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, Heneghan C. The need for randomization in animal trials: an overview of systematic reviews. PLoS One. 2014;9:e98856.
25. Silva V, Grande AJ, Martimbianco AL, Riera R, Carvalho AP. Overview of systematic reviews – a new type of study: part I: why and for whom? Sao Paulo Med J. 2012;130:398–404.
26. Silva V, Grande AJ, Carvalho AP, Martimbianco AL, Riera R. Overview of systematic reviews – a new type of study. Part II. Sao Paulo Med J. 2015;133:206–17.
27. Glass GV. Primary, secondary, and meta-analysis of research. Educ Res. 1976;5:3–8.
28. Zambon M, Biondi-Zoccai G, Bignami E, Ruggeri L, Zangrillo A, Landoni G. A comprehensive appraisal of meta-analyses focusing on nonsurgical treatments aimed at decreasing perioperative mortality or major cardiac complications. J Anesth. 2012;26:509–15.
29. Peruzzi M, De Falco E, Abbate A, Biondi-Zoccai G, Chimenti I, Lotrionte M, Benedetto U, Delewi R, Marullo AG, Frati G. State of the art on the evidence base in cardiac regenerative therapy: overview of 41 systematic reviews. Biomed Res Int. 2015;2015:613782.
30. Griebeler ML, Morey-Vargas OL, Brito JP, Tsapas A, Wang Z, Carranza Leon BG, Phung OJ, Montori VM, Murad MH. Pharmacologic interventions for painful diabetic neuropathy: an umbrella systematic review and comparative effectiveness network meta-analysis. Ann Intern Med. 2014;161:639–49.
31. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7:e49667.
32. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. J Clin Epidemiol. 2012;65:1267–73.
33. Biondi-Zoccai GG, Agostoni P, Moretti C, Meliga E, Sheiban I. Making sense of the recent meta-analytical confusion concerning the safety of drug-eluting stents. EuroIntervention. 2007;3:381–5.
34. Wald NJ, Morris JK. Teleoanalysis: combining data from different types of study. BMJ. 2003;327:616–8.
35. Paré G, Trudel M, Jaana M, Kitsiou S. Synthesizing information systems knowledge: a typology of literature reviews. Inf Manage. 2015;52:183–99.
36. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011;11:15.
37. Eden J, Levit L, Berg A, Morton S, editors. Finding what works in health care. Standards for systematic reviews. Washington: The National Academies Press; 2011.
38. Agency for Healthcare Research and Quality. Methods guide for effectiveness and comparative effectiveness reviews. Rockville: Agency for Healthcare Research and Quality (US); 2008.
39. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in healthcare. York: University of York; 2009.

# Evidence Hierarchy

# 2

Colin Ng and Umberto Benedetto

**Abstract**

Decision-making requires a delicate balance between values, expertise, resources and knowledge. Evidence is the objective dimension of knowledge which can be exploited for decision-making. As the human brain complexity is quite evident, so are evident the complexity and multifaceted feature of evidence informing decision-making in clinical medicine as well as in many different fields of humanity. In particular different sources of evidence can be identified, from less robust, precise and accurate to others which are more robust, more precise and more accurate. Yet, there is a continuum in the hierarchy of evidence, and it would be naïve to think that less formal sources of evidence should be disregarded altogether in comparison to more established and robust ones (i.e. randomised controlled trials). Similarly, other layers of evidence on top of randomised trials can be envisioned going to systematic reviews, to meta-analyses and to umbrella reviews. Only the explicit and conscientious integration of such multiple sources of evidence in a unifying framework can lead to effective and efficient decision-making.

## 2.1 Introduction

Today, medical practice is centred around the idea of knowledge being evidence-based, hence the term evidence-based medicine. What exactly is evidence-based medicine? It is said to be the "conscientious, explicit, and judicious use of current

C. Ng
Department of Cardiac Surgery,
Yong Loo Lin School of Medicine, National University of Singapore,
Singapore, Singapore

U. Benedetto, MD, PhD (✉)
Cardiothoracic Department, John Radcliff Hospital, Oxford University Hospital,
Headley Way, Headington, Oxford OX3 9DU, UK
e-mail: umberto.benedetto@hotmail.com



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_2

best evidence in making decisions about the care of individual patients" [1]. What then is evidence?

Evidence is an observation made in nature. By making observations, we can draw logical conclusions and make inferences about cause and effect. However, these conclusions can be influenced by human emotion and bias, rendering them inaccurate about the original observation. The scientific method was thus developed to minimise such error. The scientific method entails making hypotheses about a particular question, testing, documenting and making conclusions that fit. It also emphasises that experiments should be repeatable and, where appropriate, controls be used.

A fundamental concept in evidence-based medicine is to recognise that not all evidence is equally protected against error. Therefore, decisions should be made on evidence that is more protected against bias and error, by virtue of the method used. Evidence-based medicine necessarily means that higher levels of evidence, systematically and rigorously put together, are of higher value than mere personal observations and unsystematic experience [2], thus the idea of a hierarchy of evidence – different levels of evidence that present conclusions with decreasing levels of chance for bias and error. The top of the hierarchy represents the best available evidence, while the bottom represents less reliable evidence. Let us journey through the different forms of evidence, starting from the lowest, working our way up to the top.

## 2.2 Unsystematic Observations

Before evidence-based medicine became central to clinical practice, practitioners relied on personal experience to make clinical decisions regarding patient care. A physician's personal observations allowed him to form a clinical opinion and make reasonable decisions on treatment for patients. This is a very basic form of evidence and takes the lowest rank in our hierarchy of evidence.

Let us take a look at an example of a doctor who recommends prescribing beta-blockers to patients complaining of palpitations. He has successfully treated many patients with palpitations using beta-blockers. This is his experience with beta-blockers, and it forms the basis of his evidence for his claim that beta-blockers are effective for treating palpitations. He is right in saying that beta-blockers can treat palpitations, because he has made numerous observations that previous patients have recovered while on the medication.

Such forms of evidence were key to the historical practice of medicine, but we now know that it is not as valuable as other forms of evidence. The conclusions drawn from these personal observations suffer from the inherent possibility of containing errors associated with cognitive processes. Such errors include recall bias and summary of experiences [3, 4].

In this case, some patients treated with beta-blockers may have returned complaining of giddiness or other side effects. The doctor in our example did not make a tally of the exact number of patients who came back giddy as these complaints were far and few between patients. As a result, he forms an impression that these side effects are rare and proceeds to conclude that there is no major issue with using the medication. Of course, this is not necessarily a fault on his part – his

observations are sound and he used good clinical judgement. However, what he presents as a recommendation for the treatment of palpitations and the way at which he arrived at this conclusion may have been swayed by natural cognitive processes that introduce elements of uncertainty. Therefore, in the hierarchy of evidence, unsystematic personal observations rank the lowest among other forms of evidence and are considered relatively untrustworthy.

## 2.3 Physiologic and Mechanistic Studies

Knowing that personal unsystematic observations cannot be relied on with certainty, you would like to find out how beta-blockers can actually reduce palpitations by looking for evidence on how the drugs work in the human body. You now turn to a form of evidence known as physiologic and mechanistic studies. These studies focus on the molecular and systemic functions in the human body. A physiologic study entails testing a certain intervention on a population of patients with the intention of examining the mechanisms by which a certain outcome is brought about by the administration of the particular intervention. These studies are usually carried out with animal populations first, before the intervention is deemed reasonably safe to test out on a small population of human participants.

By nature of the small populations investigated in such studies, they do not carry high statistical weight. The value of such studies lies in their ability to provide us with a better understanding of the reasons certain observed effects appear in individuals treated with a particular intervention. Early studies are also useful for identifying patterns in human physiology and pharmacology. They therefore provide early evidence about a particular treatment, and the information from such studies are highly useful for future studies that provide higher levels of evidence as we will see later.

Looking at a physiologic study, you find out that beta-blockers bind to specific receptors in the autonomic nervous system. This prevents excitation of the sympathetic nervous system, thus halting certain arrhythmias that can cause palpitations. You now know that beta-blockers are useful for treating palpitations, but you can't be certain how effective they are – what is the rate of cure in a population of patients with palpitations, and what are the rates of specific side effects? Are there perhaps some serious side effects?

## 2.4 Case Reports

Case reports and case series document rare or unexpected observations in patients. A case report contains a description of a patient's background, history of presenting complaint, physical examination findings, investigation results and treatment outcomes. Most importantly, they document the particular unexpected finding and provide a discussion on the possible reasons for such an outcome. At times, a case series might pave the way for future clinical research in terms of larger studies, the types of which we shall examine in detail later in this chapter.

Owing to the inherent rarity of the case, the outcomes observed are often not readily reproducible in many other patients. A case series would be especially useful for studying trends in harmful effects of treatment or intervention. It would be unethical and unacceptable to design a study specifically investigating harmful effects on patients, thus making a case series relevant.

A series of three patients reports hallucinations associated with the use of beta-blocker therapy [5]. The majority of patients on beta-blockers do not experience this, but the rare side effects of this drug are helpful to the doctor about to prescribe the medicine. The doctor could rely on such evidence to warn the patient of the rare yet potentially more disturbing side effects, in addition to the common but less serious ones. Unlike personal experience and observation, case reports provide a more systematic way of documenting observations, as they require formal documentation of the patients' clinical parameters and circumstances.

## 2.5 Case-Control and Cross-Sectional Studies

Case-control studies are retrospective studies that analyse risk factors or protective factors in specific populations of patients with an outcome of interest. "Cases" refer to patients who have the outcome of interest. "Controls" are healthy individuals who do not have the outcome. Analysis is then performed to identify factors that are more prevalent in either group. These factors may be risk factors or protective factors, depending on which group they prevail in. To understand if a particular case-control study is worthy of being considered robust evidence, we should examine the differences between the two groups investigated [3]. Most importantly, we should ask: do the two groups differ in areas other than the outcome itself? It is also important to have a rigorous process to ensure that the diagnosis is correctly made (and conversely that those in the control group do not have the diagnosis of the particular condition) and that the exposures properly documented.

Cross-sectional studies are designed to observe for correlation between factors and outcomes in populations of patients at a specific time point, with no follow-up. A cross-sectional study can be performed on any group of individuals. The key to conducting such a study is to meticulously document the factors and outcomes, so that a thorough analysis can be performed later on. These studies provide information on the prevalence of a certain condition and can be analysed in ways that show association [6]. It is prudent to note that these studies are unable to show the direction of the association. For example, cross-sectional data may reveal that the prevalence of beta-blocker used in a palpitation-free group of patients is higher than in a group of patients with palpitations. However, we do not know if beta-blocker usage is the cause of the lack of palpitations. All we know is that there happens to be more people using beta-blockers in a group of palpitation-free individuals than a group of patients who suffer from palpitations. Therefore, we must be careful to never mistake an association found in a cross-sectional study for a cause-and-effect relationship.

In most case-control and cross-sectional studies, bigger numbers of patients are recruited, providing relatively higher statistical power than earlier forms of evidence. This certainly does not mean that the results obtained are not attributable

solely to chance; statistical tests will help to decide this. In the absence of higher levels of evidence, such studies can shed light on the association between exposures and certain outcomes as well as the factors that protect or put an individual at risk of developing certain outcomes.

## 2.6 Cohort Studies

A cohort study differs from a case-control study in that participants are enrolled and grouped based on their exposure. The risk of developing a specific outcome is compared among groups that were exposed to various treatments or even risk factors and protective factors. If the study is planned before the participants begin follow-up, the study is considered to be prospective. The participants are exposed to the treatment, and they are required to return for follow-up to see if a specific outcome is observed. A study that is planned after follow-up is complete is termed a retrospective study. Such retrospective studies rely on registries or databases of patient information with data regarding the treatment received and the subsequent outcomes. Statistical analyses are then performed to assess the incidence of a certain outcome in groups of various exposures or intervention.

In a simple example, an aspiring researcher wants to know if beta-blocker therapy can relieve palpitations. The study that he designs includes patients who were previously treated with and without beta-blockers. They were followed up and the outcome of interest was the abatement of the palpitations. It is a retrospective cohort study. Unlike a cross-sectional study, we can now start to establish some cause and effect. Many of those that were treated with beta-blockers later had good symptomatic relief, while those that were not treated continued to experience palpitations. He then concludes that beta-blockers can cure palpitations.

Some cohort studies are designed to follow up patients for long periods of time, making them effective in picking up late outcomes after prolonged exposure. Examples of which include post-marketing surveillance studies. Such studies also allow researchers to chart the progression of a particular disease and make more accurate estimations about the course of the illness in the long run. They are however prone to selection bias as the choice of intervention is influenced by a doctor's clinical judgement. Such a limitation may be better handled by randomisation of subjects, as we will soon see. Having considered this, we can still accept cohort studies to be a moderately reliable form of evidence, given the systematic approach used in studying effects in patient populations.

## 2.7 Randomised Trials

Progressing through the levels of evidence, we finally reach some of the highest levels of evidence in the hierarchy. Randomised controlled trials have long been considered the gold standard for evaluating treatment. In a randomised controlled trial, individuals are randomly assigned to receive either a particular intervention or to a control arm. The control arm does not necessarily mean that the patients in

that group receive no treatment. They can receive a placebo treatment or the previous standard therapy. Conclusions can then be drawn on whether the intervention was effective in modifying the outcome. These studies are powerful because randomisation is more likely to balance the characteristics of both groups in the study. It also prevents participants and clinicians from choosing or assigning to a particular arm of the trial. This essentially means that the chance of selection bias is kept minimal and that the groups are more representative of a real-world population.

An even more powerful form of a randomised trial involves blinding. This means that participants are not made aware to the treatment that they receive. For example, in a beta-blocker versus placebo study, both groups receive a pill – one group receives the beta-blocker and another takes the placebo pill. In some cases, blinding of participants may be prohibitively hard – for example, when investigating medical versus surgical therapy. It is unlikely for a patient to be unaware that they had gone for an operation as opposed to having been treated by drugs alone. A second level of blinding may be applied to outcome assessors and investigators, so as to reduce bias of reporting.

A double-blind randomised controlled trial therefore qualifies as a strong piece of evidence. It will show us that in randomly chosen individuals who are fairly representative of the general population, whether a certain intervention can bring about specific outcomes. Blinding further reduces the chance for patients or investigators to influence the final results.

Yet, as with other forms of studies, a randomised trial has its fair share of problems. Notably, narrow inclusion and exclusion criteria are usually employed to increase the internal validity of the study. This means that specific sample populations are chosen to demonstrate the effect of a selected intervention. It may result in a larger effect being demonstrated in the selected population than in those that were excluded from the study. This may be seen as a form of selection bias. Some trials investigating novel treatments may be funded by manufacturers or research companies. This has the potential to influence the completeness of data reported or even the fundamental methods used in the study, possibly leading to a conclusion that is in the favour of the treatment that was newly developed. Publication bias is also a factor to consider when evaluating randomised controlled trials. This refers to non-publication of trials with unfavourable results. A slightly different type of bias is outcome reporting bias, which refers to outcomes that are not reported in the final publication if they are deemed unfavourable, even though they were investigated in the study.

Having considered the weaknesses of such trials, we should appreciate that a single randomised trial does not necessarily provide the complete picture. Yet, we can see that such trials are designed with the aim of reducing selection bias and are thus good evidence in this respect.

## 2.8 Systematic Reviews and Meta-analyses

Now that we have several randomised controlled trials documenting the efficacy of beta-blocker treatment for palpitations, the next step would be to take look at these trials to see if they are in agreement. A systematic review is a review of existing

literature in a logical and ordered fashion to summarise the evidence that is available. This allows the reviewer to take a step back and observe for similarities and differences in trials and studies relating to a particular subject. By doing so, patterns can be identified and a final conclusion can be drawn about outcomes and other factors being investigated. A meta-analysis goes one step further by using statistical methods to perform a quantitative synthesis of existing data. This allows us to look for precise trends in the published data and make new conclusions or reinforce existing ones [7].

When considering meta-analyses, it is important to take into account the variability between studies (or heterogeneity). Some feel that meta-analyses should be considered if the studies used are largely homogenous in the sample populations, treatments and outcomes investigated. Otherwise, meaningful conclusion may not be drawn from the research. Systematic reviews are vulnerable to biases including publication bias [8]. This could mean that the pooled analysis inadvertently excludes studies that were not published in the first place because the results were deemed unfavourable for publication. Outcome reporting bias also affects the quality of a systematic review – the estimated effect of an intervention on an outcome may be skewed to favour an outcome that is reported by more studies.

Going beyond meta-analyses, we can make use of new ways to summarise and synthesise data from existing evidence in the form of network meta-analyses and umbrella reviews. A network meta-analysis allows for indirect comparison of treatments for a specific outcome [9]. For instance, some trials compare beta-blockers to placebo for treating palpitations, while some compare different beta-blockers. All these trials can be linked, and we can see indirect comparisons between the different treatments explored. Advanced statistical methods further allow us to rank benefits and risks of the treatments investigated and pool together huge amounts of evidence from numerous studies to draw meaningful conclusions like never before [10]. Umbrella reviews critically evaluate meta-analyses to study trends from similar studies and make conclusions about the big picture in a particular disease.

We have thus seen that such collective approaches to evidence gathering top the evidence hierarchy, especially when the statistical methods used to derive conclusions are performed properly and top quality randomised trials are included in the synthesis.

**Conclusions**

As a clinician, it is important to understand that evidence is all around us. As we strive to provide the best possible care to our patients, we should focus on evidence that is reliable, by virtue of the statistical methods used. As researchers, we aim to expand the body of medical knowledge, and we should recognise that providing good quality evidence is key to making sound conclusions. Rigorous statistical methods such as randomisation and blinding and statistical tests for bias and heterogeneity should be employed when studying effects in populations of patients. Only by doing so can the evidence be put forth in a way that is as unbiased as possible. When evaluating evidence, one should always be alert to possible sources of bias and understand the merits of each type of evidence, as well as their limitations.


**Fig. 2.1** The hierarchy of evidence, from the lowest to the highest (Adapted from EBM Pyramid by Jan Glover, David Izzo, Karen Odato and Lei Wang)

We have worked our way up from the most rudimentary forms of evidence that is available all around us to more sophisticated forms of evidence that can provide us with relevant information about treatment, outcomes, cause and effect. The hierarchy of evidence is best summarised in Fig. 2.1. Notice how the higher levels of evidence build upon evidence lower down in the hierarchy. Even though we have seen that lower-ranking evidence is less conclusive and trustworthy, they are excellent material and starting ground for new research and their utility should not be discounted.

The ultimate level of evidence, as we know it, is the synthesis of new evidence by pooling together existing high-level evidence. That is the role of the meta-analyst and systematic reviewer. If a large, well-designed randomised controlled trial provides us with good evidence on a subject matter, we can only expect a more unquestionable conclusion with the results of several of such trials collectively examined. The onus is therefore on the astute clinician researcher to examine all available evidence and choose which ones to accept, which ones to reject and which ones to advance.

## References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312:71–2.
2. Guyatt G, Rennie D, Meade MO, Cook DJ. User's guide to the medical literature: a manual for evidence-based clinical practice. 2nd ed. New York: McGraw-Hill Companies; 2008.
3. Burneo JG, Demaerschalk BM, Jenkins ME. Neurology: an evidence-based approach. New York: Springer Science & Business Media; 2011.
4. Kahneman D, Tversky A. On the reality of cognitive illusions. Psychol Rev. 1996; 103:582–91.
5. Goldner JA. Metoprolol-induced visual hallucinations: a case series. J Med Case Rep. 2012; 6:65.
6. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emerg Med J. 2003;20:54–60.
7. Guyatt G, Rennie D, Meade MO, Cook DJ. User's guide to the medical literature: a manual for evidence-based clinical practice. 2nd ed. London: AMA; 2002.

8. Montori VM, Smieja M, Guyatt GH. Publication bias: a brief review for clinicians. Mayo Clin Proc. 2000;75:1284–8.
9. Biondi-Zoccai G, Abbate A, Benedetto U, Palmerini T, D'Ascenzo F, Frati G. Network meta-analysis for evidence synthesis: what is it and why is it posed to dominate cardiovascular decision making? Int J Cardiol. 2015;182C:309–14.
10. Biondi-Zoccai G, editor. Network meta-analysis: evidence synthesis with mixed treatment comparison. Hauppauge: Nova Science Publishers; 2014.

# From Qualitative Reviews to Umbrella Reviews

3

Ana Ortega, Eduardo Lopez-Briz,
and María Dolores Fraga-Fuentes

**Abstract**

The volume of health-care literature is growing at an increasing rate, with a huge amount of studies difficult to process. Therefore, we need tools or techniques to synthesize the information to help us in clinical decision-making. In fact, the available body of evidence ranges from single studies to umbrella reviews. In this scenario, evidence-based clinical decision-making requires knowing what type of evidence to use in every situation. However, a prerequisite for optimal decision-making is a greater understanding by professionals of the different techniques used to analyse their strengths, limitations and utilities. The purpose of this chapter is to take a journey from qualitative reviews to umbrella reviews. We start the tour on a fundamental point: term definitions, showing the variability among different authors. We go on to describe the differences, advantages, disadvantages and uses of different types of evidence, from individual studies to the 'more specific methods' for knowledge synthesis, both qualitative and quantitative syntheses (systematic reviews, meta-analysis, network meta-analysis). Finally, in the last part of our journey, we compare the strengths and weaknesses of different evidence synthesis methods from the more traditional or specific to the more general or broader reviews (umbrella reviews, overviews of reviews, meta-epidemiologic reviews). Systematic reviews are at the top of the evidence pyramid. However, the number of systematic reviews published is increasing at

A. Ortega, PhD (✉)
Pharmacy Services, Clínica Universidad de Navarra, Pio XII, 36, Pamplona 31008, Spain
e-mail: aortega@unav.es

E. Lopez-Briz
Pharmacy Department, Hospital Universitario y Politécnico La Fe, Valencia, Spain

M.D. Fraga-Fuentes
Pharmacy Services, Hospital General La Mancha Centro, Alcázar de San Juan, Spain



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_3

a high rate, and decision-makers need to evaluate more evidence to answer their questions. Systematic reviews of existing systematic reviews, known as umbrella reviews, provide an overall examination of the body of information that is available for a given topic. Despite the limitations and weaknesses of tools to appraise and synthesize evidence, systematic reviews and umbrella reviews, including overviews of reviews and meta-epidemiological studies, continue to be the best tool for an approximation to the truth, in evidence-based terms.

*What information consumes is rather obvious: it consumes the attention of its recipients.*

Herbert Simon (1916–2001) – Economic Sciences Nobel Prize 1978

## 3.1 Introduction

When medical decisions are taken, there is a need to combine clinical evidence with data from the treated patients. The ever-increasing volume of evidence makes it necessary to create forms of evidence synthesis which enable the integration of diverse types or pieces of evidence into a whole which can then be usefully consulted when the decisions are taken.

When the body of the evidence comprises only a few clinical trials, it is well accepted to carry out a systematic review which may or may not include a meta-analysis. This is sometimes called secondary research, as opposed to primary research that includes clinical trials, observational studies, case studies.

However, in practice this is not the most frequent case as there are often epidemiological studies, in addition to or instead of randomized trials, or case reports with unique information. Sometimes there are no clinical trials that compare the treatments which are being considered as possible options. Furthermore, on other occasions, there are several meta-analyses or systematic reviews of the same topic, there is a need to answer different correlated questions and so on.

In these cases, there is a need for other tools or techniques to compile the information, a set of tools with an established methodology designed to avoid potential bias and different from classical systematic reviews or meta-analyses. These include network meta-analyses, umbrella reviews, overview of reviews and meta-epidemiological studies.

Each of these methods has different characteristics presenting advantages and disadvantages when compared to each other and each of which can be useful in different circumstances.

Overviews have recently become exceedingly common for different reasons [1]. One of these is the increasing number of systematic reviews and the increased use of practice guidelines that often require answers to not only one but several linked questions.

Tying questions together in an analytic framework has become the standard practice for full evidence reviews, and so overviews are a tool used to increase efficiency when managing information. This is particularly useful, as for any one clinician to gather and synthesize the evidence to answer very complex clinical issues would be very time consuming [2].

In this chapter, firstly we are going to define these synthesis methods and show the controversies in the terminology. Secondly, we will analyse the differences, advantages, disadvantages and uses of 'specific reviews' (traditional secondary research) vs single reports (primary research). In specific reviews we can include qualitative reviews, systematic reviews, meta-analysis and perhaps network meta-analysis. As single reports or primary research, we can consider case studies, clinical trials or observational studies. And, finally, we will deal with the comparison, advantages and disadvantages and applications of broad or more 'general' reviews vs the 'specific' or more 'traditional' reviews. Among these more general reviews, we include umbrella reviews, overviews of reviews and meta-epidemiologic reviews. We will not focus on the differences between the latter, as this will be the focus of the next chapter, but on the differences between the specific and more general reviews.

According to some authors [3], the common feature of all systematic reviews and overviews of reviews is that the author seeks to collate data from many studies into one product to inform and facilitate evidence-based decision-making; they differ in the unit of synthesis (primary studies vs other synthesis) and how data are analysed.

Every method has its place in the synthesis and analysis of the evidence, and each can be used to help decision-making. However, it is important to understand the strengths, limitations and utilities of each method in order to select the most suitable strategy or strategies in each situation.

## 3.2 Definitions

Definitions of primary studies are universal and widely known and can be found in PubMed [4], for example, clinical trial and observational study. A clinical trial is described as a 'pre-planned clinical study of the safety, efficacy, or optimum dosage schedule of one or more diagnostic, therapeutic, or prophylactic drugs, devices, or techniques in humans selected according to predetermined criteria of eligibility and observed for predefined evidence of favourable and unfavourable effects'. And an observational study is a 'clinical study in which participants may receive diagnostic, therapeutic, or other types of interventions, but the investigator does not assign participants to specific interventions'.

Reviews and meta-analyses are also included in PubMed. A review corresponds to 'published materials which provide an examination of recent or current literature. Review articles can cover a wide range of subject matter at various levels of completeness and comprehensiveness based on analyses of literature that may include research findings. The review may reflect the state of the art. It also includes reviews as a literary form'.

Meta-analysis is a 'quantitative method of combining the results of independent studies, usually drawn from the published literature, and synthesizing summaries and conclusions which may be used to evaluate therapeutic effectiveness, plan new studies, etc., with application chiefly in the areas of research and medicine' [4].

It should be noted that a systematic review is not included in PubMed as a Medical Subject Heading (MeSH) term but is included as a publication type.

However, the terms qualitative review, network meta-analysis, umbrella review and overview of reviews are not included as MeSH terms or publication types.

All qualitative reviews, reviews, systematic reviews, meta-analyses, network meta-analyses, umbrella reviews and overviews of reviews are all reviews and methods of evidence synthesis. To review means 'view, inspect, or examine a second time or again' [5]. And synthesis can be defined as [6] 'the contextualization and integration of research findings of individual research studies within the larger body of knowledge on the topic. A synthesis must be reproducible and transparent in its methods, using quantitative and/or qualitative methods. It could take the form of a systematic review; follow the methods developed by The Cochrane Collaboration; result from a consensus conference or expert panel and may synthesize qualitative or quantitative results'.

However, there are some differences between reviews, even if all of them can be systematic. Systematic means that all steps underlying the reviewing process are explicitly and clearly defined and can be reproduced independently by other researchers [7]. Not all systematic reviews include a meta-analysis, as not all topics are suitable for sound and robust data pooling. At the same time a meta-analysis can be conducted outside the realm of a systematic review. But in this case the result of the meta-analysis should be best viewed as hypothesis-generating only, as it carries a high risk of bias.

A systematic review is a method that collates evidence from multiple primary studies using explicit systematic methods to answer a specific question. It can analyse a single intervention or a group of interventions [3]. It combines the evidence of multiple studies by identifying relevant research, appraising study quality and summarizing findings [8]. It can be conducted using a statistical analysis or a narrative analysis. Some authors [7] consider that systematic reviews focus on a clinical problem, which can be therapeutic, diagnostic or prognostic.

The Cochrane Collaboration defines [9] systematic reviews as 'reviews of a clearly formulated question that use systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of the included studies'.

Systematic reviews can be used to address different research questions; the Canadian Institutes of Health Research (CIHR) [6] indicate some of them:

- What are the benefits and harms of treatment 'X' in humans?
- What are the benefits and harms of a new service delivery configuration?
- What are the benefits and harms of a quality improvement initiative?
- What is the accuracy of diagnostic test 'X'?
- What is the accuracy of routine coding following hospital discharge?
- What are the experiences of patients undergoing treatment 'X'?
- What is the prevalence of condition 'X'?
- How strong is the association between gene 'A' and disease 'X'?

Meta-analysis is a technique that statistically combines the results of quantitative studies to provide a more precise effect of the results [5], providing a single estimate

of the effect [10]. A good systematic review is essential for a meta-analysis of the literature. Some authors consider it a type of systematic review [8]. But, for others, meta-analysis is a study using specific statistical methods for pooling data from separate datasets [7], not necessarily a systematic review or even a review. Clinical trials and/or observational studies can be combined in a meta-analysis.

Qualitative systematic review/qualitative evidence synthesis is defined by Grant and Booth [5] and others [8] as a method for integrating or comparing the findings from qualitative studies. It looks for 'themes' or 'constructs' that lie across individual qualitative studies. The goal in general is not aggregative in the sense of 'adding studies together' as with a meta-analysis. On the contrary, it is interpretative in broadening understanding of a particular phenomenon. No method for synthesizing qualitative research has become well established [8].

Some authors [3] include the term comparative effectiveness reviews as different from other reviews. These are reviews that bring together evidence from individual studies on the relative benefits, or harms, of a range of interventions for a given problem or condition. They can use methods such as network meta-analysis. In our opinion, this methodology does not differ from systematic reviews or network meta-analysis, although comparative effectiveness research is designed to improve informed health-care decisions by choosing the best option.

Other authors identify mixed studies review/mixed methods review as any combination of methods where one significant component is a literature review (usually systematic). Within a review context, it refers to a combination of review approaches, for example, combining quantitative with qualitative research or outcome with process studies [5]. Noyes et al. [11] identified two broad approaches that can be used to integrate qualitative and quantitative findings: multilevel synthesis and parallel synthesis.

Network meta-analyses are also referred to as mixed treatment comparisons or multiple treatment meta-analysis and are reviews that include many competing treatments and combine both direct pairwise comparisons and indirect comparisons across a network of randomized trials to infer the relative effectiveness of multiple interventions [12–14]. This method allows for the ranking of treatments according to one or more criteria, effectiveness, safety, cost-effectiveness, etc.

As a generic term, overview is a summary of the medical literature that attempts to survey the literature and describe its characteristics [5].

Overviews have been referred to as umbrella reviews, meta-reviews, overviews of systematic reviews, reviews of reviews and systematic reviews of systematic reviews, among others [3]. These terms can be used without distinction; however, we think that there are some differences between some of these methods.

Overviews of systematic reviews compile data from multiple systematic reviews relevant to a single health problem using a format and methods similar to those of systematic reviews [3]. A similar definition is given in the Cochrane handbook [9]; they are reviews designed to compile evidence from multiple systematic reviews of interventions into one accessible and usable document. They aim to provide a comprehensive synthesis of the evidence examining different interventions for the same condition, different outcomes for the same intervention in the same condition, the

same intervention for different conditions, or populations, or adverse effects from the same intervention across multiple conditions [3, 8]. The majority of existing overviews provide a narrative or qualitative synthesis of the findings from each systematic review and the summary estimates from the original pairwise meta-analysis.

The Cochrane Multiple Interventions Methods Group (CMIMG) in 2011 decided to change the emphasis of an overview. Overviews of reviews should not simply summarize systematic reviews; rather, they should integrate or synthesize the evidence from existing systematic reviews and should address a well-defined clinical question [15]. Some authors consider that few systematic reviews provide much guidance on the specific forms of an intervention that should be used in different circumstances. These authors propose an intervention options table to translate review findings into evidence-based practice. The intervention options table should provide a summary of the usable and feasible interventions in the review, with information on the criteria on which users would base their choice [16]. The CMIMG considers that if an intervention review seeks to compare multiple interventions (i.e. to determine an ordering of three or more competing interventions for an outcome), this should be explicit in the protocol, and appropriate methods should be planed and implemented [15].

Umbrella reviews can be considered a broader term that includes overviews of reviews and meta-epidemiological studies.

An umbrella review specifically refers to a review of compelling evidence from multiple reviews into one accessible and usable document. It focuses on a broad condition or problem for which there are competing interventions and highlights reviews that address these interventions and results [3]. Methodologically speaking, they search for reviews, nor for primary studies, and the synthesis is graphical and tabular with a narrative commentary. Its purpose is to analyse what is known, make recommendations for practice, identify what remains unknown and make recommendations for future research [3]. According to this definition, the term can be considered to be similar to overviews of systematic reviews.

The definition of umbrella review by Ioannidis [17] is slightly different. He considers umbrella reviews to be systematic reviews that consider the many treatment comparisons for the management of the same disease. Each comparison is considered separately, and meta-analyses are performed and deemed appropriate. Umbrella reviews are clusters that encompass many reviews.

In the glossary of Joanna Briggs Institute, the term umbrella review is also considered to be synonymous with the term overview of reviews. An umbrella review is a term applied to systematic reviews that draw together evidence from a series of other systematic reviews. This type of review can be useful in providing an overview of research within a particular area [18].

Umbrella reviews have the aim of summarizing available evidence and can be used to assess similarities and differences in published reviews to summarize what is known about a topic and typically involve a number of different types of synthesis [8].

Some reviews go further and consider not only diverse interventions on a given disease but also evidence on many diseases or conditions. These are called domain analysis or meta-epidemiologic research [17].

However, Bae et al. [19] consider that meta-epidemiology is now in the process of being recognized as another epidemiological research methodology and consists of controlling meta-confounders. It is based on combinations of epidemiology and meta-analysis to describe the distribution of research evidence for a specific question, to examine heterogeneity and associated risk factors, to control bias across studies and to summarize research evidence as appropriate.

Greenhalgh et al. [20] developed an innovative interpretive method and 'meta-narrative synthesis' for synthesizing conceptual and empirical evidence from heterogeneous sources for a synthesis of diffusion of innovations in service organizations. Recognizing the breadth of relevant research from diverse research traditions and sectors that could contribute to the synthesis question, they used 'the unfolding "storyline" of a research tradition over time' as their unit of analysis.

Some authors also define scoping reviews [8] as a knowledge synthesis method to summarize a range of evidence in order to understand broadly what is known about a phenomenon. The aim is to identify rapidly the key concepts underpinning a research area and the main sources and types of evidence available. These can be undertaken as stand-alone projects, especially in a complex area or one that has not yet been comprehensively reviewed. Scoping reviews are different from systematic reviews in their broad approach to a topic, purposive sampling frame and identification of gaps in the literature.

The CIHR [6] states that scoping reviews 'aim to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available'. The CIHR defines scoping reviews as 'exploratory projects that systematically map the literature available on a topic, identifying key concepts, theories, sources of evidence and gaps in the research' and notes. 'They are often preliminary to full syntheses, undertaken when feasibility is a concern – either because the potentially relevant literature is thought to be especially vast and diverse (varying by method, theoretical orientation or discipline) or there is a suspicion that not enough literature exists'. A scoping review might consider both empirical and conceptual research and often focuses on broader questions than those considered in other syntheses.

In general, the stages of a scoping review are similar to those of a systematic review and involve the 'systematic selection, collection and summarization of existing knowledge in a broad thematic area' [3]. However, synthesis teams may reduce the scope of searches depending on the breadth of the scoping review and available resources. Likewise, scoping reviews do not often undertake detailed appraisal of identified evidence sources and detailed synthesis of the results from studies. Instead they frequently collate the identified evidence using some form of 'analytical framework or thematic construction in order to present a narrative account of the existing literature'.

Arksey and O'Malley [21] identified different purposes for scoping reviews in a broader term including realist syntheses used to explore theoretical and empirical evidence from different sectors and disciplines; examining the extent, range and nature of research activity (to provide an overview of the available literature and identify key themes and research foci); determining the value of undertaking a full systematic review (e.g. by identifying the extent of relevant literature and absence

of existing relevant reviews); summarizing and disseminating research findings across a body of research evidence; and identifying research gaps in the existing literature to aid planning and commissioning of future research (e.g. by identifying whether a research question is likely to have already been answered by existing studies and by refining the research questions and research methods for new studies to ensure that they are informed by existing studies). The scoping review has become increasingly popular as a form of knowledge synthesis. However, a lack of consensus on scoping review terminology, definition, methodology and reporting limits the potential of this form of synthesis [22].

## 3.3 From Single Studies (Primary Research) to 'Specific' Reviews (Traditional Secondary Research)

### 3.3.1 Primary Research and Secondary Research Relationship

Science is a cumulative process that develops iteratively; few studies by themselves are sufficiently persuasive to change practice or policy [6]. To make clinical decisions based on the best evidence requires the need to know what type of evidence is used in every situation. Scientific and ethical justification for new clinical trials requires them to have been designed in the light of relevant previous research or because there is no evidence. Although clinical decisions are ideally informed by a systematic review or in general by evidence synthesis, secondary research is built on primary research. The hierarchy of evidence-based medicine goes from primary research, from in vitro studies to randomized clinical trials, to the context of secondary research from editorials to meta-analyses (pairwise meta-analysis or network meta-analysis) [23]. There is, therefore, an interrelation between primary and secondary research (Fig. 3.1), and each has their advantages and disadvantages. For evidence-based practice both are important, primary results as well as a synthesized summary of all evidence within a particular domain.

Designing trials in isolation or with non-scientific priorities creates fragmented, irrelevant evidence [24]. Individual studies may be misleading due to chance or bias. Ioannidis et al. [24] have undertaken a landmark series of studies exploring the evolution of basic and applied research that highlights concerns about the reliability and interpretation of individual studies. They observed that the results of the most highly cited basic science and clinical research papers published in the most prestigious journals are frequently overturned or challenged by subsequent less prominent publications. Furthermore, they observed that the results of early publications in both basic and clinical research were often likely to report more strikingly positive or negative findings than subsequent publications. Together these studies highlight the problems of focusing knowledge translation efforts on individual studies (especially early publications with striking findings) and suggest that the evidence base in any field needs to mature and be synthesized before an observer can reliably understand its implications [6].

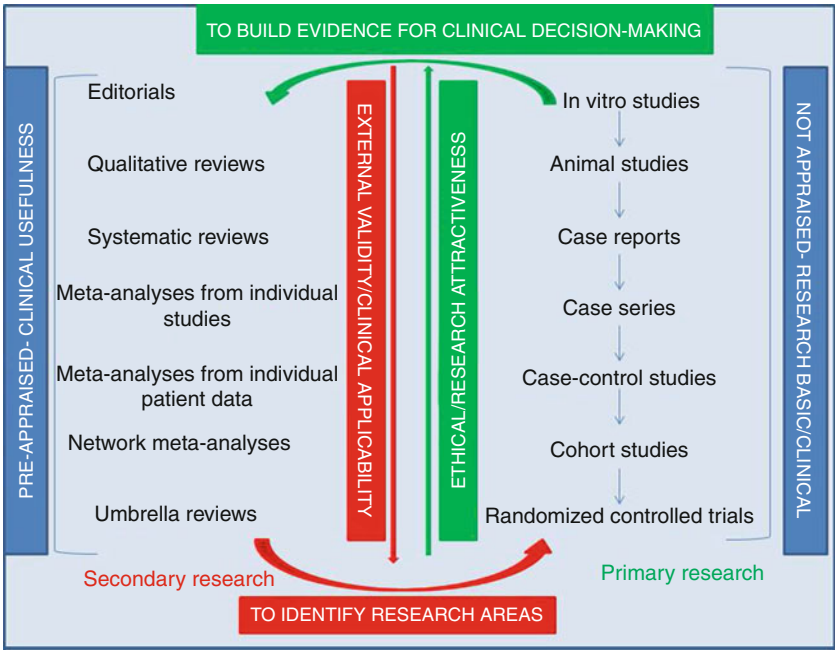



**Fig. 3.1** Relationship between primary and secondary research (Adapted from Biondi-Zoccai et al. [23])

Recently, Ebrahim et al. [25] conducted a review to identify published reanalyses of randomized clinical trials. They identified 37 reanalyses of patient-level data from previously published randomized clinical trials; approximately a third (35 %) of the published reanalyses led to changes in findings that implied conclusions different from those of the original article about which patients should be treated. It seems obvious that evidence updates are necessary, as well as having tools to synthesize the overwhelming volume of clinical research, to allow decisions based on the best evidence.

Credibility, replication and translation are all desirable properties of research findings. Reliable interpretation of the results of new clinical trials entails setting them in the context of updates of the reviews upon which they were deemed scientifically and ethically justifiable [26].

### 3.3.2 Knowledge Synthesis

Most syntheses are conducted either for the purpose of 'knowledge support' or for 'decision support'. Syntheses for knowledge support are confined to summarizing the evidence around a specific question or issue and do not undertake additional tasks to support a decision in a particular context, whereas syntheses for decision support will commonly include some or all of the following steps: engagement of

the decision-making audience in the development of the research question and synthesis protocol, consideration of several related questions using appropriate methods, deliberative process of engaging the decision-making audience to interpret and contextualize the results of the synthesis and the development of context-specific recommendations [6].

The huge amount of literature has led reviewers to perform evidence syntheses on reviews instead of primary studies [27]. For example, Bastian et al. [28] indicate that 11 systematic reviews and 75 trials would need to be read every day in order to keep up to date, when just considering the publications listed in Medline.

Among the different methods for knowledge synthesis, we can distinguish:

- Those with a focus on synthesizing the results of primary studies: systematic reviews, including narrative synthesis, meta-analysis with statistical synthesis, qualitative synthesis and mixed studies reviews
- Those which focus on broad and diverse bodies of research evidence: network meta-analyses, scoping reviews, umbrella reviews and overviews of reviews, among others, although network meta-analysis could also be included in the previous group

In the past, individuals often considered experts in the field have conducted narrative reviews of the literature associated with a particular health condition using informal and subjective methods to collect and interpret information. However, this has two major problems: the review does not provide detailed information of the process, and readers cannot replicate or verify the results and conclusions [2].

Despite the fact that narrative synthesis was one of the most common approaches to synthesis, there is surprisingly little guidance to how they should be conducted. And there is a huge risk of readers reaching inappropriate and misleading conclusions.

To make useful evidence-based clinical decisions, all knowledge synthesis methods must include a systematic and auditable approach to ensure that individual studies have trustworthy findings and that the synthesized findings accurately represent the synthesis of individual studies [8].

The purpose of systematic reviews is to collate relevant evidence from individual studies to answer a specific research question. They use explicit systematic methods to be as comprehensive as possible and to minimize the bias in the results and conclusions [3]. As they are useful for decision-making, publications including systematic reviews have increased exponentially.

Systematic reviews have nowadays been widely accepted as the backbone of good health care. They provide a synthesis of evidence for practitioners, for clinical practice guideline developers and for those designing and justifying primary research; therefore, their quality, validity and credibility are crucial for patients, professionals and society. Having an up-to-date and comprehensive review is therefore important in a context of rapidly increasing knowledge [29, 30].

There are many advantages to systematic reviews [3]:

- They are a well-established form of evidence synthesis.
- Their quality is good because of well-established methods.
- They are feasible and efficient. As they focus on specific interventions and/or populations, they include a relatively small number of studies and are therefore more feasible and efficient to conduct.
- They are the preferred method of choice, particularly when all trials are underpowered [31].
- They have the ability to point out weaknesses and fallacies in apparently sound primary studies [7].

As the number of qualitative studies in several scientific areas of interest increases, the need to systematically synthesize the findings increases as well. In a qualitative synthesis, primary qualitative studies are integrated to develop a theory or evidence-based interventions [32]. In general, the syntheses of qualitative research adopt the same steps as systematic reviews of quantitative evidence. However, systematic reviews of qualitative evidence pose considerable conceptual and methodological challenges, particularly relating to the identification of relevant studies (indexing of qualitative studies in databases remains poor, so it is necessary to develop sensitive search strategies to identify qualitative research studies), appraisal of included studies (there is insufficient evidence to inform a judgement on the rigour and added value of various approaches) and methods of synthesizing evidence (innovative approaches are emerging to synthesize qualitative evidence) [6]. The main disadvantage of qualitative synthesis is that there is no consensus on how to evaluate quality, and it has not yet been determined how to incorporate quality scores into the reporting of qualitative synthesis techniques [8].

Therefore, in general, quantitative systematic reviews are preferred to qualitative systematic reviews; however, sometimes the need for a quantitative response is urgently required, and studies are combined regardless of their appropriateness, leading to erroneous conclusions, so although they may be more precise, they may be as biased as narrative reviews [2].

Excluding these cases, in general, meta-analyses are the preferred option. However, a meta-analysis is not always possible. A good systematic review is essential to a meta-analysis of the literature. But a systematic review can be conducted without combining statistically the results of the different studies as in a meta-analysis.

Meta-analysis requires all included studies to be sufficiently similar, in the studied population, with the interventions used, and the comparisons made. Most importantly they require that the same measure or outcome be measured in the same way, at the same time intervals [5].

Meta-analysis increases the precision in estimating effects compared to individual trials and contributes to the generalizability of study results. In addition, meta-analysis may allow early detection of beneficial or harmful treatment effects where individual studies fail to provide reliable treatment estimates and can settle controversies arising from apparently conflicting studies. However, inappropriate meta-analyses can either lead to false-negative or false-positive results. Therefore, rigorously conducted

systematic reviews and meta-analyses are essential for evidence-based decision-making in clinical practice as well as at the health policy level [33].

Glasziou et al. [31] suggest, for clinicians and guideline authors, that when a systematic review or a meta-analysis is not available, those needing an answer to a clinical question, but without the time or resources to undertake a meta-analysis, might search the most precise well-conducted trial and carefully check whether the study was sufficiently large and adequately powered (that is to say, the confidence intervals exclude values that would change the clinical decision) and adequately conducted. Particular caution is needed about making negative conclusions based on small trials.

Meta-analysis can also be misleading in the presence of substantial variation (heterogeneity) in study characteristics (leading to the classic criticism of meta-analysis in that it combines apples and oranges), inclusion of individual studies with a high risk of bias and serious publication and/or reporting bias [27, 34]. As a result synthesis teams need to carefully consider the need for correctly undertaking meta-analyses, attempt to minimize these risks when planning the review (e.g. through the use of comprehensive searches to minimize risk of publication bias) and wherever possible explore the risks that a meta-analysis is misleading through the use of sensitivity analysis (e.g. whether observed effects change when studies at high risk of bias are omitted) and other diagnostics (e.g. funnel plots to explore the likelihood of publication bias).

Even if meta-analyses have some drawbacks, they have some advantages as compared to mega-trials. Meta-analyses have more external validity, more efficiency and less susceptibility to funding bias [35]. They have also increased statistical power, narrower confidence intervals for statistical inference and a large sample size and can accommodate testing post hoc hypotheses or explore the effects in selected subgroups. In addition, meta-regression offers the opportunity to test novel and hitherto unprecedented hypotheses, and insights may be gained in these cases by exploring the source of heterogeneity. However, they are more prone to susceptibility to heterogeneity and susceptibility to publication bias and have higher risk of type I error. Another drawback is the small study effect, also called publication bias. Small primary studies are more likely to be reported, quoted or published if their results are significant. And non-significant studies may be easily missed [7]. However, to conduct a mega-trial even if it can have more precision, more internal validity and prospective design, it has higher cost and less external validity, and the funding body may dominate the design and conduct [35].

Although we have commented here qualitative reviews as separate from meta-analyses, qualitative and quantitative methodologies are not always separated. Mixed methods research has the capacity to overcome problems inherent in the independent generation of quantitative or qualitative evidence alone; however, mixed methods studies are usually less likely to describe how the study was conducted, to describe procedures of qualitative data analysis and to be judged credible. Furthermore, as with other individual studies, the strength of evidence rests on the design and context of a particular study. That being said, careful inclusion of such studies into systematic reviews can prove beneficial and strengthen the conclusions.

The mixed methods approach to conducting systematic reviews is a process whereby comprehensive syntheses of two or more types of data (e.g. quantitative and qualitative) are conducted and then aggregated into a final, combined synthesis, or qualitative and quantitative data are combined and synthesized in a single primary synthesis. Mixed methods reviews represent an important development for all individuals involved in evidence-based health-care [18].

The increase in alternative medical treatment options has led to the need for comparative effectiveness research. Randomized controlled trials comparing many treatment options are usually not feasible, so other methodological approaches are needed. In general, systematic reviews focus on direct comparisons of the effects of treatments; however, there is a key limitation of standard meta-analyses; this is so that they can compare only two interventions at a time. This becomes problematic when trying to assess the comparative effectiveness of different treatment options for a health-care problem, and an optimal clinical decision needs to be taken. In addition, it is frequently the case that it is not possible to find clinical trials that compare the different treatments that can be compared directly. Therefore, another methodology is needed, one that permits this comparison in order to help the decision of which treatment to select from the different available alternatives. Network meta-analysis can sometimes solve this problem and help decision-making in these situations.

Network meta-analysis combines both direct and indirect comparisons of treatment effects. It has been developed to assess the relative effectiveness of several interventions and synthesize evidence across a network of randomized trials [13].

Network meta-analysis includes some assumptions such as homogeneity, transitivity and consistency and methods that are evolving. Homogeneity should be assessed for each pair of treatments included in the network meta-analysis for which there is direct evidence. The homogeneity assumption is satisfied when the true treatment effect is the same across all trials that allocate the two treatments of interest. If the true treatment effect is modified by a particular characteristic (e.g. duration of trial) and the trials differ with respect to the characteristic, the homogeneity assumption is violated [36].

Transitivity can be viewed as the extension of clinical and methodological homogeneity to comparisons across groups of studies that compare treatments. Transitivity cannot be tested statistically, but its plausibility can be evaluated conceptually and epidemiologically. The assumption of consistency, that the direct and indirect evidence estimates are in agreement, is a prerequisite to calculate a valid mixed estimate. Consistency is the extension of transitivity across a closed 'loop of evidence' when both direct and indirect evidences are available, and statistical methods can be used to evaluate it [12].

Therefore, network meta-analysis has some clear advantages. One advantage is the use of formal statistical tools to make comparisons across interventions, including those that are not directly compared. Normally they are performed based on individual studies. Some authors think that a network meta-analysis could be conducted combining the results of systematic reviews; in this case, it would be considered as an overview of reviews, but this would be in fact methodologically very

difficult as heterogeneity between systematic reviews is greater [3]. In addition, more data are incorporated in the analysis, and the bigger picture is tackled, while a single pairwise meta-analysis offers very fragmented picture [37].

However, network meta-analysis has also some weaknesses. The use of appropriate methods to evaluate the assumptions underlying network meta-analyses is still limited, moderating the strength of the studies' conclusions [38]. A consensus on terminology and standards for conduct and reporting would be timely [39]. And excluding treatments from a network meta-analysis can significantly change effect estimates and the probability rankings of the best treatment [40] leading to a false sense of a right conclusion.

Furthermore, the evaluation of the network meta-analysis requires careful considerations about the validity of the indirect comparisons as well as other factors that may potentially affect the interpretation of the results. Different tools are available to critically evaluate indirect comparisons and network meta-analysis [41–44] and to help inform health-care decision-making.

## 3.4 From 'Specific' or 'Traditional' Reviews to 'More General' or 'Broad' Reviews

Health-care information is growing in an almost unmanageable way. In the first quarter of 2015, about 35 systematic reviews were published daily,[1] and terms as information overload, or more graphically infobesity or infoxication, acquire a disturbing meaning for people and also for organizations, considering the difficulties in understanding and making decisions caused by the presence of too much information [45].

As has been previously stated, systematic reviews are at the top of the evidence pyramid, representing, at the moment, the highest level of evidence to be considered in health-care decision-making. In the Haynes 6S system to accessing pre-appraised evidence, specific systematic reviews are in the middle of pyramid, and synopsis of synthesis, summaries and systems are located just above it [46]. However, the growing body of evidence synthesis, with different systematic reviews on the same or similar clinical questions, suggests a logical means by which to introduce an intermediate step immediately below systems: the overviews of systematic reviews, systematic reviews of systematic reviews or umbrella reviews.

Considering the large number of systematic reviews and research synthesis available to inform decision-making in health-care, undertaking an overview of systematic reviews to compare and contrast published reviews and to provide an overall examination of a body of information that is available for a given topic may be a useful way to approach evidence-based practice [18]. They allow assessment and consideration of whether reviewers are addressing the same topic independently, observe similar results and arrive at generally the same conclusions. If contradictory, reasons for discrepancy can be explored. The intention is to provide a summary of existing research

[1] EMBASE search on April 22, 2015 (filter: systematic AND ([systematic review]/lim OR [meta analysis]/lim) AND [2015–2015]/py).

(possibly many interventions, many outcomes, different types of evidence), not just resynthesize the results of existing reviews or meta-synthesis or synthesis of meta-analyses [18]. Although being quantitatively less relevant than systematic reviews, the number of overviews of systematic reviews published increased in the last years, from 1 in 2000 to 14 in 2010 [47], maybe due to the Cochrane Collaboration endeavour to promote their dissemination and methodological normalization.

Apart from specific systematic reviews, meta-analysis or network meta-analysis, evidence synthesis can also take on other forms, as described above: umbrella reviews, overviews of reviews or meta-epidemiological studies, each one with their own strengths and limitations. In the following pages, these aspects will be discussed for systematic reviews and for overviews of reviews and meta-epidemiological studies.

### 3.4.1 Comparative Strengths of Different Evidence Synthesis Methods

Specific reviews, mainly systematic reviews, were the first attempt to combine individual information to obtain a comprehensive distillate. In the field of health-care, systematic reviews use a well-known systematic methodology to appraise primary studies (usually clinical trials or observational studies) to answer a specific research question, minimizing bias in results and conclusions [3]. Systematic reviews are a robust (and more and more popular) tool to inform decision-making, mainly due to their well-established methods and feasibility.

Systematic reviews have some advantages over other forms of evidence synthesis [3, 7, 27, 34, 48, 49]. Firstly, each systematic review provides answers to specific questions and interventions, identifying simultaneously gaps in the evidence and priorities for future research and/or synthesis. This ‘minimalist’ approximation to evidence may be useful to busy clinicians, who need at the bedside specific answers to specific questions rather than a complete overview of a topic. Secondly, the users’ perspective is oriented towards the outcomes to be achieved, rather than determining the effectiveness or not of specific interventions. The reduced scope of the research question (in terms of selected patients, few interventions to compare, limited outcomes to appraise) has made systematic reviews more feasible and efficient to carry out. Thirdly, systematic reviews imply a ‘quality control in process’, because individual studies are assessed for risk of bias and reasons are discussed. Finally, the standardized methods for search, selection and appraisal allow the user of systematic reviews to get the least biased information. In addition, statistical software to perform meta-analyses is available and is usually free and easy to use.

Although other forms of evidence synthesis can also be the subject of a quantitative approximation, systematic reviews with meta-analyses give, in a comprehensible and more robust way, the possibility to calculate if the effect of the intervention is sufficiently large in practical as well as statistical terms, allowing the determination of number needed to treat (NNT) or to harm (NNH), both of which have an important role in decision-making and educational purposes for professionals and for the patients.

Overviews of reviews, as stated previously, collate information from different systematic reviews providing a solid basis for decision-makers into one single document ('one-stop shopping'). They can provide a wider outlook on many treatments as compared to systematic reviews, and because of this, they may be more useful in the assessment of health technologies [17], lacking in the capacity to answer specific questions but winning in having a wider scope.

Overviews may be an efficient starting point to help map out the widest range of policy options available [3] and, in a similar way to systematic reviews, can identify gaps in research. In addition, by integrating all existing systematic reviews, it is possible to pay attention to those which are discordant. One crucial point is the timely delivery of relevant reports to decision-makers; this might be ensured by conducting overviews [27].

Overviews of reviews were originally introduced by Cochrane Collaboration in 2009, and this organization continues leading their dissemination and the development of methodological aspects [50]. Despite being in its early stages and despite having a less developed and disseminated methodology than systematic reviews, overviews of reviews can provide information on comparative effectiveness and can highlight different methods used across included systematic reviews [3].

The need for 'fast' evidence in reduced timeframes has also reinforced the attractiveness of undertaking this type of review. The short timeframe contrasts with the process required for a systematic review; therefore, if some have already been carried out, then it is useful to use them.

The term meta-epidemiology was first introduced in 1997 by Naylor [51], but in 2002 Sterne et al. [52] tried to best explain its purpose by referring to it as a 'statistical method' for examining the influence of qualitative problems in randomized controlled trials by means of describing the distribution of research evidence for a specific question, examining heterogeneity and controlling bias across studies and summarizing research evidence [19].

A meta-epidemiological study analyses a series of meta-analyses, in each of which the component studies have been classified according to some study-level characteristic. A first simple analysis may be to calculate a combined odds ratio from odds ratios of the individual meta-analyses in order to appraise, for instance, if non-blinded intervention has biased the results and if so by how much.

Meta-epidemiology has a robust methodological corpus, which is well established and defined, with strong theoretical background. The product of a meta-epidemiological study is easy to understand for people familiarized with systematic reviews and meta-analyses, and many often provide to clinicians and to other decision-makers a consistent way to take evidence-based decisions.

### 3.4.2 Comparative Weaknesses of Different Evidence Synthesis Methods

Advances in evidence-based methods in health care have evolved hugely over the last few years, but this evolution has been linked to a (natural) greater criticism of some aspects of this methodology and its applicability. Temporal perspective sights

of systematic reviews, overviews of reviews and meta-epidemiological synthesis put in context their usefulness and also their limitations and weaknesses. Perhaps these aspects can help them be used properly, but understanding their limitations and weaknesses does not diminish their value.

Specific reviews are by concept narrow in scope [3], because they are intended to answer a specific PICO (patient, intervention, comparison, outcome) framed question. They focus on direct pairwise comparisons, excluding some competing interventions, and may lack formal comparisons across different interventions that could be critical for informed decision-making; part of these problems can be solved with network meta-analysis [3]. On the other hand, even if perfectly executed with perfect data, a specific systematic review that addresses 1 treatment comparison for 1 outcome may offer a short-sighted view of the evidence. This may be sufficient if there is just one alternative and one outcome; however, generally there are many treatments to choose from, many outcomes to consider and many different groups of participants to be included.

Another limitation of meta-analytical techniques is given by the so-called small study effect, as stated above. Small primary studies are more likely to be reported, and, conversely, small non-significant studies are more prone to be unpublished and so missed in systematic reviews; combining these small ones with larger ones may bias the estimated effect [7].

Noble [49] has synthesized other drawbacks and limitations of systematic reviews in a very exhaustive list of 17 items. Of these, the most important could be represented by selection bias, the inclusion of poor quality non-peer-reviewed data, the possible lack of result applicability to individual patients and, mainly, coping with primary studies heterogeneity. It is beyond the scope of this chapter to analyse in depth the problems of heterogeneity in systematic reviews; quantitative statistical approximation by means of $Q$ tests or $I^2$ tests does not reflect the true heterogeneity given by differences in patient characteristics (level of risk, comorbidities, age, etc.), interventions (drugs, dosages) or outcomes.

As an elaborated 'product' of specific reviews, overviews of reviews share some limitations with these. Topics such as selective outcome reporting or publication bias could negatively affect both systematic reviews and overviews of reviews. However, some other weaknesses are privative and, in most cases, are consequences of methodological flaws in decisions, methods or outcomes of original systematic reviews. Biases or inaccuracies in relevant systematic reviews can be carried forwards and possibly exaggerated. Individual systematic reviews can use different effect measures (risk ratios, odds ratios, absolute differences, etc.), different methods of combining estimates (random vs fixed effects models), different approaches to handling missing data or variable assessment of risk of bias, and authors of overviews of reviews sometimes can try to adjust or transform results, but this is not always possible. So overviews have in general greater statistical, clinical and methodological heterogeneity than the originally appraised systematic reviews [3].

It has been said that systematic reviews with or without meta-analysis cannot be better than their included studies allow [5], so quality evaluation of primary studies is a capital fact in this kind of review. The Grading of Recommendations Assessment,

Development and Evaluation (GRADE) Working Group (http://www.grade-workinggroup.org/) has developed a system for rating the quality of evidence in systematic reviews and health technology assessments, included in Cochrane reviews but not yet in a standard way in overviews of reviews [34].

Overviews of reviews are considered appropriate when there are two or more interventions for the same condition or problem presented in separate reviews [34]. Indeed, they require systematic reviews to have been completed for all major interventions, and they also require updated systematic reviews to not be outdated [3].

Mostly, up to now, overviews of reviews summarize the evidence qualitatively making it difficult to decide between treatments. Unlike systematic reviews with meta-analyses, the quantitative synthesis is not included in most overviews of reviews, and there is little guidance on how to choose one systematic review over another, whether to include all systematic reviews and how to handle systematic reviews with discordant findings [3]. In addition, it is not clear how to combine findings from different reviews and how to consider any overlapping of primary studies [27].

However, the main weakness of overviews of reviews concerns logistical aspects. To be truly useful, they require the previous existence of the narrower (and often published) component reviews [5]. On the contrary, overviews of reviews may assume that the function be performed prospectively, defining interventions and outcomes to be addressed by individual systematic reviews [24].

Meta-epidemiological studies can offer some clues about the reliability of treatment effects, changes over time and also the influence of some study characteristics, regardless of the disease [24]. By combining individual meta-analyses in a quantitative manner, meta-epidemiologic reviews allow estimate risk factors across all meta-analyses [53].

An important weakness in meta-epidemiological studies is that primary meta-analyses must include at least one trial with and one without the risk factor of interest, and meta-analysis must include a minimum number of trials, depending on the level of heterogeneity allowed and whether multivariable analyses are undertaken [53].

Other limitations of meta-epidemiology research have been reviewed by Bae [19]: the study results that allow analysis are dichotomous and cannot handle continuous outcomes, statistical power can be limited and, moreover, indirect comparisons cannot be applied (some authors have proposed for this the term network meta-epidemiology).

Transition from a single patient to a study of many patients is a leap of faith in generalizability. A further leap is needed for the transition from a single study to a meta-analysis and from a traditional meta-analysis to a network meta-analysis, let alone wider domains (e.g. umbrella reviews). With this caveat, zooming out towards larger scales of evidence may help us to understand the strengths and limitations of the data guiding the medical care of individual patients [24].

Despite of limitations and weaknesses of tools to appraise and synthesize evidence, systematic reviews and umbrella reviews, including overviews of reviews

and meta-epidemiological studies, continue to be the best tool for an approximation to the truth, considered in terms of evidence-based level. So, criticism of systematic reviews and umbrella reviews must be considered, appraised and then used to improve the quality of reviews, but it is important to remember that at present there are no better alternative methods proposed.

## References


1. Cornell JE, Laine C. The science and art of deduction: complex systematic overviews. Ann Intern Med. 2008;148:786–8.
2. Klassen TP, Jadad AR, Moher D. Guides for reading and interpreting systematic reviews. Arch Pediatr Adolesc Med. 1998;152:700–4.
3. Hartling L, Vandermeer B, Fernandes RM. Systematic reviews, overviews of reviews and comparative effectiveness reviews: a discussion of approaches to knowledge synthesis. Evid Based Child Health. 2014;9:486–94.
4. Pubmed. National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda, MD, 20894 USA.
5. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J. 2009;26:91–108.
6. Grimshaw J. A knowledge synthesis chapter. Canadian Institutes of Health Research. 2010. Available from URL: http://www.cihr-irsc.gc.ca/e/documents/knowledge_synthesis_chapter_e.pdf. Cited 25 Apr 2015.
7. Biondi-Zoccai G, Lotrionte M, Landoni G, et al. The rough guide to systematic reviews and meta-analyses. HSR Proc Intensive Care Cardiovas Anesth. 2011;3:161–73.
8. Whittemore R, Chao A, Jang M, et al. Methods for knowledge synthesis: an overview. Heart Lung. 2014;43:453–61.
9. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from URL: www.cochrane-handbook.org.
10. Murad MH, Montori VM, Ioannidis JPA, et al. How to read a systematic review and meta-analysis and apply the results to patient care users' guides to the medical literature. JAMA. 2014;312:171–9.
11. Noyes J, Popay J, Pearson A, et al. Qualitative research and Cochrane reviews. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from URL: www.cochrane-handbook.org.
12. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Syn Meth. 2012;3:80–97.
13. Cipriani A, Higgins JPT, Geddes JR, et al. Conceptual and technical challenges in network meta-analysis. Ann Intern Med. 2013;159:130–7.
14. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005;331:897–900.
15. Salanti G, Becker L, Cadwell D, et al. Evolution of Cochrane intervention reviews and overviews of reviews to better accommodate comparisons among multiple interventions. Report from a meeting of the Cochrane Comparing Multiple Interventions Methods Groups, Milan: March 2011. Available from URL: http://cmimg.cochrane.org/Milan-report. Cited 25 Apr 2015.
16. Glasziou PP, Chalmers I, Green S, et al. Intervention synthesis: a missing link between a systematic review and practical treatment(s). PLoS Med. 2014;11(8), e1001690. doi:10.1371/journal.pmed.1001690.

17. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ. 2009;181: 488–93.
18. Joanna Briggs Institute Reviewers' Manual: 2014 edition. The Joanna Briggs Institute. ISBN 978-1-920684-11-2. Available from URL: http://joannabriggs.org/assets/docs/sumari/reviewersmanual-2014.pdf. Cited 25 Apr 2015.
19. Bae JM. Meta-epidemiology. Epidemiol Health. 2014;36, e2014019. doi:10.4178/epih/e2014019.
20. Greenhalgh T, Robert G, Macfarlane F, et al. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. Soc Sci Med. 2005;61:417–30.
21. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Social Res Method. 2005;8:19–31.
22. Colquhoum HL, Levac D, O'Brien KK, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol. 2014;67:1291–4.
23. Biondi-Zoccai G, Landoni G, Modena MG. A journey into clinical evidence: from case reports to mixed treatment comparisons. HSR Proc Intensive Care Cardiovas Anesth. 2011;3:93–6.
24. Ioannidis JP. Evolution and translation of research findings: from bench to where? PLoS Clin Trials. 2006;1(7), e36.
25. Ebrahim S, Sohani ZN, Montoya L, et al. Reanalyses of randomized clinical trial data. JAMA. 2014;312:1024–32.
26. Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. J R Soc Med. 2007;100:187–90.
27. Pieper D, Antoine SL, Morfield JC, et al. Methodological approaches in conducting overviews: current state in HTA agencies. Res Syn Meth. 2014;5:187–99.
28. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7(9), e1000326. doi:10.1371/journal.pmed.1000326.
29. Beller EM, Chen JK, Wang UL, et al. Are systematic reviews up-to-date at the time of publication? Syst Rev. 2013;2:36. doi:10.1186/2046-4053-2-36.
30. Knottnerus JA, Turgwell P. Knowledge synthesis to improve practice requires up-to-date definitions, methods, and techniques. J Clin Epidemiol. 2014;67:1289–90.
31. Glasziou PP, Shepperd S, Brassey J. Can we rely on the best trial? A comparison of individual trials and systematic reviews. BMC Med Res Methodol. 2010;10:23. doi:10.1186/1471-2288-10-23.
32. Boeije HR, van Wesel F, Alisic E. Making a difference: towards a method for weighing the evidence in a qualitative synthesis. J Eval Clin Pract. 2011;17:657–63. doi:10.1111/j.1365-2753.2011.01674.x.
33. Nordmann AJ, Kasenda B, Briel M. Meta-analyses: what they can and cannot do. Swiss Med Wkly. 2012;142:w13518.
34. Thomson D, Russell K, Becker L, et al. The evolution of a new publication 867 type: steps and challenged of producing overviews of reviews. Res Syn Meth. 2010;1:198–211.
35. Biondi-Zoccai G, Peruzzi M, Frati G, et al. Which do you like better…a bowl of Cheerios or a Big Mac? Pros and cons of meta-analyses in endovascular research. J Endovasc Ther. 2013;20:145–8.
36. Donegan S, Williamson P, D'Alessandro U, et al. Assessing key assumptions of network meta-analysis: a review of methods. Res Syn Meth. 2013;4:291–323.
37. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It depends on the distribution of effect modifiers. BMC Med. 2013;11:159. doi:10.1186/1741-7015-11-159.
38. Nikolakopoulou A, Chaimani A, Veroniki AA, et al. Characteristics of networks of interventions: a description of a database of 186 published networks. PLoS One. 2014;9(1), e86754. doi:10.1371/journal.pone.0086754.
39. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J Clin Epidemiol. 2014;67:138–43.
40. Mills EJ, Kanters S, Thorlund K, et al. The effects of excluding treatments from network meta-analyses: survey. BMJ. 2013;347:f5195. doi:10.1136/bmj.f5195.
41. Ortega A, Fraga MD, Alegre-del-Rey EJ, et al. A checklist for critical appraisal of indirect comparisons. Int J Clin Pract. 2014;68:1181–9. doi:10.1111/ijcp.12487.

42. Mills EJ, Ioannidis JP, Thourlund K, et al. How to use and article reporting a multiple treatment comparison meta-analysis. JAMA. 2012;308:1246–53.
43. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health. 2011; 14:417–28.
44. Jansen JP, Trikalinos T, Cappelleri JC, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NCP Good Practice Task Force report. Value Health. 2014; 17:157–73.
45. Yang CC, Hsinchun C, Kay H. Visualization of large category map for Internet browsing. Decis Support Syst. 2003;35:89–102.
46. DiCenso A, Bayley L, Haynes RB. Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. Evid Based Nurs. 2009;12:99–101.
47. Hartling L, Chisholm A, Thomson D, et al. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7(11), e49667. doi:10.1371/journal.pone.0049667.
48. Bartolucci AA, Hillegass WB. Overview, strengths, and limitations of systematic reviews and meta-analyses. In: Chiappelli F, Caldeira Brant XM, Neagos N, Oludawara OO, Ramchandani MH, editors. Evidence-based practice: toward optimizing clinical outcomes. Berlin: Springer; 2010.
49. Noble JH. Meta-analysis: methods, strengths, weaknesses, and political uses. J Lab Clin Med. 2006;147:7–20.
50. Smith V, Devane D, Begley CM, et al. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011;11:15.
51. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. BMJ. 1997;315:617.
52. Sterne JA, Jüni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002; 21:1513–24.
53. Trinquart L, Dechartres A, Ravaud P. Meta-epidemiology, meta-meta-epidemiology or network meta-epidemiology? Int J Epidemiol. 2013;42:1131–3.

# Umbrella Reviews, Overviews of Reviews, and Meta-epidemiologic Studies: Similarities and Differences

4

Michail Tsagris and Konstantinos C. Fragkos

**Abstract**

This chapter describes umbrella reviews, overviews of reviews, and meta-epidemiologic studies focusing on their definitions, purposes, and classifications where appropriate and then elaborating on their similarities and differences. We may consider umbrella reviews as reviews integrating several types of study designs but typically randomized controlled trials and systematic reviews of such studies in a unifying fashion in order to address a content issue (e.g., whether or not a given drug is superior to another). Overviews of reviews are reviews of systematic reviews and meta-analyses which can focus on content or methodological issues. Finally, meta-epidemiologic studies focus on potentially different types of study designs but most typically on randomized trials and systematic reviews, usually across different content domains (e.g., topics or conditions), and mainly aim at addressing methodological issues.

*Is there a magic method of determining when a meta-analysis is likely to be misleading?*

C. David Naylor [1]

## 4.1 Introduction

In the era of systematic reviews, we are now witnessing an increasing number of papers that are overviews of reviews. The spectrum is wide – ranging from simply narrative reviews to elaborate statistical theses. The nomenclature around these

M. Tsagris (✉)
Department of Computer Science, University of Crete, Heraklion, Greece
e-mail: mtsagris@yahoo.gr

K.C. Fragkos
Division of Medicine, University College London, London, UK



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_4

papers is still unclear with various names attributed many times to the same process: overviews or reviews, overviews of systematic reviews, systematic reviews of systematic reviews, umbrella reviews, umbrella reviews of systematic reviews, systematic umbrella reviews, treatment networks, multiple treatments meta-analysis, meta-analysis of meta-analyses, and meta-epidemiologic studies are some of the terms used to describe certain types of one study.

In this chapter, we give a description of these concepts and we aim to compare them. But first we will define systematic reviews and discuss the important base for many if not most reviews: the randomized controlled trial.

A systematic review attempts to identify, appraise, and synthesize all the empirical evidence that meets prespecified eligibility criteria to answer a given research question. Researchers conducting systematic reviews use explicit methods aimed at minimizing bias, in order to produce more reliable findings that can be used to inform decision-making [2]. Meta-analysis refers to statistical methods focused on contrasting and combining results from different studies of a systematic review, in the hope of identifying patterns among study results, sources of disagreement among those results, or other interesting relationships that may come to light in the context of multiple studies [3]. However, systematic research synthesis is not free of limitations.

Randomized clinical trials provide the most valid assessment of treatment effects, but bias can occur, most likely in the form of selection bias, performance bias, detection bias, and attrition bias. Empirical evidence about the influence of study design characteristics on trial results comes from meta-epidemiologic studies based on collections of meta-analyses with examples including inadequate allocation concealment and lack of blinding [4].

We will consider overviews of reviews as a more general term, with umbrella reviews being a specific type of overviews of reviews focused on a specific clinical topic (e.g., a drug or a condition). Meta-epidemiologic studies can be seen as overviews of reviews with a nonclinical first topic (e.g., they may focus on finding issues or small study effects). Thus, overviews of reviews may represent the broader environment within which we may focus on specific clinical issues (with umbrella reviews) or more on given methodological aspects (with meta-epidemiologic studies).

In this chapter, we will focus on umbrella reviews, overviews of reviews, and meta-epidemiologic studies. We will cover and explain these three terms aiming to examine their similarities and detect their differences. Subjectivity is difficult to avoid, and thus, we will try to make discourse as undetectable as possible.

## 4.2 Umbrella Reviews

Umbrella reviews are essentially systematic reviews that bring together comparisons of many treatments regarding the same disease [5]. Grant and Booth [6] mention that umbrella reviews specifically refer to a review compiling evidence from multiple reviews into one accessible and usable document. They also mention that an umbrella review focuses on a broad condition or problem for which there are competing interventions and highlights reviews that address these interventions and their results.

For example, Moe et al. [7] presented an umbrella review of 6 reviews about non-pharmacological and nonsurgical interventions for hip osteoarthritis. They found 204 papers but included only 6 in their analysis. Six studies may seem trivial compare to the initial search result of 204 studies, but considering the number of published papers that contributed to each of these reviews and the fact that these reviews are not very common to publish, it becomes clear that this number is not insignificant at all.

The term umbrella reviews is not used very frequently. The Cochrane Collaboration uses this term to describe reviews on the same topic but this doesn't clarify when one can use this term to describe overviews of reviews; and the opposite is also true, reviews which have features of umbrella reviews are not always called umbrella reviews. For example, these overviews focus on many interventions and comparisons. As opposed to a systematic review or meta-analysis, which is limited to one treatment comparison or even one outcome, an umbrella review can be adopted when many treatments are considered.

An umbrella review is probably more useful for health technology assessments which aim to inform guidelines [5]. Another good example is clinical practice [5]: in this case every management option needs to be considered before a decision is made. So we can say that some reviews or meta-analyses are more specific and address a focused range of outcomes. Thus, this explains the name umbrella reviews showing as well their advantage and importance.

Umbrella reviews have obvious drawbacks stemming from insufficiency in quality, available information, studies, and reviews in the literature. The idea of umbrella reviews is rather new and largely unexplored making rather evident that when trying to combine reviews from other studies, with different treatments and or outcomes, heterogeneity increases, the quality of each review, or even meta-analysis, remains an issue, and in general all the limitations of the reviews now add up.

Finally, could someone guarantee that all possible outcomes have been considered in the available reviews? This adds further difficulties when trying to perform a high-quality umbrella review. What if the reviews considered have overlapping studies or published papers that appear in more than one review? Thus, performing an umbrella review becomes an arduous task and now we can clearly understand why the number 6 in Moe et al. [7] is not a small number at all.

## 4.3 Overviews of Reviews

### 4.3.1 Definition and Method of Conduct

> Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up?Bastian et al. [8]

Overviews of reviews are a recent development in research synthesis with a methodology still evolving. The Cochrane Collaboration has led the area with various articles in journals such as *Evidence-Based Child Health: a Cochrane Review Journal* and specific chapters in the *Cochrane Handbook* [9]. They are defined as

reviews that compile information from individual systematic reviews relevant to a single health problem using explicit and systematic methods examining different interventions for the same condition or different outcomes for the same intervention in the same condition or the same intervention for different conditions or populations or finally adverse effects from the same intervention across multiple conditions [2, 10]. In their majority, overviews are narrative or qualitative reviews of their systematic reviews reporting on the findings and summary estimates from the meta-analysis – if occurred [9].

Initially termed as umbrella reviews, they have been subsequently been referred to in the medical literature as meta-reviews, overviews of systematic reviews, reviews of reviews, and systematic review of systematic reviews [11]. They are produced with Review Manager and are usually registered in the Cochrane Library. According to Hartling et al. [10], the key steps in producing an overview of reviews are:

1. Identification of the question
2. Establishment of the author team (clinician, researcher with experience in systematic reviews and/or training in epidemiology or related discipline, statistician, information specialist, representative from the Cochrane Review Groups that produced the underlying reviews, member of the Cochrane Umbrella Reviews Working Group)
3. Identification of outcomes of interest
4. Search for and selection of systematic reviews for inclusion
5. Presentation of findings

Quite pioneering in the field of this type of reviews is the Child Health Cochrane Group which has recently published its approach and methods for conducting overviews [12]. In their case, the managing editor of the journal *Evidence-Based Child Health* identifies topics of potential value and recruits authors from a variety of disciplines to be involved in producing the overview. The overviews are intended to be brief; the finished product focuses on primary outcomes and methods and results of the included systematic reviews.

Recent systematic reviews of overviews of reviews have captured that in the last decade, their publication rate has risen, and 126 overviews of reviews were identified up until 2012 [13]. Interventions examined were pharmacological, non-pharmacological, or both. Most overviews follow a systematic pattern with inclusion criteria which were clearly stated in over 75 % of them, with extensive searches in databases and mentioning the years and keywords searched [11]. However, quality assessment has been an issue with less than two-thirds of the overviews assessing quality of included systematic reviews [11, 13].

The value of overviews is that they can collate evidence from many high-quality systematic reviews and offer a useful reference for decision-makers. Because an overview is based on existing systematic reviews that have already identified the relevant studies and extracted data, carrying out an overview is more feasible and efficient than undertaking a systematic review.

Nevertheless, challenges remain in this new field. Pieper et al. [13] advocate that overviews do not deal with issues of discordance and many methodological challenges in the conduct of overviews remain, such as deciding whether searching for additional primary studies is necessary. They conclude that a need for a consistent nomenclature and the development of reporting guidelines for overviews would help in identifying overviews and improve their quality. There are several limitations to the overview format. One is that the authors are dependent on the decisions and methods used within the relevant systematic reviews. This can involve patient populations chosen, outcomes reported, and timing of search for systematic review. They can be affected by inconsistencies in outcomes and methods across included systematic reviews and have potential for selective outcome reporting, while biases or inaccuracies in relevant systematic reviews can be carried forward and possibly exaggerated. Finally, it requires systematic reviews to have been completed for all major interventions and be up to date or the overview will be incomplete and outdated [10, 11, 14].

### 4.3.2 Examples from the Literature

Usually an overview of review describes the systematic reviews' characteristics initially with the population, intervention and comparison, number of trials and patients analyzed in each review, and outcomes reported. Important aspects are risk assessment and presentation of findings. This includes tables with risk ratios and heterogeneity measures. For instance, Harrold et al. [15] performed an overview of Cochrane reviews to examine the evidence regarding the effectiveness and associated complications of corticosteroids used to prevent bronchopulmonary dysplasia in preterm infants. They included six reviews (67 trials and 6535 patients) and covered three main comparisons: inhaled corticosteroids versus placebo, inhaled versus systemic corticosteroids, and systemic corticosteroids versus placebo (reporting also risk ratio form reviews according to parameters offered from the reviews and also clinical significance). The pattern is more or less similar in all overviews; an indicative table with examples of topics is included below (Table 4.1).

**Table 4.1** Examples of published overviews of reviews

| Authors | Topic |
|---|---|
| Adams and Jones [16] | Respiratory: the dose-response characteristics of inhaled corticosteroids when used to treat asthma |
| Bloom [17] | Medical education: effects of continuing medical education in improving physician clinical care and patient health |
| de Vet et al. [18] | Obesity: environmental correlates of physical activity and dietary behaviors among young people |
| Cates et al. [19] | Child health: safety of regular long-acting beta2-agonists in children with asthma |
| Hillberg et al. [20] | Mental health: review of meta-analyses on the association between child sexual abuse and adult mental health difficulties |

## 4.4 Meta-Epidemiology

### 4.4.1 Definition, Purpose, and Classification

In this section, we discuss meta-epidemiology. Meta-epidemiologic studies have been coined sometimes as meta-analysis, meta-regression, or reviews of reviews. Frankly speaking, meta-epidemiology is more focused as a research paper not being a simple meta-analysis or narrative review we usually encounter in the literature; it is clearly though a sort of meta-review. The topic is quite recent in the literature, and Trinquart et al. [21] recently described a classification which we will follow in the this section. Savović et al. [4] have developed a combined database for meta-epidemiological research which has a comprehensive outline of papers having a meta-epidemiological perspective, dating the first such an endeavor back to 1995 [22]. Meta-epidemiology aims to describe the distribution of research evidence for a specific question, to examine heterogeneity and associated risk factors, and to control bias across studies and summarize research evidence as appropriate [21, 23, 24].

It is nice to start with an example which can provide perspective for the matter under discussion: let us consider the papers by Tzoulaki et al. [25] and Savović et al. [26, 27]. Both papers have taken collections of meta-analyses and have grouped them with respect to certain design characteristics of the original randomized controlled trials and subsequently compared them statistically. Tzoulaki et al. concluded that there were less promising results in the evidence derived from randomized controlled trials than from observational studies [25]; Savović et al. found that intervention effect estimates were exaggerated in trials with inadequate or unclear quality characteristics [26, 27]. An additional element of these studies is that they normally report a ratio of odds ratios which is computed from logistic regression or meta-analysis. Hence, we start to observe the pattern. Statistical methodology for this method has been described in papers by Sterne et al. [28] and Siersma et al. [29].

Therefore, meta-epidemiology, initially termed in an article by Naylor [1], is still nascent with 47 articles in PubMed mentioning the term and quite a few more applying such a method but without specifically mentioning the term. The interest of meta-epidemiology is to control potential biases in previous quantitative systematic reviews and draw appropriate inferences. With this background, diverse methods, such as meta-regression, imputation, informative missing odds ratio, two statistical models, and others, were attempted, and the term meta-epidemiology was introduced [23]. The difference from traditional epidemiology is that the subjects of traditional epidemiologic studies are individuals, whereas those of meta-epidemiologic studies are original articles that published the results of randomized controlled trials performed [24].

Trinquart et al.'s [21] classification differentiates three types of articles: simple meta-epidemiology, meta-meta-epidemiology, and network epidemiology with respect to data sources, restrictions, assessment of trial-level risk factors, assumption regarding direction of bias, impact of risk factors on intervention effect sizes, and assumption regarding exchangeability of the impact of risk factors on intervention estimates. Hence, for a meta-epidemiological article, the point of analysis are meta-analyses of randomized controlled trials; for meta-meta-epidemiology, the point are meta-epidemiologic studies, and for network epidemiology, the point are

meta-analyses of randomized controlled trials published where data had been analyzed with a valid statistical method for indirect comparisons or network meta-analysis (also called multiple-treatment meta-analysis or mixed treatment comparison) [30, 31]. More differences are shown in Table 4.2.

**Table 4.2** Features of meta-epidemiology, meta-meta-epidemiology, and network meta-epidemiology [21]

| | Meta-epidemiology | Meta-meta-epidemiology | Network meta-epidemiology |
|---|---|---|---|
| Data sources | A collection of meta-analyses of randomized trials | A collection of meta-epidemiologic studies, combined into a harmonized dataset without overlap between meta-analyses | Networks of randomized trials |
| Restrictions | Informative meta-analyses must include at least one trial with and without the risk factor of interest | The different meta-epidemiologic studies investigate various sets of risk factors, potentially assessed with different methods | Eligible networks must include more trials than interventions |
| Assessment of trial-level risk factors | Reassessment from individual trial reports or reliance on assessment from each selected meta-analysis | Assessment from each meta-epidemiologic study | Reassessment from individual trial reports or reliance on assessment from each selected network meta-analysis |
| Assumption regarding direction of bias | In active–inactive comparisons, a risk factor is expected not to favor the inactive comparator | | In star-shaped networks, a risk factor is expected not to favor the common comparator |
| | In active–active comparisons, an assumption regarding direction of bias is needed | | In networks with closed loops, an assumption regarding direction of bias is needed |
| Estimation of the impact of risk factors on intervention effect estimates | Effect estimates are compared between trials with and without the risk factor within each meta-analysis; the mean impact of the risk factor is estimated across all meta-analyses | | Effect estimates are compared between trials with and without the risk factor within each network; the mean impact of the risk factor is estimated across all networks |
| Assumption regarding exchangeability of the impact of risk factors on intervention effect estimates | Between trials within meta-analyses | | Between trials within networks |
| | Between meta-analyses | | Between networks |

**Table 4.3** Examples of meta-epidemiologic studies

| Contributing meta-epidemiologic study | Clinical areas/types of interventions | Type of characteristic analyzed |
|---|---|---|
| Als-Nielsen et al. [32], Siersma et al. [29], Balk et al. [33], Contopoulos-Ioannidis et al. [34], Egger et al. [35], Kjaergard et al. [36], McAuley et al. [37], Moja et al. [38], Royle and Milne [39], Sampson et al. [40], Schulz et al. [22], Wood et al. [41], Zhang et al. [42], Nüesch et al. [43, 44], Valdes et al. [45], Oliver et al. [46], Herbison et al. [47], Tzoulaki et al. [25], Savović et al. [26, 27], Dechartres et al. [48, 49], Zhang et al. [50] | Various topics in most papers but certain have a specific focus (circulatory, pediatrics, infection, surgery, mental health, digestion, pregnancy and childbirth, osteoarthritis, cardiovascular, critical care) | Sequence generation, allocation concealment, blinding, placebo control vs. untreated control, genetic polymorphism, exclusion of patients, randomization and effect size, single center vs. multicenter, experimental vs. observational design, study design, sample size |

### 4.4.2 Meta-Epidemiology in the Literature

Examples of studies having employed meta-epidemiology are shown in Table 4.3. Design issues that were analyzed in the papers were sequence generation, allocation concealment, blinding, concealment, placebo control versus untreated control, genetic polymorphism, exclusion of patients, randomization and effect size, study design, and sample size.

## 4.5 Similarities and Differences

The similarities and differences of all three types of studies are shown in Table 4.4. The differences and similarities are described with respect to the unit of analysis, statistics, comparison of interventions, assessment of quality, feasibility, and the targeted audience.

We summarize in a less structured format such similarities and differences highlighting that umbrella reviews can be specifically defined as reviews integrating several types of study designs but typically randomized controlled trials and systematic reviews of such studies in a unifying fashion in order to address a content issue (e.g., whether or not a given drug is superior to another). Overviews of reviews can be identified as reviews of systematic reviews and meta-analyses which focus on content or methodological issues. Finally, meta-epidemiologic studies may be considered to focus on potentially different types of study designs but most typically on randomized trials and systematic reviews, usually across different content domains (e.g., topics or conditions), and mainly aim at addressing methodological issues.

**Table 4.4** Similarities and differences

| | Umbrella reviews | Overviews of reviews | Meta-epidemiology |
|---|---|---|---|
| Unit of analysis | Systematic reviews but also reviews (e.g., critical or systematized) | Systematic reviews | Meta-analysis |
| Statistics | Not imperative to use them but if feasible may involve certain metrics | Its use is investigated early in the design and used when feasible | Primarily statistical analysis (meta-analysis or logistic regression) |
| Comparison of interventions | Primary aim | Primary aim | Examines design effects not interventions |
| Assessment of quality | Not definitely done | Important part of the overview | Not necessarily a part of the design |
| Feasibility | Yes | Yes, however might require statistical expertise | Requires statistical expertise |
| Audience | Decision-makers (e.g., policy makers, guideline committees, practitioners) | Decision-makers and researchers | Researchers |

**Conclusion**

In this chapter, we described three types of reviews based predominantly on systematic reviews and meta-analyses. Similarities and differences were discussed. However, the topic is quite nascent, still evolving and new issues keep arising in this type of research creating new questions and challenges that require tackling.

## References

1. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. Meta-analysis is an important contribution to research and practice but it's not a panacea. BMJ. 1997;315(7109):617–9. doi:10.1136/bmj.315.7109.617.
2. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008. doi:10.1002/9780470712184.
3. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Chichester: Wiley; 2009. doi:10.1002/9780470743386.
4. Savović J, Harris RJ, Wood L, Beynon R, Altman D, Als-Nielsen B, Balk EM, Deeks J, Gluud LL, Gluud C, Ioannidis JPA, Jűni P, Moher D, Pildal J, Schulz KF, Sterne JAC. Development of a combined database for meta-epidemiological research. Res Synth Methods. 2010;1(3–4):212–25. doi:10.1002/jrsm.18.
5. Ioannidis JPA. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. Can Med Assoc J. 2009;181(8):488–93. doi:10.1503/cmaj.081086.

6. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J. 2009;26(2):91–108. doi:10.1111/j.1471-1842.2009.00848.x.
7. Moe RH, Haavardsholm EA, Christie A, Jamtvedt G, Dahm KT, Hagen KB. Effectiveness of nonpharmacological and nonsurgical interventions for hip osteoarthritis: an umbrella review of high-quality systematic reviews. Phys Ther. 2007. doi:10.2522/ptj.20070042.
8. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7(9):e1000326. doi:10.1371/journal.pmed.1000326.
9. Becker LA, Oxman AD. Overviews of reviews. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008. doi:10.1002/9780470712184.ch22.
10. Hartling L, Vandermeer B, Fernandes RM. Systematic reviews, overviews of reviews and comparative effectiveness reviews: a discussion of approaches to knowledge synthesis. Evid Based Child Health Cochrane Rev J. 2014;9(2):486–94. doi:10.1002/ebch.1968.
11. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7(11):e49667. doi:10.1371/journal.pone.0049667.
12. Thomson D, Russell K, Becker L, Klassen T, Hartling L. The evolution of a new publication type: steps and challenges of producing overviews of reviews. Res Synth Methods. 2010;1(3–4):198–211. doi:10.1002/jrsm.30.
13. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. J Clin Epidemiol. 2012;65(12):1267–73. doi:10.1016/j.jclinepi.2012.06.015.
14. Thomson D, Foisy M, Oleszczuk M, Wingert A, Chisholm A, Hartling L. Overview of reviews in child health: evidence synthesis and the knowledge base for a specific population. Evid Based Child Health Cochrane Rev J. 2013;8(1):3–10. doi:10.1002/ebch.1897.
15. Harrold J, Ali S, Oleszczuk M, Lacaze-Masmonteil T, Hartling L. Corticosteroids for the prevention of bronchopulmonary dysplasia in preterm infants: an overview of Cochrane reviews. Evid Based Child Health Cochrane Rev J. 2013;8(6):2063–75. doi:10.1002/ebch.1955.
16. Adams NP, Jones PW. The dose–response characteristics of inhaled corticosteroids when used to treat asthma: an overview of Cochrane systematic reviews. Respir Med. 2006;100(8):1297–306. doi:10.1016/j.rmed.2006.04.015.
17. Bloom BS. Effects of continuing medical education on improving physician clinical care and patient health: a review of systematic reviews. Int J Technol Assess Health Care. 2005;21(3):380–5. doi:10.1017/S026646230505049X.
18. de Vet E, de Ridder DTD, de Wit JBF. Environmental correlates of physical activity and dietary behaviours among young people: a systematic review of reviews. Obes Rev. 2011;12(5):e130–42. doi:10.1111/j.1467-789X.2010.00784.x.
19. Cates CJ, Stovold E, Wieland S, Oleszczuk M, Thomson D, Becker L. The Cochrane Library and safety of regular long-acting beta2-agonists in children with asthma: an overview of reviews. Evid Based Child Health Cochrane Rev J. 2012;7(6):1798–806. doi:10.1002/ebch.1889.
20. Hillberg T, Hamilton-Giachritsis C, Dixon L. Review of meta-analyses on the association between child sexual abuse and adult mental health difficulties: a systematic approach. Trauma Violence Abuse. 2011;12(1):38–49. doi:10.1177/1524838010386812.
21. Trinquart L, Dechartres A, Ravaud P. Commentary: meta-epidemiology, meta-meta-epidemiology or network meta-epidemiology? Int J Epidemiol. 2013;42(4):1131–3. doi:10.1093/ije/dyt137.
22. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408–12. doi:10.1001/jama.1995.03520290060030.
23. Bae J-M. Meta-epidemiology. Epidemiol Health. 2014;36:e2014019. doi:10.4178/epih/e2014019.

24. Zhang W. Meta-epidemiology: building the bridge from research evidence to clinical practice. Osteoarthritis Cartilage. 2010;18(Supplement 2):S1. doi:10.1016/S1063-4584(10)00293-1.
25. Tzoulaki I, Siontis KC, Ioannidis JP. Prognostic effect size of cardiovascular biomarkers in datasets from observational studies versus randomised trials: meta-epidemiology study. BMJ (Clin Res Ed). 2011;343:d6829. doi:10.1136/bmj.d6829.
26. Savović J, Jones H, Altman D, Harris R, Juni P, Pildal J, Als-Nielsen B, Balk E, Gluud C, Gluud L, Ioannidis J, Schulz K, Beynon R, Welton N, Wood L, Moher D, Deeks J, Sterne J. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess (Winchester, England). 2012;16(35):1–82. doi:10.3310/hta16350.
27. Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, Ioannidis JPA, Schulz KF, Beynon R, Welton NJ, Wood L, Moher D, Deeks JJ, Sterne JAC. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012;157(6):429–38. doi:10.7326/0003-4819-157-6-201209180-00537.
28. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002;21(11):1513–24. doi:10.1002/sim.1184.
29. Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med. 2007;26(14):2745–58. doi:10.1002/sim.2752.
30. Chaimani A, Vasiliadis HS, Pandis N, Schmid CH, Welton NJ, Salanti G. Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. Int J Epidemiol. 2013;42(4):1120–31. doi:10.1093/ije/dyt074.
31. Salanti G, Dias S, Welton NJ, Ades AE, Golfinopoulos V, Kyrgiou M, Mauri D, Ioannidis JP. Evaluating novel agent effects in multiple-treatments meta-regression. Stat Med. 2010;29(23):2369–83. doi:10.1002/sim.4001.
32. Als-Nielsen B, Chen W, Lotte Gluud L, Siersma V, Hilden J, Gluud C. Are trial size and reported methodological quality associated with treatment effects? Observational study of 523 randomised trials. In: 12th Cochrane Colloquium: bridging the gaps. Cochrane, Ottawa, Canada 2004. 2–6.
33. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, Lau J. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA. 2002;287(22):2973–82. doi:10.1001/jama.287.22.2973.
34. Contopoulos-Ioannidis DG, Gilbody SM, Trikalinos TA, Churchill R, Wahlbeck K, Ioannidis JPA; Project TE-P. Comparison of large versus smaller randomized trials for mental health-related interventions. Am J Psychiatry. 2005;162(3):578–84. doi:10.1176/appi.ajp.162.3.578.
35. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess (Winchester, England). 2003;7(1):1–76.
36. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001;135(11):982–9. doi:10.7326/0003-4819-135-11-200112040-00010.
37. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet. 2000;356(9237):1228–31. doi:http://dx.doi.org/10.1016/S0140-6736(00)02786-0.
38. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. BMJ. 2005;330(7499):1053. doi:10.1136/bmj.38414.515938.8F.
39. Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. Int J Technol Assess Health Care. 2003;19(04):591–603. doi:10.1017/S0266462303000552.

40. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, St. John PD, Viola R, Raina P. Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol. 2003;56(10):943–55. doi:http://dx.doi.org/10.1016/S0895-4356(03)00110-0.
41. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ (Clinical Res Ed). 2008;336(7644):601–5. doi:10.1136/bmj.39465.451748.AD.
42. Zhang W, Robertson J, Jones AC, Dieppe PA, Doherty M. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis. 2008;67(12):1716–23. doi:10.1136/ard.2008.092015.
43. Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Burgi E, Scherer M, Altman DG, Juni P. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ (Clinical Research Ed). 2009;339:b3244. doi:10.1136/bmj.b3244.
44. Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, Egger M, Juni P. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. BMJ (Clinical Res Ed). 2010;341:c3515. doi:10.1136/bmj.c3515.
45. Valdes AM, Arden NK, Tamm A, Kisand K, Doherty S, Pola E, Cooper C, Muir KR, Kerna I, Hart D, O'Neil F, Zhang W, Spector TD, Maciewicz RA, Doherty M. A meta-analysis of interleukin-6 promoter polymorphisms on risk of hip and knee osteoarthritis. Osteoarthritis Cartilage. 2010;18(5):699–704. doi:http://dx.doi.org/10.1016/j.joca.2009.12.012.
46. Oliver S, Bagnall AM, Thomas J, Shepherd J, Sowden A, White I, Dinnes J, Rees R, Colquitt J, Oliver K, Garrett Z. Randomised controlled trials for policy interventions: a review of reviews and meta-regression. Health Technol Assess (Winchester, England). 2010;14(16):1–165, iii. doi:10.3310/hta14160.
47. Herbison P, Hay-Smith J, Gillespie WJ. Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. J Clin Epidemiol. 2011;64(10):1070–5. doi:10.1016/j.jclinepi.2010.12.018.
48. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ (Clinical Res Ed). 2013;346:f2304. doi:10.1136/bmj.f2304.
49. Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med. 2011;155(1):39–51. doi:10.7326/0003-4819-155-1-201107050-00006.
50. Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. Crit Care (London, England). 2013;17(1):R2. doi:10.1186/cc11919.

# Part II

# Sound Design, Conduct, and Reporting

# Designing the Review

# 5

Spyridon N. Papageorgiou and Giuseppe Biondi-Zoccai

**Abstract**

Overviews of systematic reviews are a relatively new research synthesis method that has emerged, due to the massive publication of systematic reviews and the need to answer complex clinical questions in a timely manner. Although the number of the overviews of systematic reviews has increased, no definitive guidelines regarding their conduct and reporting exist up to now. Some guidance regarding overview procedures can be, however, extrapolated from systematic review methodology, given the similarities that these two methods of evidence synthesis possess. On the other hand, considerable differences between overviews and systematic reviews exist regarding their scope, their eligibility criteria, and their analysis. It is therefore the aim of this chapter to provide a comprehensive guide through the steps of an overview of reviews, which will guide both interested readers and researchers willing to embark on such a journey themselves. This chapter's main emphasis is given to the most widely known type of overview, the umbrella reviews, as the same principles apply to almost all overviews. Finally, specific considerations are given for the other main type of overviews of reviews, namely, meta-epidemiological studies.

S.N. Papageorgiou, DDS (✉)
Department of Orthodontics, School of Dentistry, University of Bonn, Welschnonnenstr. 17, Bonn 53111, Germany

Department of Oral Technology, School of Dentistry, University of Bonn, Bonn, Germany
e-mail: snpapage@gmail.com

G. Biondi-Zoccai, MD, MStat
Department of Medico-Surgical Sciences and Biotechnologies, Sapienza University of Rome, Corso della Repubblica 79, Latina 04100, Italy

Eleonora Lorillard Spencer Cenci Foundation, Rome, Italy
e-mail: snpapage@gmail.com



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_5

## 5.1 Introduction

Akin to the ever-expanding medical literature, systematic reviews in healthcare have increased in numbers, as decision makers try to utilize the best available evidence to inform decision making. Systematic reviews apply a rigorous scientific approach to an existing body of research evidence in an attempt to identify and critically appraise studies and summarize the results of high-quality research ultimately informing in a single manuscript.

In the last years, a number of organizations like the Agency for Healthcare Research and Quality in the USA, the National Institute for Health and Care Excellence in the UK, and international organizations such as the Cochrane and Campbell Collaborations and the Joanna Briggs Institute have been dedicated in producing systematic reviews to inform on evidence-based decision making. As a consequence, the number of published systematic reviews has increased exponentially. Eleven systematic reviews are published every day, while 5000 systematic reviews are indexed annually in MEDLINE [1]. Considering the large numbers of systematic reviews and research syntheses available to inform many topics in health care and the complexity of some clinical questions, systematic reviews of existing systematic reviews (overviews of systematic reviews) have emerged as a means to compare and contrast published reviews and to provide a summary of the body of information that is available for a given topic [2].

Beyond the impetus for overviews driven by the sheer volume of systematic reviews being published, the need for "fast" evidence in reduced time frames has also reinforced the attractiveness of overviews. Decision makers are increasingly required to make evidence informed policy decisions and often require evidence in short time frames [3]. In this sense, overviews of reviews can provide an intermediate link between systematic reviews and rapid reviews, as a streamlined approach to informed decision making from the totality of evidence in health as quickly as possible.

In addition, practice guidelines often require answers to not only one but several linked questions. When systematic overviews address interlocking questions, it is valuable for reviewers to tie the questions together in an analytic framework; this has become the standard practice for full evidence reviews commissioned by the US Preventive Services Task Force [4]. Another example of a complex clinical question is when those responsible for developing clinical guidelines need to focus on data for both the effectiveness and adverse effects of an intervention, outcomes that are sometimes addressed by different systematic reviews. Overviews are valuable in outlining the range of available policy and program options [5] and promoting evidence-based treatment by bringing the evidence to the attention of a new audience [6]. Other valuable aspects of overviews include providing a comprehensive list of the currently available relevant systematic reviews, summarizing the available evidence and its implications, and highlighting areas where more research is needed [7, 8].

The aim of this chapter is to provide a brief summary of the procedures of an overview of systematic reviews that aims to synthesize evidence from multiple systematic reviews. We begin by explaining the terminology that will be adopted throughout the chapter and differentiate between the two main types of overviews,

namely, umbrella reviews and meta-epidemiological studies. As umbrella reviews are found more often in the literature and the two designs share some common methodology, the chapter is focused on the general methodology of overviews. Each step of the overview is briefly described, from setting the overview scope up to the final research synthesis and its presentation, while points that must be taken into consideration before the review's initiation are highlighted.

This chapter is intended as a practical guide for readers wishing to understand the process of an overview of reviews and for readers considering undertaking an overview themselves. Although sometimes overviews might lead to meta-analyses of networks of interventions, this chapter does not pertain to this topic, and readers are prompted to specialized sources on the subject [9]. Finally, this chapter does not pertain to overviews produced by the Cochrane Collaboration (Cochrane Overviews), as these have a very strict scope and default methodology [7], and do not reflect the broad concept of overviews.

## 5.2 Terminology

Terminology in this field remains to a great degree ambiguous, due to the absence of relevant guidelines, as shown by the somewhat different takes of contributors to this book. Starting from the widely known systematic reviews, the purpose of systematic reviews is to summarize all relevant evidence regarding a clinical question, by searching, including, and appraising appropriate clinical studies (hereon "primary studies"), usually randomized controlled trials (RCTs). Overviews of systematic reviews (hereon "overviews"), on the other side, search, include, and appraise systematic reviews rather than primary studies. Overviews can be found described with various terms in the literature, with the most frequent being "overview of reviews," "overview of systematic reviews," and "umbrella reviews" [2]. For practicality reasons, the terminology of the Cochrane Collaboration is adopted and built upon as follows: overview is the broader term, where the aim is to summarize all relevant evidence from multiple systematic reviews. A subset of overviews are umbrella reviews, where the intent is to summarize existing evidence from all systematic reviews on a specific intervention (used for various conditions/health problems) or to summarize evidence from all systematic reviews on a specific condition/health problem (including many interventions). Meta-epidemiological studies are a subset of overviews and aim to summarize evidence from multiple systematic reviews, either on a specific field or not, and to associate quantitatively their results with a specific characteristic of the primary studies [10]. Finally, there are also overviews that do not fall either in the category of umbrella reviews or meta-epidemiological studies; methodological overviews, for example, include multiple systematic reviews and assess their quality, without, however, trying to associate it with their results.

The chapter hereon pertains to overviews, with a predilection for umbrella reviews, except from the last part that list several considerations specific for the conduct of meta-epidemiological studies.

## 5.3 Overview Team/Organizational Issues

The team needed for the conduct of an overview is similar to the team needed for a systematic review and ideally consists of at least three people, as certain procedures (literature search, study selection, data extraction, and assessment of study limitations) should be done by at least two persons independently, and after prior calibration, while disagreements between them are settled by a third uninvolved person. Persons of specific expertise might be needed to complete the overview team including expertise on clinical management of specific diseases or specific treatments, patient acceptance of the administered treatments, and policy issues of an intervention (i.e., clinicians, patients, and policy makers, respectively), expertise on searching electronic databases (i.e., a librarian), translations of papers (in case of searching and including non-English studies), and statistical expertise (i.e., epidemiologist/statistician).

Further organizational issues to take into account include institutional access to electronic databases, author expenses, secretarial support for communication with authors of systematic reviews or authors of primary studies, additional costs for statistical expertise and advanced statistical software, and costs for publication in an open access journal, which is highly recommended for policy-/decision-shaping papers intended for the wide audience, like overviews.

Additionally, as with conventional systematic reviews, it is advisable that duplication of overviews and research waste is avoided at any cost [11, 12]. For this reason, it is advisable that before planning in detail a new overview, its authors search for already existing overviews on the same topic in main databases [MEDLINE, the Cochrane Database of Systematic Reviews, or the Database of Abstracts of Review of Effects (DARE)] or for relevant overviews that are in the production stage [e.g., in the PROSPERO register of systematic reviews website (http://www.crd.york.ac.uk/PROSPERO/)].

In order to increase the transparency of an overview, it is important that a detailed protocol for the overview has been constructed and has been registered in an open depository (e.g., in the PROSPERO register, see following chapter). This protocol should be constructed *a priori* and should include all planned procedures (including qualitative and quantitative synthesis), and all changes to the protocol decided after initiation of the overview should be clearly described and justified. It is therefore advised that the overview authors are familiar to some extent with the actual stand of the literature in the field by conducting a preliminary search of the literature prior to the finalization of the protocol. This will allow them to vaguely anticipate the kind of studies that might be identified, as well as let them decide how to better analyze them. In the case of overviews, a preliminary search will ensure that an up-to-date overview of the research question of interest does not already exist and will let the authors gauge the likely number of systematic reviews (and thus, amount of work) that will be included in the overview.

## 5.4 Scope/Objective

As in all types of research, framing the research question is perhaps the most important foundational step, as it guides the methods and processes of the overview. When planning an overview, it is often helpful to discuss the research objectives widely to ensure that the overview is relevant and addresses the needs of the different potential stakeholder audiences. The objectives of an overview may be broad or focused and may try to answer a clinical question, form a theory based on observation, or instruct the conduct of future studies. In general, the more specific the objectives of the overview, the more acquiescent the question will be to conventional systematic review methodology. Conversely, the broader the objectives of the overview, the more amendable they will be to possible changes/adaptations that might arise during the overview process, while also providing generalizability of the overview findings to various settings and study populations. Some examples of overviews with different scopes are listed in Table 5.1, including conventional systematic reviews, to make the comparison.

To start from familiar grounds, standard systematic reviews usually aim to assess evidence from multiple primary studies regarding the effectiveness or adverse effects of an intervention used on patients of the same condition, ideally through quantitative synthesis (meta-analysis). Going one step further, network meta-analyses or mixed treatment comparison meta-analyses aim to assess evidence from multiple primary studies regarding the effectiveness or adverse effects of multiple similar interventions used on patients of the same or similar conditions.

Per definition, overviews aim to appraise information from multiple systematic reviews on the same or different fields. According to the terminology adopted in this chapter, umbrella reviews are a subset of overviews that synthesize clinical evidence from multiple systematic reviews that answer the same or similar questions, thereby facilitating a quick overview of existing evidence and enabling policy makers to take appropriate decisions. They provide a wide range of possibilities for the overview's scope according to the chosen population/condition (same or different conditions), the chosen interventions to be compared (one or multiple interventions), and the chosen outcome (one or various outcomes pertaining to the intervention effectiveness or adverse effects).

Meta-epidemiological studies aim to synthesize data from multiple systematic reviews, in order to provide empirical evidence about the association between characteristics of primary studies with the results of the primary studies. Briefly, after including eligible systematic reviews, they reorganize the primary studies included in these reviews according to a specific characteristic and calculate the influence of the characteristic on the results within each meta-analysis, before pooling these effects across all included meta-analyses. From such studies, empirical evidence regarding various characteristics of primary studies has been provided, including indexing in MEDLINE [13, 14], language [14–16], or geographic origin [17, 18], study design [14, 19–21], sample size [14, 22–24], or methodological aspects of

**Table 5.1** Various scopes for systematic reviews and overviews of systematic reviews and their corresponding criteria for eligible studies

| Category | Scope | Condition/population | Compared interventions (incl. control/placebo) | Outcome | Study design |
|---|---|---|---|---|---|
| *Systematic reviews* | | | | | |
| Systematic review | Effectiveness/adverse effects of (usually) one intervention for (usually) one condition | Same | Same | Same | Primary studies[a] |
| Network or MTC meta-analysis | Effectiveness/adverse effects of multiple interventions for (usually) one condition | Same | Multiple | Same | Primary studies[a] |
| *Overviews of reviews* | | | | | |
| Umbrella reviews | Effectiveness of an intervention for multiple conditions | Multiple | Same | Same (effectiveness) | SRs[b] |
| | Adverse effects of an intervention for multiple conditions | Multiple | Same | Same (adverse effects) | SRs[b] |
| | Effectiveness of multiple interventions for a specific condition | Same | Multiple | Same (effectiveness) | SRs[b] |
| | Effectiveness of multiple interventions for a specific condition | Same | Multiple | Multiple (effectiveness) | SRs[b] |
| | Adverse effects of multiple interventions for a specific condition | Same | Multiple | Same (adverse effects) | SRs[b] |
| | Adverse effects of multiple interventions for a specific condition | Same | Multiple | Multiple (adverse effects) | SRs[b] |
| Methodological overview of reviews | Methodological issues of SRs, possibly limited to a specific field | Multiple | Multiple | Multiple | SRs |
| Meta-epidemiological study | Association of a characteristic of SRs/primary studies with the results | Same/multiple | Same/multiple | Same/multiple | SRs |

*MTC* mixed treatment comparisons, *SR* systematic reviews

[a]Including screening of reference lists from existing systematic review

[b]Possibly including searching for primary studies to update the included systematic reviews

primary studies [14, 21, 25]. This information can be used for the appraisal of existing evidence, for minimizing bias by appropriately designing a new systematic review or for performing bias adjustment in meta-analyses, through special analytical frameworks [26, 27].

A special kind of overviews is methodological overviews of systematic reviews, where multiple systematic reviews from one or various fields are assessed for various criteria, including, among others, methodological adequacy, compliance with reporting guidelines, or signs of reporting biases [28–35]. This kind of overview aims to provide a cross-sectional description of the current stand in systematic review literature in terms of quality. Umbrella reviews and methodological overviews often overlap to a large degree, as, for example, in the umbrella review of Seida et al. [36]. There is, however, no clear distinction between what is meant as an umbrella review and methodological overviews, except for one: umbrella reviews are focused more on the clinical outcomes reported in the included systematic reviews and assess the review quality with this in mind. Pure methodological overviews on the other side investigate mostly the methodological quality or limitations of systematic reviews in a field, without trying to associate this with their results or answer a clinical question.

Similar to standard systematic reviews, overviews should ideally include a clear and descriptive research question, which will be reflected in the title of the overview and will provide the basis for setting the eligibility criteria for study inclusion.

## 5.5 Title

Given the various terms that are used to describe overviews in the literature, it is important that all overview papers are clearly described in a consistent way in their title, in order to facilitate indexing and quick identification. It is therefore advised that the title of an overview always include the phrase "…: an overview of systematic reviews," "…: an umbrella review," or "…: a meta-epidemiological study," as appropriate. The title of the overview must be concise enough to reflect the interventions or the phenomena of interest as a whole; however, it should also be as descriptive as possible, ideally incorporating as many elements as possible from the PICOS (Participants-Intervention-Comparison-Outcome-Study design) framework.

## 5.6 Background

The background section of an overview paper or protocol should be comprehensive and cover all the main elements of the topic under review. It should cover the knowledge extent addressing the question of the overview and the existence (or lack thereof) of any previous overview with their shortcomings. The reason for undertaking the overview should be clearly stated together with the target audience and whom the overview is intended to inform.

## 5.7 Eligibility Criteria for Inclusion/Exclusion of Studies

The eligibility criteria for inclusion of systematic reviews into the overview should be predefined at the protocol stage, reflect the overview's question, and be constructed with the help of the PICOS framework.

### 5.7.1 Participants

The authors of the overview must define the specific characteristics of the participants (age, sex, ethnicity, condition of interest, severity of the disease, etc.) that might be used to judge the eligibility of the systematic reviews. If the overview authors choose to include multiple patient populations, they must clearly describe each of them. If they choose not to use any limitation regarding the patient populations or their condition, they should likewise state this explicitly. Justification for any inclusion or exclusion of participants should be explained in the overview's text.

### 5.7.2 Interventions/Comparisons

The interventions or phenomena of interest for an overview should be defined in detail and should be congruent with the review objectives. Overviews that aim to address multiple interventions and treatments should define each potential intervention of interest clearly or alternatively state that they aimed to include all existing interventions for a condition. Sometimes, it might make sense to include only a category of interventions in an overview (e.g., "surgical interventions for…" or "pharmacological interventions for…"), and stating this clearly in the title and the eligibility criteria of the overview will enhance the text readability. Another point to consider regarding the interventions that will be compared in the overview is whether or not to include placebo or no treatment as a treatment alternative. Including studies that compare active interventions with placebo or no treatments might strengthen the evidence network [37] and enable better direct and indirect comparisons among treatments.

### 5.7.3 Outcomes

As with standard systematic reviews, outcomes of interest in an overview must be defined in a clear and coherent way. Outcomes should be relevant to the question of the overview, and patient-relevant outcomes should be favored, in general, to surrogate endpoints. If multiple outcomes are considered, it is useful to differentiate between primary and secondary outcomes. Finally, whenever possible and especially in overviews of a single intervention, overview authors should try to include outcomes regarding both effectiveness and adverse effects of an intervention.

### 5.7.4 Study Design

Most overviews limit their search to published systematic reviews [2], as this speeds up the identification and extraction procedure considerably and enables policy making based on the current existing evidence. Eligibility criteria for the selection of appropriate systematic reviews for the overview include usually:

1. The definition of a systematic review that will be used to distinguish between systematic and narrative reviews.
2. The appropriateness of the systematic review and the included primary studies in terms of interventions, condition, populations, and outcomes, according to the overview protocol.
3. The design of the included primary studies, namely, RCTs, observational studies, or both, according to the nature of the research question [21, 38].
4. The provision of complete data at the level of the primary studies, in order to facilitiate replication of the meta-analysis.

However, when the majority of identified systematic reviews are outdated, something that is often seen [39], updates of the systematic reviews by searching for primary studies published after the systematic review's last search are often undertaken. This might be judged appropriate in a wide selection of cases, as the need to keep systematic reviews up-to-date has been clearly documented [40], and in this direction, various surveillance methods have been proposed to identify, if a review has been outdated [41]. There is evidence that approximately 15 %, 25 %, and 50 % of published reviews are out of date 1, 2, and 5.5 years after their publication, respectively [42]. On the other hand, only 5 % of the conducted overviews update their included systematic reviews by additional searches for primary studies [43].

When searches for primary studies are deemed appropriate for up-to-dateness reasons, eligibility criteria for the additional primary studies must be defined, which are usually the same criteria used in the identified systematic reviews. Also, usually the original search strategies of the included systematic reviews are used for their updates, possibly, with slight modifications.

However, *de novo* search for primary studies as a whole is usually not performed in overviews, as this falls out of their scope and increases the workload for the overview team. Nor do overviews seek to replicate the study selection, data extraction or quality assessment already performed by the systematic review authors. This might be appropriate for a new systematic review of multiple treatments that checks previous reviews for eligible studies, but is not what is envisaged for overviews.

## 5.8 Search for Studies to Be Included in the Overview

Similar to systematic reviews, the literature search is an important step in the overview, as it ensures that all existing studies that could contribute with valuable information are identified. The overview protocol should include the detailed search

strategy used for study identification including the resources (databases) checked, the key terms, any search filters or limitations used for each database, and the time frame of each search. It is also advisable that the overview authors provide the exact search algorithm for at least one of the assessed databases, if not for all, in order to enable their replication.

A number of specific validated search strategies have been developed in order to identify systematic reviews in MEDLINE or other databases [44–48]. The overview authors should construct their search in order that a common place among the strategy's sensitivity, specificity, and precision is met, thereby reducing the workload for the study selection and maximizing output. One of the most promising search strategies for identifying systematic reviews is the balanced one provided by Montori et al. [46], which possesses high sensitivity (90.2 %), high specifity (98.4 %), and acceptable precision (46.5 %).

It is generally accepted that searching only MEDLINE is not adequate, since 20 to 70 % of published papers in various fields remain unidentified by MEDLINE [49], while sometimes even studies identified from MEDLINE cannot be retrieved through it [50]. The thoroughness of the conducted literature searches varies considerably among overviews [2]. Overviews produced by the Cochrane Collaboration are as a rule restricted to Cochrane Reviews and to searches in the Cochrane Database of Systematic Reviews [7]. Even when additional searches for non-Cochrane reviews are performed, they are usually restricted to the DARE database [51]. It is, however, advisable that more than one to two databases are searched for an overview, ideally without limiting the search by geographical origin or language of systematic reviews. Literature searches in overviews rarely need to extend prior to 1990, as there are very few systematic reviews published prior to that time [52], and the information loss is typically minimal (unless the interest is on psychological interventions or ancillary subjects). Apart from this start date that can be used to restrict the search, all other restrictions should be kept to a minimum for risk of missing eligible studies. Additionally, gray literature (reports that have not been formally published in a journal) and non-English literature should also be searched and checked for eligibility, whenever possible. Finally, if manual searches are undertaken to supplement the electronic searches, they should be followed by listing of the journals and the specific issues that were searched.

The final step of the literature search is to export the search results from each database in an appropriate format for further assessment, including removal of duplicates and construction of the final list that will be used for study selection. Usually, commercial reference management software programs such as EndNote (http://endnote.com/), Reference Manager (http://www.refman.com/), RefWorks (http://www.refworks.com/), DistillerSR (http://distillercer.com/products/distillersr-systematic-review-software/), or the non-commercial Mendeley (http://www.mendeley.com/) and Systematic Review Assistant-Deduplication Module can be used [53], which contain algorithms designed to automatically identify and remove duplicate records. However, the detection of duplicates can be thwarted by inconsistent citation details, missing information or errors in the records, and is usually only partially successful [54], making a manual check at the end unavoidable.

## 5.9 Study Selection

Selection of eligible studies for overviews should be done in a transparent and objective way that minimizes possible sources of error. It is advisable, therefore, that selection procedures are described in detail and that study selection is performed by at least two independent researchers, while a third uninvolved person is consulted in case of disagreements. Unfortunately, this is not always the case, as in a recent assessment of published overviews, only 49 % of them reported details about the study selection. From those that reported the study selection, however, the vast majority (78 %) of the overviews used two independent researchers to screen potentially eligible papers [2].

Usually, as in systematic reviews, the selection procedure is conducted by checking sequentially first the title, then the abstract, and last the full text of each record, until an exclusion reason can be found. If no exclusion can be justified, then the record is deemed eligible for inclusion. From the practical side of the procedure, the more records that are excluded by title or abstract, the less the workload for the overview team. Assessing the full texts of possibly eligible records is, however, necessary for most of the identified papers and can be encumbered by difficulties in acquiring or translating the full texts of the papers. In the end, a list of reviews included and excluded from the overview (with reasons) must be provided in the overview. Finally, the agreement between the two independent assessors that conduct the duplicate procedures of an overview (study selection, data extraction, and quality assessment) can be measured by a kappa statistic or an intraclass correlation coefficient.

## 5.10 Data Collection

### 5.10.1 Data Collection at the Level of the Review, the Primary Studies, and the Authors

Overviews are generally based on the material of the included systematic reviews and do not replicate their data extraction. Therefore, data are usually collected directly from the published review papers and can be additionally augmented by inquiries to the authors of the systematic reviews. The overview authors might also choose to go one step further and extract data directly from the published reports of each primary study or contact their authors for additional data. This is, however, time-consuming and can heavily increase the workload, thereby slowing the production of the overview, which might go against the overview's scope.

Data collection from the included systematic reviews is performed with specific extraction forms and ideally by two independent researchers. These data collection forms should be constructed prior to the overview initiation and should have been discussed and piloted prior to launching into extraction, in order to maximize consistency between assessors and ensure that sufficient details are extracted.

### 5.10.2 Basic Data to Exclude from Each Included Systematic Review

It is up to the overview authors to decide what data will be exctracted from each included systematic review, but commonly collected information includes:

- A unique identifier for the review and publication year
- The dates covered by the literature search including indications of up-to-dateness
- PICOS information for the primary studies included in the review
- Number of included primary studies, number of included participants, and number of included cases (in cases of binary outcomes)
- Sample size and effect estimate from the largest primary study included in the review
- Overall pooled effect estimate with 95 % confidence interval for the main outcome
- Statistical model used for the analysis
- Heterogeneity diagnostics (including $\tau^2$ and/or $I^2$ with 95 % confidence intervals and *P* value for homogeneity testing)
- Various sources of bias (conflict of interest, funding sources, reporting bias, overall quality of evidence, etc.)

## 5.11 Assessment of Methodological Limitations ("Quality")

One of the most important steps in appraising existing evidence is the assessment of existing limitations and sources of bias that could affect the credibility of the review. As we move from primary studies to systematic reviews and then to overviews, the complexity of the research question increases, which is reflected in the complexity of quality assessments. Quality assessments in overviews should ideally be conducted separately for each included outcome and can be either adopted from the published systematic reviews or be directly done by the overview. Going from the narrowest to the widest focus, the assessments that can be included in an overview are:

1. Assessments of methodological limitations of the primary studies included in the systematic review (their risk of bias, i.e., the extent to which their results might be influenced)
2. Assessments of methodological limitations of the systematic reviews
3. Assessments of the overall quality of existing evidence

### 5.11.1 Limitations at the Level of the Primary Studies

The appraisal of primary studies included in the systematic reviews should be done by the authors of the systematic review and should be incorporated in the review's

conclusions [55]. The authors of the overview might then directly extract a risk of bias summary for each outcome and incorporate it in the overview tables. A problem that can be encountered when relying on assessments performed by the review authors is that often different tools are used, making comparisons among reviews challenging [2]. The alternative is the reassessment of all included primary studies by the overview authors with the same tool, which is, however, time-consuming and beyond the scope of most overviews.

The assessment of internal validity (risk of bias) for RCTs can be done with the Cochrane Collaboration risk of bias tool [56], which has been widely accepted and is supported to a certain degree by empirical evidence. For nonrandomized designs, on the other hand, several tools exist to assess the studies' internal validity [57]. Among them, two comprehensive tools that are usually preferred are the Downs-Black checklist [58] and the Newcastle-Ottawa checklist [59], but no empirical evidence currently exists to support them. It must be here noted that the use of scales for the assessment of primary studies included in a systematic review should be avoided based on theoretical and emerging empirical evidence [55], and checklists or component approaches (like in the Cochrane tool) might be more appropriate.

### 5.11.2 Limitations at the Level of the Systematic Reviews

Authors of overviews should identify possible methodological limitations of the identified systematic reviews, as these might influence the review results. These could include limited literature searches, questionable selection, and extraction procedures, failing to take into account the risk of bias of the included primary studies, inadequate statistical synthesis methods, and reporting biases (including publication bias). There exist various tools that can be used to assess the quality of included systematic reviews, with the most widely used ones being the Oxman and Guyatt Overview Quality Assessment Questionnaire (OQAQ) [60], followed by the A Measurement Tool to Assess Systematic Reviews (AMSTAR) tool [61], and the Quality of Reporting of Meta-analysis (QUOROM) statement [2, 28, 62]. Other tools that are used include, among others, Glenny et al. checklist [63], the Critical Appraisal Skills Programme (CASP) tool [64], the Rapid Assessment Protocol [65], and the National Center for the Dissemination of Disability Research (NCDDR) guidelines [66].

Although OQAQ is one of the earliest and most used tools for evaluating the scientific quality of a review article, methods for evaluating systematic reviews have evolved since the instrument introduction, and OQAQ does not address several methodological domains thought to be important nowadays [67]. On the other hand, AMSTAR has been recommended for assessing the methodological quality of systematic reviews among others, from the World Health Organization and by the Canadian Optimal Medication Prescribing and Utilization Service, as it possesses good reliability in clinical settings and has underwent internal and external validation [61, 68, 69]. It must be noted here that the use of reporting guidelines, like

QUOROM or Preferred Items for Systematic Reviews and Meta-Analyses (PRISMA), for the assessment of methodological quality is inappropriate and discouraged by the authors of these guidelines [55, 62]. Additionally, contrary to the quality assessment at the level of primary studies, for which empirical evidence has linked the risk of bias with the magnitude and direction of effect estimates [25], such a linkage has not yet been established for the methodological quality of systematic reviews. It must also be noted here that the abovementioned tools pertain more to the systematic review methodology (qualitative part) of reviews, rather than the meta-analytical part [70, 71].

Finally, the methodological quality of systematic reviews can also be used as eligibility criterion for the overview, by including only systematic reviews deemed to be of high quality. Although this is possible, most overviews include all systematic reviews and incorporate their methodological quality in the formulation of the conclusions or conduct sensitivity analyses by limiting to high-quality reviews. Indeed, analysis of the potential association between review validity and results can provide important insights on the topic at hand and also provide results with larger implications.

### 5.11.3 Quality of the Overall Body of Evidence

Quality assessment for the overall body of evidence has been introduced in the last years with the GRADE (Grades of Recommendation, Assessment, Development and Evaluation) approach [72], has quickly gained widespread acceptance, and is nowadays incorporated into most systematic reviews. Ideally, the GRADE assessment should be conducted by the authors of the systematic reviews, as they are more likely to be familiar with the study-level details that are needed for assessing the risk of bias, consistency, precision, directness, and reporting bias in each review case [2]. If, however, the GRADE assessment is missing from the original review, the overview authors might decide to conduct it themselves, if the overview team has this capacity. The GRADE approach, although very useful and widely accepted as a means of rating the body of evidence originating from single primary studies or systematic reviews, is not directly transferable to overviews of systematic review. A modification of this approach to fit the needs of overviews would be highly desirable and should be easy for people already familiar with the GRADE principles, but is not yet available.

## 5.12 Research Synthesis

Research synthesis is the ultimate aim of the overview and can be either qualitative or quantitative. The extent to which the overview authors reanalyze some or all of the data included in the systematic reviews varies considerably among overviews and reflects the overview scope, as set by its authors.

### 5.12.1 Qualitative (Narrative) Synthesis of the Included Reviews

Many overviews will simply extract data from the included systematic reviews and reformat them in tables or figures in an abstracted format [7]. Authors of overviews should clearly state, if the results of the reviews have been modified in any way, including converting summary statistics (e.g., odds ratios to relative risks) or standardizing for different baseline risks in order to facilitate comparisons across reviews. If no meta-analyses are conducted in the included reviews, the overview authors can report the concluding remarks of the review authors and their reasons for not conducting meta-analyses.

### 5.12.2 Semiquantitative Synthesis of the Included Reviews

Ideally, each included review should be presented in enough detail to facilitate a quick and effective overview of the overall evidence landscape formed by the included reviews. This often goes further than just reporting the results of each meta-analysis and also includes some minor reanalysis of the meta-analyses, but without altering their scope or analysis concept, so that each piece of the puzzle fits together. For example, if both meta-analyses of continuous and binary outcomes are included in the overview, the overview authors might re-express all effect sizes with the same metric (e.g., odds ratios or standardized mean difference) [73]. Also, while some overview authors prefer to just report the analysis model used in every meta-analysis [74, 75], as this choice is based on clinical and statistical reasoning [76], other overview authors prefer to be consistent and re-express each meta-analysis both with fixed- and random-effects models [77, 78].

Additionally, a number of further assessments or sensitivity analyses are conducted by some authors of overviews, which, however, are no formal meta-analysis of the various review results in the sense of a network meta-analysis. For example, some researchers question whether it is better to draw conclusions from a single well-powered study, rather than by pooling many smaller studies [79–81]. This debate entails many factors that come into play like internal validity, external validity, efficiency, publication bias, the role of funding, etc. In this sense, some overview authors perform a sensitivity analysis, whereby the overall pooled estimate of the meta-analysis is compared with the estimate from the largest included primary study (judged by either the largest sample size or by the lowest standard error of the estimate) [77].

As far as diagnostics of small-study effects are concerned, most overview authors decide to use the Egger regression asymmetry test (with or without extrapolation of the review's results to an infinite sample size) [78, 82], instead of alternative tests that outperform the Egger test [83, 84], as $2 \times 2$ contingency tables that are needed for the latter are often missing. Some overviews also conduct another investigation for publication bias, which includes investigating if the observed number of studies with nominally significant results is different from the expected number of

significant results in each meta-analysis [85], by assuming that the true effect in each meta-analysis coincides with the effect of the most precise study [86].

Finally, in an attempt to reduce the number of false-positive findings and identify reviews with credible findings, specific criteria to categorize the observed associations as credible or not have been proposed, including:

1. Strong statistical significance ($P < 0.001$)
2. Inclusion of more than 1000 cases (for binary outcomes)
3. Absence of extreme heterogeneity ($I^2 < 75$ %)
4. 95 % prediction intervals excluding the null value
5. Significant effects after elimination of small-study effect (by extrapolating to an infinite sample size)
6. Absence of excess significance [75, 77, 78]

### 5.12.3 Dealing with Overlapping and Discordant Reviews

Many times there is some overlap between the identified systematic reviews, which can be either at the level primary study level (one primary study included in two or more reviews) or at the review level (two or more identified reviews on the same topic). It is important that overview authors clearly state, if these overlaps were managed and how, as well as state how discordances between reviews were settled.

Reasons for discordant reviews include different search strategies, methods, populations, study designs, and outcomes. Furthermore, different interpretations of the same data can lead to conflicting conclusions because of different judgments [87]. When, however, discordant results among the identified reviews on a clinical question exist, the overview authors should try to explore subtle differences that may explain discrepancies (such as in questions, methods for study selection, data extraction, data synthesis, or funding source) [43] or might conduct a de novo review, if discrepancies in findings are not apparent [88]. Jadad et al. provide an algorithm that can help the authors of overviews to resolve issues of discordance [89]. Another alternative for the elimination of overlaps at the review level is by selecting only one review per research question – usually the most recent or the most comprehensive one. It is advisable, however, that in such cases overview authors conduct a sensitivity analysis to compare the results of the selected review with the results of the excluded reviews to test the robustness of their choice.

### 5.12.4 Quantitative Synthesis of Review Results

In some cases, overviews may include formal statistical analyses including direct or indirect comparisons based on the data of the included reviews (see corresponding chapter in this book). Analytic frameworks such as mixed treatment comparisons or network meta-analysis can provide valuable information by

using all existing interventions, even if some of them were not originally compared head to head. By including both direct and indirect evidence, mixed treatment comparisons can strengthen inferences about the relative efficacy of interventions as well as inferences about the relative value across a range of interventions [90]. Indirect comparisons might be judged appropriate, especially if there is no evidence on direct comparisons [91]. However, guidance for the use of such methods in the context of overviews is missing [2]. Although these analyses provide valuable information, they might contradict the original aim of the overview, as additional data extraction on the level of the primary studies might be needed. Additionally, contrary to a new systematic review, which would be specifically designed for network or mixed treatment comparisons, meta-analysis in the context of overviews includes the risk of unidentified primary studies having escaped the searches of the original systematic reviews. For these reasons, caution might be warranted when interpreting formal network analyses from overviews, as they might not be exhaustive.

## 5.13 Presentation of Findings

Detailed and transparent presentation of the overview results is of paramount importance, as the aim of overviews is to provide a friendly front end to multiple systematic reviews and an exhaustive list of them. The corresponding chapter for overviews from the Cochrane Handbook provides some basic guidance on presenting the results of overviews, although the choice of how to better present the results is left to the authors of each overview.

As far as tables are concerned, the included review details can be accommodated in a single "characteristics of included reviews" table (in accordance to the "characteristics of included trials" table of systematic reviews). This table can include all information that was chosen to be extracted in the overview's protocol as well as any additional information judged appropriate for extraction during the overview. Additionally, the use of an "overview of reviews" table is suggested by the Cochrane Collaboration, in accordance with the "summary of findings" table for systematic reviews, and an appropriate template is provided.

As far as figures are concerned, overview authors might want to limit the number of included figures to those really necessary and avoid replicating the figures of all included reviews. The original flow diagram proposed by the PRISMA statement for systematic reviews can efficiently be used to guide readers through the identification and selection procedures of an overview paper [55]. Furthermore, if a forest plot is included in an overview, it might be advisable to use slight modifications compared to forest plots of systematic reviews, in order to make the difference between reviews and overviews clear. Overview authors could use each row in the forest plot to report the summary pooled estimate from each included review (instead of the estimate of each primary study, as in systematic reviews), while differentiating between direct/indirect and originally reported/calculated-for-the-overview analyses.

## 5.14 Limitations

Despite their usefulness, overviews might still present certain caveats, which must be taken into account while designing the overview, in order to minimize them. First of all, likewise to normal systematic reviews, overviews too are dependent on the quality of the included systematic reviews. Therefore, included reviews should be both of high quality and up-to-date. The latter is exceptionally important, as new studies might completely change the existing conclusions. The same holds true for overviews of reviews, which must be regularly updated [39]. Additionally, in order to keep the overview up-to-date until the time of their publication, the overview procedures must be kept as short as possible. Finally, overviews of reviews provide a zoom-out picture of the current landscape of evidence [8]. Readers interested in the fine details and implication of the included research will always have to consult the primary trials directly [92].

## 5.15 Considerations for Meta-epidemiological Studies

Although both umbrella reviews and meta-epidemiological studies belong to the big family of overviews, they have considerable differences regarding their planning, conduct, analysis, and implications.

As far as the study scope is concerned, umbrella reviews serve as a rapidly produced friendly front end to the evidence from multiple systematic reviews on a subject, with the aim to inform evidence-based policy making. Meta-epidemiological studies, on the other side, aim to identify phenomena and underlying theories that might influence the results of primary studies as a whole. Although both reviews depend on systematic reviews, the intake for umbrella reviews, as a general rule, is the already "digested" information already extracted, assessed, and presented by the systematic review authors. On the other hand, meta-epidemiological studies embark on direct reassessment and categorization of the primary studies included in each systematic review according to the characteristic of interest.

From the logistic side of the study conducted, the burden for the overview team is considerably greater for meta-epidemiological studies compared to umbrella reviews. A quick search in MEDLINE with the keywords ("meta-epidemiological" or "meta-epidemiologic"), performed on 01/04/2015, gives 36 meta-epidemiological studies, with a mean of 6.4 authors (range 3–18 authors), which is higher than the majority of systematic reviews. Whereas authors of umbrella reviews typically have to acquire the full texts of the systematic review and sometimes contact the authors of the included reviews for clarifications, authors of meta-epidemiological studies must in most cases also acquire the full texts of each primary study included in every review and, possibly, contact their authors for clarifications. As can be imagined, this can multiply the financial and time costs for the production of such research papers. Additionally, whereas only a small minority of systematic reviews is published in non-English languages, there is a bigger part of primary studies that

are in languages other than English, making the management of translations from the overview team often necessary.

From the side of the evidence synthesis, methods used in umbrella reviews can vary from plain narrative synthesis of the results reported from each review to additional minor analyses conducted to assess the result robustness and to complete reanalysis of the review results with mixed treatment comparison frameworks. For meta-epidemiological studies, there are two main alternatives that are used. The first alternative is comparing indirectly the characteristic of interest within each meta-analysis, by creating subgroups of primary studies according to the characteristic of interest. In the example of Bowater et al., the authors compared the results from meta-analyses of cardiovascular trials originating from North America and Europe [18]. The authors calculated separately for each meta-analysis the relative risks for mortality from studies originating from North America and from studies originating from Europe and compared them. The second alternative is to directly express within each meta-analysis the effect of the characteristic of interest and then pool the effects among all meta-analyses with common meta-analytic methods. In the example of Papageorgiou et al. the results of meta-analyses from orthodontic interventions are categorized according to the basic study design of the included primary studies [21]. After acquiring the identified meta-analyses and the subsequent primary studies, the primary studies are characterized according to the study design (e.g., randomized or nonrandomized), and the effect of randomization on the trial results is calculated by means of ΔSMD (difference in standardized means difference between the randomized and nonrandomized subgroups) within each meta-analysis. Then the ΔSMDs from the included meta-analyses are pooled in a meta-analytical framework, which allows for heterogeneity both between trials and between meta-analyses. This has the advantages that a summary effect among all studies can be calculated and that further statistical analyses are possible, including assessments of consistency, reporting biases, subgroup differences (e.g., between reviews of effectiveness or adverse effects), and sensitivity analyses.

Finally, the implications from these two types of overviews can differ considerably. The results of umbrella reviews are usually content- specific and pertain to a specific clinical question (e.g., "which treatment is better for this condition?" or "for which conditions can this treatment be effectively used?"). In this sense, umbrella reviews are invaluable for policy makers, but it might not be possible to extrapolate their results in different clinical scenarios or fields. On the other hand, the implications of meta-epidemiological studies are more on a general methodological basis. It is therefore possible that the results of meta-epidemiological studies can be extrapolated to various fields, either because the meta-epidemiological study covers various fields or because the effect of the characteristic in question is consistent among fields. However, answering a specific clinical question falls outside the scope of a meta-epidemiological study.

For all these reasons, umbrella reviews and meta-epidemiological studies are both overviews, in the sense that they synthesize data from multiple systematic reviews, but differ on a number of levels.

**Conclusions**

To summarize, designing an overview of reviews is very similar to designing a systematic review, and many steps in the procedure can be directly extrapolated or with slight modifications [2]. The aim of overviews, and especially umbrella reviews, is to provide a user-friendly summary of existing systematic review evidence that reflects the actual literature stand, while weighing their pros and cons. For this reason, their production should be swift, but still fortified against bias and exact replication of the systematic review procedures should be avoided, in general. Given the large volume of published systematic reviews, the complexity of some clinical questions, and the need for appraisal of the totality of evidence, overviews can be a useful resource for policy makers in developing clinical practice guidelines, decision support systems, and drug formularies.

## References

1. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7:e1000326.
2. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7:e49667.
3. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher M. Evidence summaries: the evolution of a rapid review approach. Syst Rev. 2012;1:10.
4. Guirguis-Blake J, Calonge N, Miller T, Siu A, Teutsch S, Whitlock E. Current processes of the U.S. Preventive Services Task Force: refining evidence-based recommendation development. Ann Intern Med. 2007;147:117–22.
5. Lavis JN. How can we support the use of systematic reviews in policymaking? PLoS Med Public Libr Sci. 2009;6:e1000141.
6. Bambra C, Gibson M, Sowden AJ, Wright K, Whitehead M, Petticrew M. Working for health? Evidence from systematic reviews on the effects on health and health inequalities of organisational changes to the psychosocial work environment. Prev Med. 2009;48:454–61.
7. Becker LA, Oxman AD. Chapter 22. Overview of reviews. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available from www.cochrane-handbook.org.
8. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. Can Med Assoc J. 2009;181:488–93.
9. Biondi-Zoccai G, editor. Network meta-analysis: evidence synthesis with mixed treatment comparison. Hauppauge: Nova; 2014.
10. Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002;21:1513–24.
11. Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. Syst Rev. 2012;1:7.
12. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, Howells DW, Ioannidis JP, Oliver S. How to increase value and reduce waste when research priorities are set. Lancet. 2014;383:156–65.
13. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess. 2003;7:1–76.

14. Papageorgiou SN, Antonoglou G, Tsiranidou E, Jepsen S, Jäger A. Bias and small-study effects influence treatment effect estimates: a meta-epidemiological study in oral medicine. J Clin Epidemiol. 2014;67:984–92.
15. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol. 2000;53:964–72.
16. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol. 2002;31:115–23.
17. Panagiotou OA, Contopoulos-Ioannidis DG, Ioannidis JP. Comparative effect sizes in randomised trials from less developed and more developed countries: meta-epidemiological assessment. BMJ. 2013;346:f707.
18. Bowater RJ, Hartley LC, Lilford RJ. Are cardiovascular trial results systematically different between North America and Europe? A study based on intra-meta-analysis comparisons. Arch Cardiovasc Dis. 2015;108:23–38.
19. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med. 2011;8:e1001026.
20. Tzoulaki I, Siontis KC, Ioannidis JP. Prognostic effect size of cardiovascular biomarkers in datasets from observational studies versus randomised trials: meta-epidemiology study. BMJ. 2011;343:d6829.
21. Papageorgiou SN, Xavier GM, Cobourne MT. Basic study design influences the results of orthodontic clinical investigations. J Clin Epidemiol. 2015. (Epub ahead of print). doi:10.1016/j.jclinepi.2015.03.008.
22. Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, Egger M, Jüni P. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. BMJ. 2010;341:c3515.
23. Chaimani A, Vasiliadis HS, Pandis N, Schmid CH, Welton NJ, Salanti G. Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. Int J Epidemiol. 2013;42:1120–31.
24. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ. 2013;346:f2304.
25. Savović J, Jones H, Altman D, Harris R, Jűni P, Pildal J, Als-Nielsen B, Balk E, Gluud C, Gluud L, Ioannidis J, Schulz K, Beynon R, Welton N, Wood L, Moher D, Deeks J, Sterne J. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess. 2012;16:1–82.
26. Turner R, Spiegelhalter D, Smith G, Thompson S. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc. 2009;172:21–47.
27. Welton N, Ades A, Carlin J, Altman D, Sterne J. Models for potentially biased evidence in meta-analysis using empirically based priors. J R Stat Soc Ser A. 2009;172:119e36.
28. Biondi-Zoccai GG, Lotrionte M, Abbate A, Testa L, Remigi E, Burzotta F, Valgimigli M, Romagnoli E, Crea F, Agostoni P. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: case study. BMJ. 2006;332:202–9.
29. Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. PLoS One. 2013;8:e83138.
30. Papageorgiou SN, Papadopoulos MA, Athanasiou AE. Evaluation of methodology and quality characteristics of systematic reviews in orthodontics. Orthod Craniofac Res. 2011;14:116–37.
31. Tunis AS, McInnes MD, Hanna R, Esmail K. Association of study quality with completeness of reporting: have completeness of reporting and quality of systematic reviews and meta-analyses in major radiology journals changed since publication of the PRISMA statement? Radiology. 2013;269:413–26.

32. Papageorgiou SN, Papadopoulos MA, Athanasiou AE. Reporting characteristics of meta-analyses in orthodontics: methodological assessment and statistical recommendations. Eur J Orthod. 2014;36:74–85.
33. Papageorgiou SN, Papadopoulos MA, Athanasiou AE. Assessing small study effects and publication bias in orthodontics: a meta-epidemiological study. Clin Oral Investig. 2014;18:1031–44.
34. Onishi A, Furukawa TA. Publication bias is underreported in systematic reviews published in high-impact-factor journals: metaepidemiologic study. J Clin Epidemiol. 2014;67:1320–6.
35. Papageorgiou SN, Dimitraki D, Coolidge T, Kotsanos N. Publication bias & small-study effects in pediatric dentistry meta-analyses. J Evid Based Dent Pract. 2015;15:8–24.
36. Seida JK, Ospina MB, Karkhaneh M, Hartling L, Smith V, Clark B. Systematic reviews of psychosocial interventions for autism: an umbrella review. Dev Med Child Neurol. 2009;51:95–104.
37. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med. 2013;159:130–7.
38. Schünemann HJ, Tugwell P, Reeves BC, Ekl EA, Santesso N, Spencer FA, Shea B, Wells G, Helfand M. Non-randomized studies as a source of complementary, sequential, or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. Res Synth Methods. 2013;4:49–62.
39. Pieper D, Antoine SL, Neugebauer EA, Eikermann M. Up-to-dateness of reviews is often neglected in overviews: a systematic review. J Clin Epidemiol. 2014;67:1302–8.
40. Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. When and how to update systematic reviews. Cochrane Database Syst Rev. 2008;(1):MR000023.
41. Pattanittum P, Laopaiboon M, Moher D, Lumbiganon P, Ngamjarus C. A comparison of statistical methods for identifying out-of-date systematic reviews. PLoS One. 2012;7:e48894.
42. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007;147:224–33.
43. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. J Clin Epidemiol. 2012;65:1267–73.
44. Hunt DL, McKibbon KA. Locating and appraising systematic reviews. Ann Intern Med. 1997;126:532–8.
45. Shojania KG, Bero LA. Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy. Eff Clin Pract. 2001;4:157–62.
46. Montori VM, Wilczynski NL, Morgan D, Haynes RB; Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. BMJ. 2005;330:68.
47. Wilczynski NL, Haynes RB. EMBASE search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. J Clin Epidemiol. 2007;60:29–33.
48. Lee E, Dobbins M, Decorby K, McRae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and metaanalyses. BMC Med Res Methodol. 2012;12:51.
49. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ. 1994;309:1286–91.
50. Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. J Clin Epidemiol. 2008;61:357–64.
51. Petticrew M, Song F, Wilson P, Wright K. Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (DARE). Int J Technol Assess Health Care. 1999;15:671–8.
52. Smith V, Devane D, Begley C, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011;211:11–5.
53. Rathbone J, Carter M, Hoffmann T, Glasziou P. Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. Syst Rev. 2015;4:6.

54. Qi X, Yang M, Ren W, Jia J, Wang J, Han G, Fan D. Find duplicates among the PubMed, EMBASE, and Cochrane library databases in systematic review. PLoS One. 2013;8:e71838.
55. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009;62:e1–34.
56. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. BMJ. 2011;343:d5928.
57. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7:iii–x, 1–173.
58. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52:377–84.
59. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed 24 Mar 2014.
60. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44:1271–8.
61. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
62. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF; for the QUOROM Group. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet. 1999;354:1896–900.
63. Glenny AM, Esposito M, Coulthard P, Worthington HV. The assessment of systematic reviews in dentistry. Eur J Oral Sci. 2003;111:85–92.
64. Guyatt G, Rennie D. Users' guides to medical literature: a manual for evidence-based clinical practice. Chicago: American Medical Association; 2002.
65. The Joanna Briggs Institute. Rapid appraisal protocol internet database. Rapid user guide 1.2. Adelaide: The Joanna Briggs Institute; 2006.
66. Task Force on Systematic Review and Guidelines. Guidelines for assessing the quality and applicability of systematic reviews. Austin: National Center for the Dissemination of Disability Research; 2011. Available at: http://www.ncddr.org/guidelines.
67. Shea B, Dube C, Moher D. Assessing the quality of reports of systematic reviews: the QUORUM statement compared to other tools. In: Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing Group; 2001. p. 122–39.
68. Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. Health Res Policy Syst. 2006;4:20.
69. Canadian Coordinating Office of Health Technology Assessment. COMPUS procedure. Evidence-based best practice recommendations. Available: http://www.cadth.ca/media/compus/pdf/COMPUS_%20procedure_e.pdf. Accessed 28 Mar 2015.
70. Higgins JPT, Lane PW, Anagnostelis B, Anzures-Cabrera J, Baker NF, Cappelleri JC, Haughie S, Hollis S, Lewis SC, Moneuse P, Whitehead A. A tool to assess the quality of a meta-analysis. Res Synth Methods. 2013;4:351–66.
71. Papageorgiou SN. Meta-analysis for orthodontists: part II – is all that glitters gold? J Orthod. 2014;41:327–36.
72. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011;64:380–2.

73. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med. 2000;19:3127–31.
74. Contopoulos-Ioannidis DG, Ioannidis JP. Claims for improved survival from systemic corticosteroids in diverse conditions: an umbrella review. Eur J Clin Invest. 2012;42:233–44.
75. Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JP. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ. 2014;348:g2035.
76. Papageorgiou SN. Meta-analysis for orthodontists: part I--how to choose effect measure and statistical model. J Orthod. 2014;41:317–26.
77. Belbasis L, Bellou V, Evangelou E, Ioannidis JP, Tzoulaki I. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. Lancet Neurol. 2015;14:263–73.
78. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JP. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ. 2015;350:g7607.
79. Shrier I, Platt RW, Steele RJ. Mega-trials vs. meta-analysis: precision vs. heterogeneity? Contemp Clin Trials. 2007;28:324–8.
80. Inthout J, Ioannidis JP, Borm GF. Obtaining evidence by a single well-powered trial or several modestly powered trials. Stat Methods Med Res. 2012. [Epub ahead of print].
81. Biondi-Zoccai G, Peruzzi M, Frati G. Commentary: which do you like better…a bowl of Cheerios or a Big Mac? Pros and cons of meta-analyses in endovascular research. J Endovasc Ther. 2013;20:145–8.
82. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629–34.
83. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med. 2006;25:3443–57.
84. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA. 2006;295:676–80.
85. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials. 2007;4:245–53.
86. Ioannidis J. Clarifications on the application and interpretation of the test for excess significance and its extensions. J Math Psychol. 2013;57:184–7.
87. Strech D, Tilburt J. Value judgments in the analysis and synthesis of evidence. J Clin Epidemiol. 2008;61:521–4.
88. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med. 2008;148:776–82.
89. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. CMAJ. 1997;156:1411–6.
90. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004;23:3105–24.
91. Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, Bradburn M, Eastwood AJ. Indirect comparisons of competing interventions. Health Technol Assess. 2005;9:26.
92. Thomson D, Russell K, Becker L, Klassen T, Hartling L. The evolution of a new publication type: steps and challenges of producing overviews of reviews. Res Synth Methods. 2010;1:198–211.

# Registering the Review

# 6

Alison Booth

**Abstract**

Good quality systematic reviews of any study type involve good design and careful planning. To minimise the risk of bias, methods should be pre-specified in a protocol with subsequent deviations and changes from what was planned being recorded and explained in the completed review report. Transparency in conduct and reporting enables those using systematic review findings to judge the quality of a review and assess for themselves the potential impact of any deviation from what was planned initially. This chapter presents the case for systematic review protocol registration and introduces PROSPERO, an open register designed specifically for prospective registration of systematic reviews. Considerations when registering a systematic review of reviews are illustrated with examples from PROSPERO.

## 6.1 Introduction

Systematic reviews of any type of evidence should involve a consistent, transparent and reproducible approach to identifying, evaluating and summarising the evidence on a topic. Protocol registration is a key step to providing transparency.

Conflicts of Interest: Alison Booth has developed and manages PROSPERO.

A. Booth
Centre for Reviews and Dissemination, University of York, Heslington, York YO10 5DD, UK
e-mail: alison.booth@york.ac.uk



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_6

This chapter explains the principles and purpose of protocol registration and presents PROSPERO, a purpose built register. The requirements for prospective registration of reviews of reviews are illustrated with examples from the register. Much of this chapter is based on information provided on the PROSPERO website and is included here by kind permission of the Centre for Reviews and Dissemination (CRD) [9].

## 6.2 Why Register a Protocol

As systematic review methodology developed, concerns grew about the conduct and reporting of clinical trials [13, 17, 31]. Concerns that the biases seen in trials were also appearing in the conduct of reviews were confirmed by investigations into outcome reporting and publication biases in reviews. In 2002 a comparison of outcomes stated in the protocols of 47 Cochrane reviews with those reported in the full publication showed that 43 of the 47 reviews contained the addition or deletion of outcomes [30]. However reporting in the reviews was poor, so it was not clear if the changes were the result of some form of bias or legitimate but undocumented changes made as the review methods were developed. In 2007 Moher et al. found the overall quality of 300 published systematic reviews to be disappointing [24]. Issues identified included missing details of risk of bias assessment of the included studies in about a third of the reviews; and only a quarter reported undertaking any analysis to look for publication bias. Most of the reporting failures were in non-Cochrane reviews: only 11 % of the non-Cochrane reviews examined mentioned having a protocol. The absence of a protocol raises concerns about the rigour of the conduct of the review.

Evidence of selective outcome reporting biases and poor reporting of reviews prompted leaders in the field to compile and publish the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement [20, 23]. One of the items in PRISMA identified access to the protocol and a registration number as desirable. At the time the PRISMA statement was published, access to systematic review protocols was limited to the outputs of individual organisations such as the Cochrane and Campbell Collaborations and the Joanna Briggs Institute [32–34]. There were also limited options for publishing protocols in journals.

In 2010 Kirkham et al. looked at the prevalence of outcome reporting bias in RCTs and the impact of those biases on 288 Cochrane reviews [19]. When the protocol and the published review were compared, 22 % contained discrepancies in at least one outcome measure; 75 % were in the primary outcome. Potential bias from changes being made after seeing the results from individual trials was found in 29 % (8/28) of these reviews. Only 6 % of the 64 reviews with an outcome discrepancy explained the reason for the change in the review. The study also found that outcomes that were promoted from secondary in the protocol to primary in the review were more likely to be significant than if there was no discrepancy (relative risk 1.66 95 % CI (1.10–2.49), $p=0.02$).

A recent Cochrane review examined empirical studies into selective inclusion and/or reporting of outcomes in systematic reviews of RCTs [27]. The review found that 38 % of the systematic reviews included in four studies, added, omitted, upgraded or downgraded at least one outcome between the protocol and final report. However, the association between statistical significance and discrepancies in reporting of outcomes was unclear. The reason for discrepancies was rarely reported in the published review. It was also unclear whether the decision to make these changes was related to how statistically convincing the treatment effect for that outcome was. There was evidence that 32 % of the systematic reviews did not report all of the outcomes in the abstract of the review. Outcomes with more statistically convincing results were more likely to be completely reported in the abstract than other outcomes.

Bias in research can arise from a variety of sources [35]. Even the most rigorous approach to undertaking a systematic review cannot eliminate bias. But it is possible to minimise some of the risks, and transparency in the process allows the reader to assess the remaining potential risk and influence of bias on the findings of the review. Making key details of the protocol publicly available through registration provides such transparency.

Another major driver for prospective registration of systematic review protocols is to help avoid unplanned duplication. Minimising waste is high on research agendas around the world [18]. There are justifiable reasons for repeating or undertaking complementary systematic reviews, but these should be planned and undertaken in the full knowledge of existing and ongoing reviews [21, 22]. A database of ongoing reviews provides reviewers, funders and commissioners with searchable access to details of what is already being addressed and when the results are likely to be available. This helps to avoid unplanned duplication and has the potential to promote collaborations.

## 6.3 Registration Options

Organisations such as the Cochrane and Campbell Collaborations and the Joanna Briggs Institute only publish protocols for their own reviews on their websites. The Cochrane Collaboration includes protocols for overviews of reviews. Major funders of reviews such as the US Agency for Healthcare Research and Quality (AHRQ) and the UK National Institute for Health Research (NIHR) make protocols for research they fund available on their websites. There are also now more opportunities to get protocols published in journals such as the BMC journal Systematic Reviews. However, a register provides a single site to search for ongoing reviews. Some clinical trials registers have accepted registration of systematic review protocols, but they are not a logical place to search for protocols of systematic reviews.

Growing concerns about and evidence of potential bias in the conduct of systematic reviews, and the lack of an open register, prompted CRD to develop a database for the prospective registration of systematic review protocols [8]. Funded by the NIHR, PROSPERO was launched in February 2011, as the first free, open access

international prospective register of systematic review protocols [11]. PROSPERO has the advantage that it was designed for the registration of systematic review protocols. Details of protocols from the major organisations producing reviews such as the Cochrane Collaboration are now included in PROSPERO. The NIHR has mandated registration of all the reviews they fund which meet PROSPERO inclusion criteria. Other major funders such as the Canadian Institute for Health Research (CIHR) also strongly encourage protocol registration. As register content has grown, so has usage: in 2014 PROSPERO usage statistics showed that well over 3.5 million pages were viewed by over 132,000 unique client Internet providers. Internet provider addresses can represent either a single user or a whole organisation (e.g. the National Health Service in England), so we know that these numbers represent a conservative estimate of actual users. PROSPERO therefore offers international exposure.

Another advantage for PROSPERO is that records are permanent, ensuring that even if the findings are never published and/or referenced in the record, contact details are available for users to follow-up enquiries.

## 6.4 Reviews of Reviews in PROSPERO

The minimum data set for registration was agreed through consultation with international experts in systematic reviewing, methodology, commissioning and guideline development in health and social care and journal editors around the world [7].

At launch the focus was on reviews of the effects of interventions; this has now expanded to include any systematic review for which there is a health-related outcome in the broadest sense [4, 6]. Systematic reviews of reviews were not included in PROSPERO to begin with for practical reasons: it was not clear if the initial registration template would be suitable for broader types of knowledge syntheses [5]. A user survey after a year of operation elicited requests for reviews of reviews to be included on the basis that they have similar methodological issues to systematic reviews of single studies, a similar dataset could be expected, and they are equally prone to bias [4]. In September 2012, the PROSPERO Advisory Group agreed it an appropriate time to start accepting registrations of reviews of reviews.

The term ‘reviews of reviews’ is used in the PROSPERO guidance notes, but other descriptors in submissions are accepted, provided the registration details demonstrate a systematic approach to the review. New Cochrane protocols are automatically uploaded from the Cochrane Library so to avoid duplication of records, protocols for Cochrane overviews of reviews should not be registered independently on PROSPERO. At the end of March 2015, there were 40 reviews of reviews registered in PROSPERO; a list of the variations in descriptors used in records is given in Table 6.1. The results of a search of Ovid MEDLINE for all these descriptors are shown in Fig. 6.1, together with the number of reviews of reviews registered in each year since acceptance. The records in PROSPERO can provide a useful learning tool, though it should be remembered that submissions are not peer reviewed, only checked to ensure they meet the inclusion criteria.

**Table 6.1** Descriptors used for reviews of reviews in PROSPERO (March 2015)

| Main descriptor used in title/submission | PROSPERO records |
|---|---|
| Review of reviews | 13 records (used in title by 12) |
| Overview of reviews | 8 records (used in title by 7) |
| Systematic review of systematic reviews | 6 records (used in title by all) |
| Umbrella review | 6 records (used in title by all) |
| Synthesis of reviews | 2 records (used in title by both) |
| Overview of systematic reviews | 2 records (used in title by both) |
| Review of systematic reviews | 2 records (used in title by both) |
| Systematic review of reviews | 1 record (used in title) |

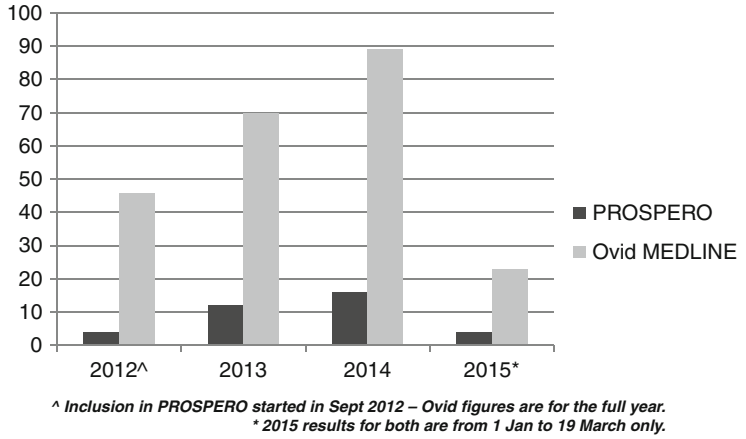



**Fig. 6.1** Reviews of reviews in PROSPERO and Ovid Medline

## 6.5 Registering a Review of Reviews on PROSPERO

Registration on PROSPERO involves the prospective submission and publication of key information about the design and conduct of a systematic review. Registration is free of charge. Registrants are responsible for the information entered in the registration form and by submitting this, agree to be accountable for the accuracy and timeliness of the record and its content. The person submitting the completed form, known as the Named contact, is also expected to keep the record up to date, including provision of a citation and a link to the report on completion of the review. Major publishers such as BioMed Central, BMJ, BMJ Open and PLoS have endorsed prospective registration of systematic reviews and ask for the PROSPERO registration number to be included in submissions of final reports [12].

PROSPERO records are permanent, so can be referred to in the final report to save word space in a journal article. Prospective registration is essential for the comparison of what was planned with what is reported on completion, so retrospective registrations are not accepted. It is not possible to account for all potential

biases, so from a practical perspective, the earliest point in the review process where bias may potentially be accounted for is during screening against eligibility criteria. For this reason, registration forms should be completed and submitted before screening commences. Protocols are iterative documents and during the course of a review amendments may become necessary. PROSPERO facilitates the documentation of revisions and updates on progress, so transparency can be maintained throughout the review. Submissions must be in English for practical reasons, but search strategies and protocols attached to a record may be in any language.

### 6.5.1 Completing Methods Fields for a Review of Reviews

The PROSPERO dataset contains 22 required items and 18 optional items (Table 6.2). The more administrative fields are common to all types of reviews and should be self-explanatory. Details of the information expected in the review methods fields are given here and illustrated with examples from reviews of reviews registered in PROSPERO.

#### 6.5.1.1 Review Question and Title

The title for a review of reviews protocol should state succinctly the interventions or exposures being reviewed and the associated health or social problem being addressed. Including the type of the review in the title helps identification by users and search engines, for example, 'Family-based interventions for substance misuse: a systematic review of systematic reviews' [1]. The objectives of the review, often in the form of questions to be addressed, should be clearly stated. This is ideally done as a statement of the primary aim of the review of reviews, including the interventions and issue to be addressed. Specific objectives may then be listed. For reviews of reviews, this is often quite broad, for example, the objectives from the title example specify in the objectives the inclusion of alcohol and illicit drugs as types of substance misuse [1]. The reviewers then itemise their intention to: review reviews of psychological interventions, assess the effectiveness of specified interventions, identify gaps in the evidence and disseminate findings.

#### 6.5.1.2 Searches

Details of the sources to be searched and any restrictions should be provided. The full search strategy is not required but may be supplied as a link or attachment. Search strategies for reviews of reviews will be simpler than those for systematic reviews of single studies as the included reviews will already have undertaken these searches. However there are still a range of sources that could be appropriately searched depending on the topic. For example, a review of reviews of non-surgical interventions for improving symptoms of obstructive sleep apnoea syndrome in adults includes a broad range of resources to be searched. These include: ASSIA, Campbell Library, DoPHER, CDSR, DARE, EMBASE, NHS EED, PROSPERO, PubMed, PsycINFO, MEDLINE (Ovid), Sports Discus, SCOPUS and CINAHL. As is good practice, they also clearly state in advance the date range for the searches

**Table 6.2** The PROSPERO dataset

| Section | Field number and title |
|---|---|
| Review title and time scale | 1. Review title[a] |
| | 2. Original language title |
| | 3. Anticipated or actual start date[a] |
| | 4. Anticipated completion date[a] |
| | 5. Stage of review at time of this submission[a] |
| Review team details | 6. Named contact[a] |
| | 7. Named contact email[a] |
| | 8. Named contact address |
| | 9. Named contact phone number |
| | 10. Organisational affiliation of the review[a] |
| | 11. Review team members & their organisational affiliations |
| | 12. Funding sources/sponsors[a] |
| | 13. Conflicts of interest[a] |
| | 14. Collaborators |
| Review methods | 15. Review question(s)[a] |
| | 16. Searches[a] |
| | 17. URL to search strategy |
| | 18. Condition or domain being studied[a] |
| | 19. Participants/population[a] |
| | 20. Intervention(s), exposure(s)[a] |
| | 21. Comparators(s)/control[a] |
| | 22. Types of study to be included[a] |
| | 23. Context |
| | 24. Primary outcome(s)[a] |
| | 25. Secondary outcomes[a] |
| | 26. Data extraction, (selection and coding) |
| | 27. Risk of bias (quality) assessment[a] |
| | 28. Strategy for data synthesis[a] |
| | 29. Analysis of subgroups or subsets[a] |
| General information | 30. Type of review |
| | 31. Language |
| | 32. Country |
| | 33. Other registration details |
| | 34. Reference and/or URL for published protocol |
| | 35. Dissemination plans |
| | 36. Keywords |
| | 37. Details of any existing review of the same topic by the same authors |
| | 38. Current review status[a] |
| | 39. Any other information |
| | 40. Details of final report/publication(s) |

[a]Indicates a required field

and the study type as systematic reviews only and that there will be no language restriction [3]. In another example, in addition to searching standard databases, the reviewers describe searching for protocols and registered titles of reviews and state they will not contact authors. They also plan to handsearch relevant guidelines and reference lists [28].

In the future, we expect to see search strategies include repositories of extracted data such as the Systematic Reviews Data Register (SRDR). The SRDR platform facilitates the extraction and management of data for systematic reviews and meta-analyses, creating a central database that can be critiqued, updated and augmented on an ongoing basis. By facilitating open access to extracted data, SRDR helps to provide transparency and reliability in the systematic review process and a more efficient means of producing and updating systematic reviews. SRDR also aims to promote cooperation and utilisation across related resources, including PROSPERO.

#### 6.5.1.3 Condition, PICO and Context

The core PICO items, population, intervention, comparator, outcomes and the condition or domain being studied are all required fields for any systematic review. Reviews of reviews often aim to provide a broad overview of the evidence on a topic, but it is still important to be specific about the scope. A short description of the disease, condition or healthcare domain being studied should include health and wellbeing outcomes where relevant. Summary criteria for the participants or population of interest should include inclusion and exclusion criteria. Full and clear descriptions of the nature of the interventions or exposures of interest and the same for the alternatives against which the interventions should have been compared are important. Again include the inclusion and exclusion criteria for both the interventions and the comparators. In a review of reviews, it may be that a group of interventions is described, for example, all drug interventions will be included, and non-drug interventions will be excluded. Atkinson et al. describe the prevalence and characteristics of obstructive sleep apnoea (OSA); specify participants as adults (18 years and over) diagnosed with OSA; and plan to include any non-surgical intervention (with examples), and any comparator [3]. They put inclusion in the context of current interventions by limiting studies to post 1980 only.

The primary, most important outcome, and secondary, additional outcomes, should be clearly stated and details of the timing and effect measures given as appropriate. Nathan et al. have the primary outcome of 'change in vegetable and/or fruit consumption' [25]. The measures for assessment include change in grams, number of portions at follow-up as measured by a variety of diet records and by changes in biomedical markers. Their secondary outcomes include cost-effectiveness and adverse events. It is essential to give clear and accurate details about outcomes in the registration form as it is this record of what was planned that will be compared with what is reported in the final publication.

#### 6.5.1.4 Types of Reviews

It is also important to identify the types of reviews to be included, for example, by giving details, at least in broad terms, of how the reviews for inclusion and

exclusion will be defined and what types of studies they will include. Atkinson et al. specify they will include reviews and/or meta-analyses of RCTs, observational studies, case-controlled or other quasi-experimental studies but will exclude reviews of diagnostic studies [3].

#### 6.5.1.5 Data Extraction

Give the procedure for selecting reviews for the review, extracting data and coding where relevant. List the data to be extracted, the number of researchers involved and how discrepancies will be resolved. Arango et al. have two reviewers, one from each collaborating centre independently screening all abstracts using an inclusion screening form [2]. The reviewers will be blinded to publisher, journal and authors, making their judgements on the title, year and abstract only. Discrepancies will be discussed by all members of the review team. Full papers for included abstracts will again be reviewed. Similar rigor is described for extraction of data from the included reviews and coding and is specified separately for the two types of reviews to be included. Data extraction lists are then given for each type of review.

#### 6.5.1.6 Quality Assessment

Registration requires a statement of whether and how risk of bias will be assessed, how the quality of individual studies will be assessed and whether and how this will influence the planned synthesis. For the assessment of the methodological quality of systematic reviews, the main tool cited in 28 of the 40 reviews of reviews in PROSPERO is the AMSTAR checklist [29]. Other tools referred to include the Overview Quality Assessment Questionnaire (OQAQ) [26] and the Critical Appraisal Skills Programme (CASP) Systematic review checklist [14]. Akram et al. plan for one reviewer to use AMSTAR and a second to do an independent check and go on to say how they will use the quality assessment scores to decide on which reviews to include in the results [1].

#### 6.5.1.7 Data Synthesis and Subgroup Analyses

Plans for data synthesis may have to change depending on the data identified as meeting the inclusion criteria; however pre-specification of intent is required in PROSPERO. The planned general approach is asked for, and in the case of reviews of reviews, this may be a short description of a structured approach to a narrative synthesis or details of a meta-analysis. Dherani et al. describe how they will map the evidence on interventions to prevent falls, undertake a narrative synthesis, contextualise the data to settings, identify key themes and present quantitative results in tables [15]. By contrast, Wells et al. are looking at reviews of atypical antipsychotics with a single comparator. They therefore can expect to find the data for their planned pairwise and Bayesian network meta-analyses (both fixed and random effects), for which they give analytical software details [36].

Where there are plans to explore subgroups or subsets within the review of reviews, these should also be stated a priori. A registered review of reviews with a systematic review of more recent RCTs and observational studies about non-invasive ventilation includes planned subgroup analyses such as grouping by

type/mode of ventilation, number of hours of use per day, severity of disease and co-morbidities [16].

#### 6.5.1.8 Dissemination

Although this is an optional field, dissemination is an essential part of any research project, and brief details of plans for communicating essential messages from the review to the appropriate audiences should be given. Where appropriate, plans should not be confined to the publication of a report and/or academic journal article. For example, one registered review of reviews plans a workshop for key stakeholders [15], and another will use a variety of media for specific audiences [1].

On completion and publication, the PROSPERO record should be updated to reflect the status of the review of reviews. Details of final publications, of any type, can be added to the PROSPERO record at any time.

## 6.6 Practicalities of Registration

To register any review, including a review of reviews, users 'Join' to obtain a login name and password so they can 'Sign in'. This allows creation of a new record and/or access to existing records for updating or amending. The dataset is divided into a four pages and 'Required' fields, marked with a red asterisk, must be completed before the Submit button becomes active. The form can be saved and exited at any time and revisited at a later date to add or edit information. Personal 'Join' information can be updated and PROSPERO password changed in 'My details'.

Forms can be printed or saved as a portable document (pdf) or as a word processing document to enable sharing and collaboration on development of the submission. Fields can be completed by cutting and pasting information from a prepared protocol. Alternatively, the PROSPERO form has been used as a template for developing the review protocol [6]. Records need to be fully searchable, so information needs to be entered in the specified fields: it is not sufficient to refer to an attached protocol or publication. Brief guidance is given for each field, and further information and examples can be accessed via the icon or downloaded as a pdf [10].

When all the required fields have been completed, the 'Submit' button becomes active and the form can be sent to the PROSPERO administrators. Access to your record is suspended during the administrative process. Receipt of submissions is acknowledged in an automated email sent to the named contact. Application forms are checked against the inclusion criteria for PROSPERO and for clarity of content. They are then approved and published on the register and returned for clarification or rejected. Submissions are turned round within 5 working days of receipt.

Once published on the register, the record becomes accessible again in 'My records'. This allows amendments and updates within the record. On submitting changes, you will be asked to give brief details in a revision note of the changes made. The information entered here will appear in the public record and should inform users of the register of the nature of the changes made (e.g. removed one of the outcome measures after initial screening as not relevant to review question; changed the anticipated completion

date as data extraction is taking longer than anticipated). All submitted edits and changes to a PROSPERO record are recorded, dated and made available within the public record audit trail. The most recent version appears and previous versions are accessible from dated archive links in the record together with the revision notes.

Records remain permanently on PROSPERO. Once the review is completed, the status should be updated in the record and the anticipated publication date given. Once available, the bibliographic reference and electronic links to final publications should be added to the record. In the absence of a publication, details of availability of the review's unpublished results, or reasons for the termination of the review, should be documented. The named contact will receive reminder emails on the anticipated completion and the anticipated publication dates, with detailed instructions on what to do.

When it comes to updating a completed review that has already been registered on PROSPERO, details should be added to the existing record by selecting the 'Update of a review' status option. This ensures that the history and previous versions are all linked and available in the same record and the unique identification number links all records.

**Conclusion**

The value of time spent on designing, planning and then registering a systematic review of reviews should never be underestimated. Good design and a well thought through protocol is the basis of good research. Making detailed planned methods publicly available via PROSPERO provides transparency in the process and helps avoid unplanned duplication. Registration is best practice and demonstrates that sound, reproducible methods have been used, giving confidence in the reliability of the findings.

## References

1. Akram Y, Copello A, Moore D. Family based interventions for substance misuse: a systematic review of systematic reviews. PROSPERO. 2014:CRD42014006834.
2. Arango J, Ellsberg M, Morton M, et al. Interventions to prevent or reduce violence against women and girls: a systematic review of reviews. PROSPERO. 2013:CRD42013004422.
3. Atkinson G, Batterham A, Ells L, et al. Systematic review of reviews on non-surgical interventions for improving symptoms of obstructive sleep apnoea syndrome in adults. PROSPERO. 2013:CRD42013006052.
4. Booth A. PROSPERO's progress and activities 2012/13. Syst Rev. 2013;2:111. doi:10.1186/2046-4053-2-111.
5. Booth A, Clarke M, Dooley G, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev. 2012;1:2. doi:10.1186/2046-4053-1-2.
6. Booth A, Clarke M, Dooley G, et al. PROSPERO at one year: an evaluation of its utility. Syst Rev. 2013;2:4. doi:10.1186/2046-4053-2-4.
7. Booth A, Clarke M, Ghersi D, et al. Establishing a minimum dataset for prospective registration of systematic reviews: an international consultation. PLoS One. 2011;6(11):e27319. doi:10.1371/journal.pone.0027319.

8. Booth A, Clarke M, Ghersi D, et al. An international registry of systematic review protocols. Lancet. 2011;377(9760):108–9. doi:10.1016/S0140-6736(10)60903-8.
9. Centre for Reviews and Dissemination. CRD website. 2015. http://www.crd.york.ac.uk. Accessed 27 Mar 2015.
10. Centre for Reviews and Dissemination. Guidance notes for registering a systematic review protocol with PROSPERO. 2013. http://www.crd.york.ac.uk/PROSPERO/documents/Registering%20a%20review%20on%20PROSPERO%202%20Sept%202013.pdf. Accessed 27 Mar 2015.
11. Centre for Reviews and Dissemination. PROSPERO International prospective register of systematic reviews. 2015. http://www.crd.york.ac.uk/PROSPERO. Accessed 27 Mar 2015.
12. Centre for Reviews and Dissemination. PROSPERO International prospective register of systematic reviews. Support for prospective registration of systematic reviews. 2015. http://www.crd.york.ac.uk/PROSPERO/support.asp. Accessed 28 Mar 2015.
13. Chan A, Hróbjartsson A, Haahr M, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004;291(20):2457–65. doi:10.1001/jama.291.20.2457.
14. Critical Appraisal Skills Programme (CASP). CASP checklists. 2014. http://www.casp-uk.net/#!casp-tools-checklists/c18f8. Accessed 27 Mar 2015.
15. Dherani M, Buckner S, Pope D, et al. Preventing falls and associated mortality in older people: an umbrella review of systematic reviews. PROSPERO. 2015:CRD42015010571.
16. Dretzke J, Moore D, Turner A, et al. Systematic review and economic evaluation of the clinical and cost-effectiveness of community based non-invasive ventilation (NIV) in patients with stable end stage COPD with hypercapnic respiratory failure. PROSPERO. 2012:CRD42012003286.
17. Dwan K, Altman D, Arnaiz J, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008;3(8):e3081. doi:10.1371/journal.pone.0003081.
18. Ioannidis J, Greenland S, Hlatky M, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383(9912)166–75. doi:http://dx.doi.org/10.1016/S0140-6736(13)62227-8.
19. Kirkham J, Altman D, Williamson P. Bias due to changes in specified outcomes during the systematic review process. PLoS One. 2010;5(3):e9810.
20. Liberati A, Altman D, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6(6). doi:10.1371/journal.pmed1000097.
21. Moher D. The problem of duplicate systematic reviews. BMJ. 2013;347. doi:http://dx.doi.org/10.1136/bmj.f5040.
22. Moher D, Booth A, Stewart L. How to reduce unnecessary duplication: use PROSPERO. BJOG. 2014;121(7):784–6. doi:10.1111/1471-0528.12657.
23. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;339:b2535.
24. Moher D, Tetzlaff J, Tricco A, et al. Epidemiology and reporting characteristics of systematic reviews. PLoS Med. 2007;4(3):e78. doi:10.1371/journal.pmed.0040078.
25. Nicole Nathan N, Wolfenden L, Wiggers J, et al. Effectiveness of school-based interventions on increasing children's consumption of vegetables and fruit: a review of reviews. PROSPERO. 2014:CRD42014013082.
26. Oxman A, Guyatt G. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44(11):1271–8.
27. Page M, Mckenzie J, Kirkham J, et al. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. Cochrane Database Syst Rev. 2014;(10):MR000035. doi:10.1002/14651858.MR000035.pub2.
28. Plaszewski M, Bettany-Saltikov J. Effects of nonsurgical management for patients with idiopathic scoliosis: an overview of systematic reviews. PROSPERO. 2013:CRD42013003538.

29. Shea B, Hamel C, Wells G, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009;62(10):1013–20.
30. Silagy C, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. JAMA. 2002;287(21):2831–4. doi:10.1001/jama.287.21.2831.
31. Song F, Eastwood A, Gilbody S, et al. Publication and related biases. Health Technol Assess. 2000;4(10):1–115. doi:10.3310/hta4100.
32. The Campbell Collaboration. The Campbell Library. 2015. http://www.campbellcollaboration.org/lib/. Accessed 28 Mar 2015.
33. The Cochrane Collaboration. The Cochrane Library. 2015. http://www.cochranelibrary.com/. Accessed 28 Mar 2015.
34. The Joanna Briggs Institute. The Joanna Briggs Institute register of titles. 2015. http://joannabriggs.org/research/registered_titles.aspx. Accessed 28 Mar 2015.
35. Tricco A, Pham B, Brehaut J, et al. An international survey indicated that unpublished systematic reviews exist. J Clin Epidemiol. 2009;62(6):617–23.e5. doi:10.1016/j.jclinepi.2008.09.014.
36. Wells G, Cameron C, Klarenbach S, et al. Clinical benefits and harms of atypical antipsychotics for schizophrenia in patients inadequately managed with one or more atypical antipsychotics at recommended doses: a systematic review of reviews and network meta-analysis. PROSPERO. 2013:CRD42013005487.

# Searching Evidence

7

Su Golder and Kath Wright

**Abstract**

The quest for reviews suitable for inclusion in an umbrella review, an overview of reviews or in a meta-epidemiological study is a crucial step to ensure the validity of findings of evidence syntheses as well as to minimise publication bias or similar threats to accuracy. Dedicated databases are obviously a key starting point, such as the Cochrane Database of Systematic Reviews (CDSR) and the Database of Abstracts of Reviews of Effects (DARE). Then, established filters can be used to search the more comprehensive, but potentially too extensive, general literature databases, such as MEDLINE/PubMed and Embase. Careful and early collaboration with an information specialist is recommended to optimise the yield and minimise the effort of searching evidence for an umbrella review.

## 7.1 Introduction

Literature searching is a key component of any review, be it a traditional systematic review or an umbrella review. To be useful to decision-makers, reviews should aim to be as comprehensive as is feasible under the constraints of time, resources and the difficulties of identifying some data (such as unpublished material, industry data, non-English or older studies). In addition, the methods used should be reproducible and transparent to enable the reader to assess the quality of the review and, if need be, to update the review.

S. Golder, PhD (✉)
Department of Health Sciences, University of York, Area 2, York YO10 5DD, UK
e-mail: su.golder@york.ac.uk

K. Wright
Centre for Reviews and Dissemination, University of York, York, UK



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_7

## 7.2 Identifying Studies for Inclusion in Traditional Systematic Reviews versus Identifying Reviews for Inclusion in Umbrella Reviews

Literature searching carried out to identify potential studies to populate traditional systematic reviews usually consists of a range of activities that can include searching electronic databases, checking reference lists of included studies, citation searching, contacting experts and handsearching [1, 2]. Usually there is also some attempt made to identify unpublished materials, such as conference presentations or information only available via regulatory agencies, the Internet or pharmaceutical and manufacturing industries; this can be done by searching databases that focus on this kind of publication, by scanning relevant websites or by contacting authors and industry. An additional option may be to carry out general Internet searches using a search engine such as Google or Google Scholar. The potential for retrieving an unmanageable number of records from Internet searches and the unreliable nature of searching the Internet may mean that the review team agree to scan a predefined number of the Internet hits retrieved and only use this method as a supplement to other forms of searching.

The search strategies used in the searches of electronic databases may be complex, combining both free text terms, subject headings and, if required, search filters designed to restrict search results to specific types of study such as randomised clinical trials, case series or evaluation studies. The aim is to create a strategy with a high degree of sensitivity to ensure that as many potentially relevant records as possible are identified from the searches, with as few irrelevant records as possible.

It is usual practice to search several, possibly overlapping, databases to take account of the differences in journal and topic coverage and also to compensate for poor or inconsistent indexing that can result in a search strategy failing to identify records [1, 2]. Inevitably this approach leads to some records being identified multiple times, but duplicates are usually identified and removed using bibliographic software, such as endnote.

Overall, the effort and time involved in carrying out a robust search to identify studies for inclusion in a traditional systematic review is considerable, and several organisations have published guidance that underpins literature searching practice [1–3]. By contrast, there is comparatively little discussion or guidance currently available on how to search for reviews to include in an umbrella review and which sources to use.

Robinson et al. (2014) [4] seek to develop "detailed methods to explicitly and transparently integrate existing reviews into systematic reviews" (p. 1) and as part of that process have identified existing guidance on locating systematic reviews that recommend "using specific databases and search filters to aid in locating existing systematic reviews" (p. 6). Commonly recommended databases include Database of Abstracts of Reviews of Effects (DARE), Cochrane Database of Systematic Reviews (CDSR), Health Technology Assessment (HTA) database, MEDLINE and Embase. We will describe these, and other potential resources, in this chapter.

In addition to offering a list of databases to search, Robinson also suggests "Some organizations promote limiting searches for existing systematic reviews to selected sources (for example, CDSR), with the idea that these systematic reviews would be expected to meet sufficient quality standards" (p. 6). The effect of doing this would be to have fewer sources to search, a simpler search strategy (as no systematic review filter will be required) and less time spent in identifying reviews to include in an umbrella review.

Searching as comprehensively as possible for all previous systematic reviews on a related topic can be challenging. There are two main types of sources that are useful to identify systematic reviews:

1. Databases specifically aimed at providing large collections of systematic reviews
2. More generic sources and databases which contain a wide array of evidence types – including systematic reviews

The first source does not require the use of a systematic review search filter (combination of search terms aimed at identifying systematic reviews). However, for the second source, a search filter may be useful to focus the search on systematic reviews and thus limit the potentially large number of irrelevant material to sift.

## 7.3 Sources to Use to Identify Reviews for Inclusion in Umbrella Reviews

There are a small number of databases and other resources that comprise collections of systematic reviews that can be used to identify potential studies for inclusion in umbrella reviews. Some of them, such as the Cochrane Database of Systematic Reviews (CDSR), have a search interface that allows the user to develop sophisticated search strategies incorporating free text search terms alongside Medical Subject Headings (MeSH). However, others are primarily web-based resources with a basic search interface more suited to using one or two search terms. What is common to all of the resources described below is that it is adequate to use the topic terms relevant to the question without search terms for systematic review study type. The electronic searches developed, therefore, tend to be less extensive and complex than those conducted for traditional systematic reviews, while the numbers of potentially relevant records are also much smaller.

### 7.3.1 Cochrane Database of Systematic Reviews (CDSR): http://www.thecochranelibrary.com/

The Cochrane Database of Systematic Reviews (CDSR) contains only reviews and review protocols published by the Cochrane Collaboration – an international organisation composed mostly of volunteers. CDSR includes systematic reviews of healthcare interventions, diagnostic test accuracy reviews and methodology reviews.

Their review methods incorporate approaches designed to minimise potential bias such as using several sources to identify studies, predefined inclusion criteria and validity assessment of included studies [2]. A recent study [5] provides evidence that Cochrane Reviews can have an impact upon healthcare policy. The reviews are regularly updated by Cochrane Review groups as new evidence becomes available and covers a wide range of topic areas.

CDSR is one of the resources that make up the Cochrane Library produced by John Wiley & Sons Ltd. It is available on subscription in various formats, although some countries and regions have set up national agreements making it free at the point of use.

Database of Abstracts of Reviews of Effects (DARE): http://www.crd.york.ac.uk/CRDWeb/

Reviews included in the Database of Abstracts of Reviews of Effects (DARE) need to be clearly health related and to have evaluated the outcome(s) of an intervention such as quality of life or mortality. Reviews of adverse effects and risk of disease are included as are reviews of the organisation and delivery of health care, reviews of diagnostic and prognostic tests and reviews of screening programmes.

DARE includes critical abstracts and citations for systematic reviews that have been identified and assessed against a set of quality criteria as below:

- Were inclusion/exclusion criteria reported?
- Was the search adequate?
- Were the included studies synthesised?
- Was the quality of the included studies assessed?
- Are sufficient details about the individual included studies presented?

Systematic reviews must meet four of these criteria, with criteria one to three being mandatory.

DARE is available via a number of different routes. It can be searched via the Centre for Reviews and Dissemination's (CRD) website as well as being incorporated into the Cochrane Library, PubMed Health and Ovid's Evidence-Based Medicine Reviews (EBMR). The search interfaces available on the Cochrane Library and the Centre for Reviews and Dissemination website both allow users to create complex search strategies using free text terms and MeSH, while search results can be exported in formats suitable for importing into bibliographic software.

The database was produced by the Centre for Reviews and Dissemination, University of York, from 1994 to March 2015 and in October 2014 contained over 30,000 records. No new records have been added to the database since March 2015, although it is still available to search from CRD's website.

### 7.3.2 Health Technology Assessment (HTA) Database: http://www.crd.york.ac.uk/CRDWeb/

The Health Technology Assessment (HTA) database provides international coverage of completed and ongoing health technology assessments and much of its

content is supplied by the members of the International Network of Agencies for Health Technology Assessment (INAHTA) and other HTA organisations around the world. It is a particularly good source for identifying grey literature as much of the information it contains is not published in the peer-reviewed literature accessible via bibliographic databases.

Database records provide weblinks to reports, project pages and organisation websites wherever possible so database users can easily access further information.

Like DARE, the HTA database is available via a number of different routes such as the Centre for Reviews and Dissemination's website and the Cochrane Library, so it is possible to undertake complex searches. Unlike DARE, search results will require sifting to identify systematic reviews as the database contains a much wider range of study types.

The database is produced by the Centre for Reviews and Dissemination (CRD), University of York.

### 7.3.3 Campbell Library: http://www.campbellcollaboration.org/lib/

The Campbell Library contains systematic reviews produced by the Campbell Collaboration covering crime and justice, education, social welfare and international development. They are also available in a peer-reviewed monograph series called Campbell Systematic Reviews (ISSN: 1891-1803).

### 7.3.4 The Database of Promoting Health Effectiveness Reviews (DoPHER): http://eppi.ioe.ac.uk/webdatabases/Intro.aspx?ID=2

The Database of Promoting Health Effectiveness Reviews (DoPHER) currently contains details of over 2500 reviews (systematic and nonsystematic) of health promotion and public health effectiveness.

The reviews in the register have been assessed and tagged with generic keywords (e.g. health focus and population group) and keywords indicating the quality of the review (aims, search, inclusion criteria, quality assessment methods, data extraction methods and analysis methods).

It is produced by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) which is part of the Social Science Research Unit at the Institute of Education, London.

### 7.3.5 Health Systems Evidence: http://www.mcmasterhealthforum.org/hse/

Health Systems Evidence describes itself as "a continuously updated repository of syntheses of research evidence about governance, financial and delivery

arrangements within health systems, and about implementation strategies that can support change in health systems".

Example reviews include decentralised versus centralised governance of health services and repeat dispensing of prescriptions in community pharmacies.

For each systematic review, there is a description of how recently the search for studies was conducted, the quality of the synthesis (measured using the AMSTAR criteria) and the countries in which the included studies in the synthesis were conducted. The content is available in a number of languages including Spanish, French and Portuguese as well as English.

The resource is freely available, although users need to register and then subsequently log in with a username and password. There are several search options available, and users can create their own one-page summary reports of records matching their search requirements. Currently the database contains approximately 3000 records.

Updating of the resource is supported by Cochrane Canada, the Program in Policy Decision-Making, Health Information Research Unit's McMaster PLUS, the Canadian Agency for Drugs and Technologies in Health (CADTH) Rx for Change, as well as by the McMaster Health Forum at McMaster University.

### 7.3.6 The JBI Database of Systematic Reviews and Implementation Reports: http://joannabriggslibrary.org/index.php/jbisrir

The database publishes both systematic reviews and protocols as well as implementation reports undertaken by the Joanna Briggs Institute and its international collaborating centres and groups.

Example reviews include the experience of self-care: a systematic review and systematic review of evidence on the impact of nursing workload and staffing on establishing healthy work environments.

There are a number of browsing and search options available on the webpages.

The Joanna Briggs Institute (JBI) is the international not-for-profit research and development arm of the School of Translational Science based within the Faculty of Health Sciences at the University of Adelaide, South Australia.

### 7.3.7 3ie Systematic Review Database: http://www.3ieimpact.org/en/evidence/systematic-reviews/

The 3ie database is a small collection of systematic reviews of the effectiveness of social and economic interventions of relevance to low- and middle-income countries. Examples of reviews available in the database include face washing promotion for preventing active trachoma and access to electricity for improving health, education and welfare in low- and middle-income countries: a systematic review.

As at October 2014, it contains summaries of over 200 reviews identified from a range of sources. The summary of each review includes details of the findings and


**Fig. 7.1** Screenshot of the interface for searching the 3ie systematic review database


**Fig. 7.2** Home page of PROSPERO

the methods used as well as a quality appraisal of the review. The search interface provides both basic and advanced options where there are a number of ready-made filters that can be selected by database users (Fig. 7.1).

The database producer, 3ie, is an international non-governmental organisation (NGO) that promotes the development of policies and programmes in low- and middle-income countries based on the best available evidence. The three main funders of 3ie are the Bill & Melinda Gates Foundation, UKaid, through the Department for International Development and the William and Flora Hewlett Foundation.

### 7.3.8 PROSPERO: http://www.crd.york.ac.uk/prospero/

PROSPERO is the open access prospective register for systematic review protocols launched by NIHR Centre for Reviews and Dissemination in February 2011. The key aims of the register are to facilitate transparency in the review process by making planned methods permanently and publicly available and provide a searchable database of ongoing reviews that will assist in the avoidance of unplanned duplication of work (Fig. 7.2).

Researchers enter a minimum dataset, agreed through international consultation; submissions are checked against the inclusion criteria and for sense but are not peer reviewed before being published on the register. The number of registrations was over 4700 in October 2014, making it an increasingly useful resource for finding out about reviews currently in progress. Records are permanent and registrants are encouraged to add links when their review is published.

A number of search options can be selected or the user can simply browse all of the available records. Search results can be saved individually in pdf format or printed as individual documents.

### 7.3.9 Social Care Online: http://www.scie-socialcareonline.org.uk/

Coverage of this resource includes social work, social care and services such as home care, child protection, safeguarding, adoption and fostering. Broad issues (e.g. integration, access and equality) that potentially impact on these services are also included in the database. In addition to journal articles, the database includes government policy, legislation and statistics. The primary focus of the database is the UK although comparative studies are also collected. Using the "content-type" filter in the advanced search interface, it is possible to restrict search results to records tagged as systematic reviews. Social Care Online is produced by the Social Care Institute for Excellence (SCIE) and is the UK's largest database of information on all aspects of social care and social work.

### 7.3.10 The Medion Database: http://www.mediondatabase.nl/

This resource focuses on systematic reviews of diagnostic studies but also includes systematic reviews of genetic tests. It has a simple search interface that incorporates a useful set of topic areas allowing users to limit their search results. As of January 2015, this database had over 3000 records.

### 7.3.11 Epistemonikos: http://www.epistemonikos.org/

This database contains umbrella reviews, systematic reviews and primary studies. The content of this database is gathered from multiple sources including the Cochrane Library, DARE, PubMed and the Biblioteca Virtual en Salud.

### 7.3.12 TRIP: http://www.tripdatabase.com/

TRIP compiles information from numerous sources such as the Cochrane Library, DARE and PubMed.

### 7.3.13 Sumsearch: http://sumsearch.org/

The content of this database is gathered from multiple sources including DARE, MEDLINE and the National Guideline Clearinghouse.

### 7.3.14 Guideline Collections

Some, but not all, clinical practice guidelines are based upon systematic reviews of the literature so resources such as the National Guidelines Clearinghouse (NGC) that collect guidelines and websites from organisations such as NICE (National Institute for Health and Care Excellence) and SIGN (Scottish Intercollegiate Guidelines Network) that list guidelines can also be used to identify relevant systematic reviews.

#### 7.3.14.1 The National Guidelines Clearinghouse (NGC): http://www.guideline.gov/index.aspx

Since June 2014, the clinical practice guidelines included in this resource are restricted to those that have been based upon a systematic review of the evidence. The NGC creates a structured summary that provides a standard set of information about each individual guideline to enable comparisons between guidelines.

Content can be accessed by browsing listings of topics or producers or by scanning the complete index of included summaries (2695 as at October 2014). A basic one word search option is provided as well as an advanced search facility that enables users to select various filters to limit results in various ways such as publication year, target population and clinical speciality. It is also possible to restrict to "methods used to analyze the evidence" by selecting from a number of options including, e.g. meta-analysis of individual patient data, meta-analysis of randomised controlled trials and systematic review.

The National Guidelines Clearinghouse is an initiative funded by the Agency for Healthcare Research and Quality (AHRQ) and the US Department of Health and Human Services.

#### 7.3.14.2 The National Institute for Health and Care Excellence (NICE) Website: https://www.nice.org.uk/guidance

The NICE website includes publication of several series of guidelines for clinical practice, public health, social care, safe staffing and medicines practice.

Guidelines can be accessed by using the website's search engine or by browsing through predefined selections of topics. There are six top level categories (conditions and diseases; health protection; lifestyle and wellbeing; population groups; service delivery, organisation and staffing; settings and environment) that are further subdivided into more specific topic areas. At this level a range of publications and documentation is available to download including pathways, technology appraisals, advice and guidelines.

NICE provides national guidance and advice to improve health and social care in the UK.

#### 7.3.14.3 Scottish Intercollegiate Guidelines Network (SIGN): http://www.sign.ac.uk/

The SIGN website includes a list of 141 guidelines that have been developed since the organisation was formed in 1993. All of the guidelines developed by SIGN are based upon a systematic review of the evidence.

As well as browsing the complete listing, there are some broad categories in place (e.g. obstetrics and gynaecology, mental health) that direct users to the relevant area of the website. SIGN develops and disseminates national clinical guidelines to improve the quality of health care in Scotland.

### 7.3.15 Using Bibliographic Databases Such as MEDLINE and Embase

Another method of identifying systematic reviews to include in an umbrella review is to use one of the many bibliographic databases that are available. If this approach is used, the most efficient method of searching may be to combine the relevant topic search terms with a systematic reviews search filter.

Filters are search strategies designed to retrieve specific types of records, and there are a number of these available to retrieve systematic reviews from databases such as a MEDLINE or Embase. Each filter will have been designed for a specific database and search interface, so this needs to be checked before use. The performance of individual filters will be measured by the authors who devise them, although performance is often subsequently tested by other researchers and sometimes published as a research paper.

Some database interfaces helpfully include the option of selecting predefined search filters so that the user can simply check a box to implement a study type restriction on their search results (Fig. 7.3).

### 7.3.16 McMaster University's Hedges Project: http://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx

The McMaster University's Hedges project includes a selection of filters that can be used in conjunction with topic terms to identify systematic reviews. The search filters for MEDLINE, Embase and PsycINFO [6, 7] presented on the website have all been derived from research projects. There are versions for use with OvidSP as well as MEDLINE PubMED interface. In addition, there are a range of sensitivity and precision options available, so if you want to ensure comprehensive coverage of your topic area, you can select a highly sensitive search filter, while if precision is of more importance, you can select a search filter that will retrieve a smaller, more focused group of records. Choosing the latter option could mean that you may

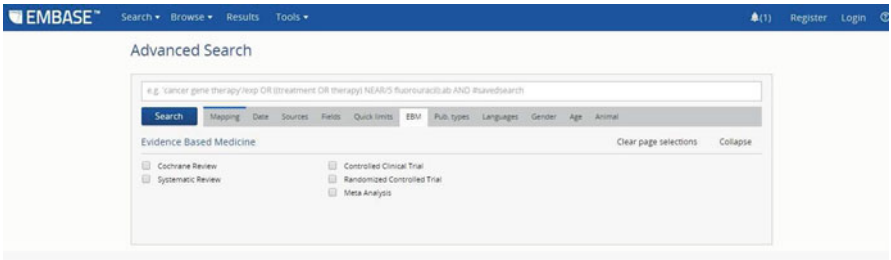



**Fig. 7.3** Search tool available in Embase

inadvertently omit relevant records. Users need to cut and paste the search filters from the McMaster website into their own database search interface.

### 7.3.17 McMaster's Health Evidence Website: http://www.healthevidence.org/default.aspx

Another website that provides a selection of systematic review search filters is McMaster's Health Evidence website. The filters described here for the MEDLINE and Embase databases are those that are used to populate the Health Evidence database that covers public health systematic reviews. The filter given here for CINAHL is now out of date as it was designed to work with the OvidSP interface rather than the current EBSCO interface. Lee et al. [8] describe the range of systematic review search filters including the ones used for Health Evidence. The paper is published in an open access journal, and the filters are accessible in the tables and additional material can be downloaded from the website.

### 7.3.18 Other Websites: https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home

One of the largest collections of search filters available is the ISSG Search Filter Resource. This is produced by a group of information professionals supporting research groups within England and Scotland providing technology assessments to the National Institute of Health and Care Excellence (NICE). It aims to identify, assess and test search filters designed to retrieve research by study design. It also contains a critical appraisal tool enabling the reader to assess the evidence on the performance of the search filters produced and assist in the selection of the most appropriate search filter.

Other websites listing systematic reviews search filters are the:

- The Canadian Agency for Drugs and Technologies in Health (CADTH) website, which included filters for MEDLINE, EMBASE and PsycINFO: http://www.cadth.ca/en/resources/finding-evidence-is/string

- The Centre for Reviews and Dissemination, University of York, (CRD) website lists filters for MEDLINE, Embase, PsycINFO and CINAHL that are used to populate the Database of Abstracts of Reviews of Effects (DARE): http://www.crd.york.ac.uk/crdweb/searchstrategies.asp

When using a filter, it is good to exercise caution and check that the filter is both up to date and the most appropriate for your search. In addition, you may like to read the performance reviews of the search filters to help you select which filter to use.

## 7.4 Summary

There are a range of specialist systematic review databases that may be useful when searching for systematic reviews, and we would recommend these databases as a good starting point. In order to conduct comprehensive searches, generic databases (such as MEDLINE and Embase) which include a wide range of study types may need to be searched. In such instances, using a carefully selected search filter can save time and reduce the number of irrelevant records to sift.

**Acknowledgements We would like to thank Julie Glanville, YHEC, for comments on an earlier draft.**

## References

1. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking systematic reviews in health care. York: University of York/Centre for Reviews and Dissemination; 2009. Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm.
2. Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]: the Cochrane Collaboration. 2011. Available from: www.cochrane-handbook.org.
3. Institute of Medicine. Finding what works in health care: standards for systematic reviews. Washington, DC: The National Academies Press; 2011.
4. Robinson KA, Whitlock EP, O'Neil ME, Anderson JK, Hartling L, Dryden DM, et al. Integration of existing systematic reviews into new reviews: identification of guidance needs. Syst Rev. 2014;3:60.
5. Bunn F, Trivedi D, Alderson P, Hamilton L, Martin A, Iliffe S. The impact of Cochrane Systematic Reviews: a mixed method evaluation of outputs from Cochrane Review Groups supported by the UK National Institute for Health Research. Syst Rev. 2014;3:125.
6. Montori V, Wilczynski N, Morgan D, Haynes R. Optimal search strategies for retrieving systematic reviews from MEDLINE: analytical survey. BMJ. 2005;330:68.
7. Wilczynski N, Haynes R, for the Hedges Team. EMBASE search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. J Clin Epidemiol. 2007;60:29–33.
8. Lee E, Dobbins M, DeCorby K, McRae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and meta-analyses. BMC Med Res Methodol. 2012;12:51.

# Abstracting Evidence

8

Giuseppe Biondi-Zoccai, Giuseppe La Torre,
Leonardo Roever, and Fabrizio D'Ascenzo

**Abstract**

Retrieval and collection of accurate and detailed data is an obviously crucial aspect of any umbrella review, overview of reviews, or meta-epidemiologic study. Yet there is limited evidence guiding this key reviewing step, and many reviewers overlook its importance and ensuing need for accurate planning and undertaking. Nonetheless, the available evidence and expert opinion is coherently supporting a set of best practices to ensure the validity, thoroughness, and usability of retrieved data. In particular, data abstraction should be performed by two or more independent reviewers, on formally developed and piloted report forms. The utmost transparency should be sought, for instance, storing reviewing details in online data repositories for scrutiny or subsequent use. Finally, the risk of duplication when conducting an umbrella review or overview of reviews (e.g., considering twice the same trial results because of being reported by two separate

G. Biondi-Zoccai, MD, MStat (✉)
Department of Medico-Surgical Sciences and Biotechnologies,
Sapienza University of Rome, Corso della Repubblica 79, Latina 04100, Italy

Eleonora Lorillard Spencer Cenci Foundation, Rome, Italy
e-mail: giuseppe.biondizoccai@uniroma1.it

G. La Torre
Eleonora Lorillard Spencer Cenci Foundation, Rome, Italy

Department of Public Health and Infectious Diseases,
Sapienza University of Rome,
Rome, Italy

L. Roever
Department of Clinical Research, Federal University of Uberlândia, Uberlândia, Brazil

F. D'Ascenzo
Divisione di Cardiologia, Dipartimento di Scienze Mediche, Città della Salute e della Scienza, Turin, Italy



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_8

systematic reviews) should be minimized, unless this is one of the meta-epidemiologic goals of the reviewing and research synthesis effort.

*It is a capital mistake to theorize before one has data.*

Arthur Conan Doyle (1859–1930) [1]

## 8.1 Introduction

Data are the backbone of any reviewing effort, and their careful yet efficient abstraction is crucial to enable the production of valid and similarly efficient umbrella reviews, overview of reviews, or meta-epidemiologic studies. Despite their obvious role, data and their collection means have often been overlooked or treated superficially. This is unfortunate as no statistical technique, even if highly sophisticated, can remedy bias due to suboptimal data abstraction (e.g., information bias). Indeed, the key aims which should inform data abstraction for tertiary research should be transparency, error minimization, and efficiency [2, 3].

Despite such importance and the self-evident relevance of the above goals, the evidence informing on best practices in data collection for systematic reviews is quite limited and practically absent for umbrella reviews, overview of reviews, or meta-epidemiologic studies [4–6]. We thus need to borrow from sources of evidence focused more on systematic reviews and meta-analyses and concomitantly rely largely on expert opinion (combining experience and expertise). Nonetheless, there is still room for credible recommendations on effective and efficient data collection methods. Careful application of such methods will typically yield high-quality data and maximize the validity of any reviewing exercise.

In general and apparently tautological terms, the minimum set of data requiring abstraction are those that are essential for the review of interest (e.g., details on populations, interventions/comparisons/exposures of interest, and outcomes), supplemented by those ancillary data which are required to let the reader and decision-maker put the findings into context and check their plausibility and applicability. In addition, a thorough data collection process will ensure that the data already abstracted may be useful to other researchers in the future, a practice which will become more and more common in the era of open dataset access and online data repositories, such as the Systematic Review Data Repository (SRDR) [7, 8]. On top of the following recommendations, the reader is referred to other important documents, and in particular the Cochrane Handbook for Systematic Reviews of Interventions and the Joanna Briggs Institute Reviewers' Manual sections devoted to overviews of reviews and umbrella reviews, respectively, for other important insights [4, 5].

## 8.2 Collection of Review Data

In keeping with our default definitions of umbrella reviews, overviews of reviews, and meta-epidemiologic studies, the key unit for data collection in such reviewing efforts is a systematic review, a meta-analysis, or any other type of secondary

research. Accordingly, great attention should be given to the peculiarities of this study design, not withstanding the ultimate focus on the goal of summarizing, in most cases, the available evidence in terms of populations, interventions/comparisons/exposures of interest, and outcomes (Table 8.1) [4, 9–14]. In other words, while apparently the unit of study is a review, ultimately the focus of the reader and user will likely be on patients. In addition, details on the persons undertaking the umbrella review, bibliographic and bibliometric features of the review under analysis, and any ancillary information which may be considered relevant for confounding and/or effect modification (e.g., funding and conflicts of interest) can be

**Table 8.1** Checklist for data abstraction for an umbrella review, overview of reviews, or meta-epidemiologic study, having systematic reviews as primary object of research

| Domain | Elaboration |
|---|---|
| Identification | Bibliographic details (e.g., authors, title, journal, year), database ID, review/trial registration identifier, report details, and review/study details (in case of complex clusters of reviews and trials) |
| Eligibility | Features relevant for the inclusion or exclusion of the review (or the primary study) |
| Methods[a] | Methodological and reporting features impacting on results interpretation relevant to the review (or the primary study) |
| Participants | Details on patients included in the review (or in the primary study) |
| Interventions/comparisons | Details on interventions and comparisons included in the review (or in the primary study) |
| Outcomes | Details on outcomes included in the review (or in the primary study) |
| Results | Details on the specific results provided in the review (or in the primary study), including, when applicable, summary effect estimates, inconsistency estimates, and other ancillary analytical results (e.g., small study effect assessment) |
| Interpretation | Details on the specific discussion of the review (or of the primary study) findings |
| Additional details | Important data potentially impacting on the application or interpretation of the results (e.g., funding, conflicts of interest, or bibliometric data) and tracking of contacts with authors |
| Reviewer details | Data on reviewer identification, time of review, and any update |

[a]Including items relevant to validity and reporting appraisal for reviews or primary studies, including those required by A Measurement Tool to Assess Systematic Reviews (AMSTAR), the Overview Quality Assessment Questionnaire (OQAQ), the Rapid Appraisal Protocol (RAP), the Guidelines for Assessing the Quality and Applicability of Systematic Reviews of the National Center for the Dissemination of Rehabilitation Research (NCDRR), the Documentation and Appraisal Review Tool (DART) or the International Narrative Systematic Assessment tool for reviews, and the Cochrane Risk of Bias Tool or the Newcastle-Ottawa Scale (NOS) for primary studies [2–5, 9–14]

abstracted [15]. Finally, the data abstraction process should be planned in keeping with the preferred approach for validity assessment, as a one-stop collection process is usually preferable to minimize time loss and inconsistency. Accordingly, the specific items of the different tools to appraise the validity of reviews should be borne in mind in this phase of planning and reviewing [9–13].

## 8.3 Collection of Study Data

In many cases, reviewers aiming for an umbrella review, overview of review, or meta-epidemiologic study may rely on the data presented in the shortlisted reviews to gather details on primary sources of evidence (e.g., randomized controlled trials, observational studies). In such instances, it is important though to make sure that the original data collection process had been valid and that no systematic error had been entered in the processes taken to complete and report the review (e.g., typographical errors in publishing the review). Thus, it may be useful to double check atleast some of the shortlisted trials, either from the original source document or through another review including the same study.

In other cases, it may be needed to collect additional details from the primary studies, either because they had not been collected or because some primary studies had not been included at all. In such settings then an umbrella review, overview of review, or meta-epidemiologic study entails the same methodological aspects and skills required for a traditional systematic review. This specific topic is largely beyond our present scope, and high-quality recommendations are available elsewhere [3–5, 16–18]. As previously clarified, it is paramount to collect details informing on populations, interventions/comparisons/exposures of interest, and outcomes, as well as reviewer, bibliographic, and validity data, to enable comprehensive and detailed reporting and analysis. In addition, validity appraisal of such primary studies may be of interest, and in such cases details sufficient to enable the application of established appraisal, such as the Cochrane Risk of Bias Tool or the Newcastle-Ottawa Scale, are also required [4, 14].

## 8.4 Troubleshooting

Besides the minimal and optimal set of data to be collected, it is important to define and plan how to best abstract data [8]. In most cases, data retrieval should be performed by two or more reviewers, independently, with divergences traced and solved after consensus or involvement of another reviewer [19]. It may occasionally be accepted that one reviewer extracts data, while a different one checks all of them for accuracy. In any case, divergences should be solved constructively, tracking explicitly the consensus development. In addition, agreement between different reviewers may be explicitly appraised with specific tests, such as the Cohen's kappa coefficient, to ensure consistency.

Data collection forms, being them on paper or electronic, should be piloted in order to ensure validity and agreement among different reviewers, especially if the reviewing effort is substantial (e.g., >10 reviews included). Reviewers should be trained, though, not withstanding the apparently limited impact of reviewer experience on data quality [20]. The transition from paper forms to spreadsheets, all-purpose databases, web surveys, and, eventually, specialized software is already occurring, and the benefits of the latter type of tools are clear especially when the number of included reviews and trials is substantial [9, 21–24]. Use of standardized forms and definitions is also going to be particularly welcome in the, hopefully near, future, when authors will most likely upload their data on public repositories, such as the SRDR, to enable other researchers checking and using them at their will [7, 8].

While blinding reviewers to the identification details of the authors of the shortlisted reviews is conceivable, it may be logistically challenging, and to date there is no irrefutable evidence to support it [25, 26]. In addition, it is recommended to quote verbatim in the abstraction forms those sections of the shortlisted reviews which guide a specific labeling. Moreover, quantitative data as originally reported should be preferred to back-computed values, which may lead to biased results if mathematical transformations are used inconsistently [27]. Accordingly, denominators, counts, and samples at risk should always be explicitly stated, especially when different outcomes at varying risk of attrition are considered. Whenever data are missing, inconsistent, or need confirmation, authors of the shortlisted reviews or even authors of the original primary studies can be contacted. Such efforts should be based on an explicit plan, recorded and transparently reported to avoid duplicate efforts. Whereas there is an ongoing push to perfect automated data abstraction and validity appraisal in systematic reviewing efforts, to date there is no room for their prime time use [28, 29].

Another very important issue is the potential clustering of multiple systematic reviews including different reports all stemming from the same initial trial [30]. This phenomenon, which can be considered similar to a snowball effect, should be taken into account when describing the umbrella review results and even more importantly by means of hierarchical models in case of formal inferential analysis [31]. Indeed, meta-epidemiologic studies may aptly exploit duplicate reviewing efforts to highlight important mediators of review results [32–34], but such duplicate entries may instead significantly bias effect estimates in umbrella reviews if not recognized and managed correctly.

Finally, and more pragmatically, it is best to limit the amount of collected data to a reasonably small to moderate set, at least when the number of included systematic reviews is large, to avoid creating an exceedingly extensive dataset which only dilutes the core inferential message and can also raise the temptation of multiplicity and cherry picking.

**Conclusion**

Data collection represents a very important aspect of any reviewing effort. Given the lack of a large evidence base on best practices in this step of an umbrella review, overview of review, or meta-epidemiologic study, it is necessary to apply

sound judgment and balance the desire for detailed datasets with pragmatism, without however risking being superficial. It is clear that bias forced into a reviewing effort at this stage is very difficult to recognize later on, and thus, robust methods must be employed. In the future, it is likely that the duplication of data collection efforts will be minimized by uploading standardized data abstraction forms for clinical studies and systematic reviews into dedicated online data repositories.

## References


1. Arthur Conan Doyle. BrainyQuote.com, Xplore Inc; 2015. Available at: http://www.brainyquote.com/quotes/quotes/a/arthurcona131991.html. Last accessed 21 Aug 2015
2. Guyatt G, Rennie D, Meade MO, Cook DJ. Users' guide to the medical literature. A manual for evidence-based clinical practice. 2nd ed. New York: McGraw-Hill Professional; 2008.
3. Biondi-Zoccai G, editor. Network meta-analysis: evidence synthesis with mixed treatment comparison. Hauppauge: Nova; 2014.
4. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011. Available from: www.cochrane-handbook.org. Last accessed 18 Aug 2015.
5. The Joanna Briggs Institute. The Joanna Briggs Institute Reviewers' Manual. Methodology for JBI umbrella reviews. Adelaide: The University of Adelaide; 2014.
6. Li L, Tian J, Tian H, Sun R, Liu Y, Yang K. Quality and transparency of overviews of systematic reviews. J Evid Based Med. 2012;5:166–73.
7. Systematic Review Data Repository (SRDR). Available at: srdr.ahrq.gov. Last accessed 18 Aug 2015.
8. Li T, Vedula SS, Hadar N, Parkin C, Lau J, Dickersin K. Innovations in data collection, management, and archiving for systematic reviews. Ann Intern Med. 2015;162:287–94.
9. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44:1271–8.
10. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
11. Diekemper RL, Ireland BK, Merz LR. Development of the Documentation and Appraisal Review Tool for systematic reviews. World J Meta Anal. 2015;3:142–50.
12. Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. J Clin Epidemiol. 2015;68:574–83.
13. La Torre G, Backhaus I, Mannocci A. Rating for narrative reviews: concept and development of the International Narrative Systematic Assessment tool. Senses Sci. 2015;2:31–5.
14. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol. 2010;25:603–5.
15. Jørgensen AW, Maric KL, Tendal B, Faurschou A, Gøtzsche PC. Industry-supported meta-analyses compared with meta-analyses with non-profit or no support: differences in methodological quality and conclusions. BMC Med Res Methodol. 2008;8:60.
16. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in healthcare. York: University of York; 2009.
17. Eden J, Levit L, Berg A, Morton S, editors. Finding what works in health care. Standards for systematic reviews. Washington, DC: The National Academies Press; 2011.

18. Agency for Healthcare Research and Quality. Methods guide for effectiveness and comparative effectiveness reviews. Rockville: Agency for Healthcare Research and Quality (US); 2008.
19. Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. J Clin Epidemiol. 2006;59(7):697–703. Epub 2006 Mar 15.
20. Horton J, Vandermeer B, Hartling L, Tjosvold L, Klassen TP, Buscemi N. Systematic review data extraction: cross-sectional study showed that experience did not increase accuracy. J Clin Epidemiol. 2010;63:289–98.
21. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J Am Med Inform Assoc. 2002;9:653–8.
22. Elamin MB, Flynn DN, Bassler D, Briel M, Alonso-Coello P, Karanicolas PJ, Guyatt GH, Malaga G, Furukawa TA, Kunz R, Schünemann H, Murad MH, Barbui C, Cipriani A, Montori VM. Choice of data extraction tools for systematic reviews depends on resources and review complexity. J Clin Epidemiol. 2009;62:506–10.
23. Ip S, Hadar N, Keefe S, Parkin C, Iovin R, Balk EM, Lau J. A web-based archive of systematic review data. Syst Rev. 2012;1:15.
24. Doctor evidence. Available at: drevidence.com. Last accessed 18 Aug 2015.
25. Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. Lancet. 1997;350:185–6.
26. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17:1–12.
27. Gøtzsche PC, Hróbjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses that use standardized mean differences. JAMA. 2007;298:430–7.
28. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Syst Rev. 2015;4:78.
29. Marshall IJ, Kuiper J, Wallace BC. Automating risk of bias assessment for clinical trials. IEEE J Biomed Health Inform. 2015;19:1406–12.
30. Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. BMJ. 1997;315:635–40.
31. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol. 2010;63:875–82.
32. Biondi-Zoccai GG, Lotrionte M, Abbate A, Testa L, Remigi E, Burzotta F, Valgimigli M, Romagnoli E, Crea F, Agostoni P. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: case study. BMJ. 2006;332:202–9.
33. Nowbar AN, Mielewczik M, Karavassilis M, Dehbi HM, Shun-Shin MJ, Jones S, Howard JP, Cole GD, Francis DP, DAMASCENE writing group. Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis. BMJ. 2014;348:g2688.
34. Peruzzi M, De Falco E, Abbate A, Biondi-Zoccai G, Chimenti I, Lotrionte M, Benedetto U, Delewi R, Marullo AG, Frati G. State of the art on the evidence base in cardiac regenerative therapy: overview of 41 systematic reviews. Biomed Res Int. 2015;2015:613782.

# Appraising Evidence

9

William L. Baker, Meg Bennetts, Craig I. Coleman, and Joseph C. Cappelleri

**Abstract**

Appraisal of the quality of evidence when performing an umbrella review remains an integral part of the process. However, specific recommendations for evaluating systematic review quality have not conveniently coalesced for those interested in performing umbrella reviews. In this chapter, we will coalesce and discuss the tools available to critically appraise systematic reviews when performing an umbrella review. While assessing individual study risk of bias is not a standard part of an umbrella review, understanding the tools used by the included systematic reviews for this step is important. Guidelines exist for reporting systematic reviews of both randomized controlled trials and observational studies. Methods for assessing quality of a systematic review and the strength of the evidence are well established. The expanding field of network meta-analysis has also provided good practices to appraise reporting and quality of its reviews. Future guidance documents on performance of umbrella reviews should give careful consideration to providing recommendations for formal risk of bias assessment as part of the evidence appraisal process.

W.L. Baker (✉)
Department of Pharmacy Practice, University of Connecticut School of Pharmacy, 69 N. Eagleville Rd, Unit 3092, Storrs, CT 06269-3092, USA
e-mail: william.baker_jr@uconn.edu

M. Bennetts
Global Pharmacometrics, Pfizer Limited, Sandwich, Kent, UK

C.I. Coleman
Department of Pharmacy Practice, University of Connecticut School of Pharmacy, Storrs, CT, USA

J.C. Cappelleri
Statistics, Pfizer Inc, Groton, CT, USA



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_9

## 9.1 Introduction

Assessment of individual study quality or risk of bias (or both) is a well-recognized, integral step in the systematic review process [1]. This formal evaluation of the literature base allows for appropriate interpretations and conclusions to be made regarding the findings, particularly when meta-analysis is performed [2]. The process of critically evaluating the literature base similarly extends to the performance of umbrella reviews whereby quality of included systematic reviews must be performed. The importance of this step in the process is underlined by research showing not only the inconsistent manner in which systematic reviews are reported in the medical literature [3], but also the lack of methodological rigor for overviews of reviews [4]. More specifically, Pieper and colleagues [4] showed that one-third of included overviews of reviews did not perform a quality assessment for their included systematic reviews. Hartling and colleagues [5] showed that only 37 % of overview of reviews identified reported performing quality assessment of the included systematic reviews.

The challenge with appraising the evidence as part of an umbrella review is identifying the most suitable tool to use. The Cochrane Collaboration recommends use of their own Risk of Bias Tool for assessing randomized controlled trials in a systematic review and avoiding use of scales that provide aggregate quality scores [2, 6]. Others have agreed with avoidance of summary scores to delineate studies as high quality or not [7]. However, specific recommendations for evaluating systematic review quality are not available for those interested in performing umbrella reviews. The Cochrane Collaboration recommends that some assessment of the methodological quality of the reviews in an overview should be provided, but does not recommend a specific instrument [2]. Therefore, a review of the tools that can be used by researchers performing overview of reviews would be of use.

In this chapter, we will discuss the tools available to critically appraise systematic reviews when performing an umbrella review. Specifically, we will describe tools to evaluate the quality of individual studies (Sect. 9.2), tools to inform reporting of systematic reviews (Sect. 9.3), tools to appraise the quality of systematic reviews (Sect. 9.4), and tools for to appraise network meta-analyses (NMA) (Sect. 9.5).

## 9.2 Tools to Evaluate Quality of Individual Studies

While the primary focus of this chapter will be to discuss tools for appraising systematic reviews when included in an umbrella review, it is also important for investigators to be familiar with strategies used to assess quality of individual studies included in the systematic reviews, as a prerequisite step for the building blocks of an umbrella review. A large number of tools are available to assess quality or risk of bias (or both) of randomized controlled trials and observational studies [8, 9]. A thorough review of these is beyond the scope of this chapter. Instead, we present a few of the most commonly used and currently recommended tools that may be encountered when performing an umbrella review.

**Table 9.1** Cochrane risk of bias tool

| Bias classification | Description | Relevant risk of bias domains | Domain number |
|---|---|---|---|
| Selection bias | Systematic differences between baseline characteristics of the groups that are compared | Sequence generation | 1 |
| | | Allocation concealment | 2 |
| Performance bias | Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest | Blinding of participants and personnel | 3 |
| | | Other potential threats to validity | 7 |
| Detection bias | Systematic differences between groups in how outcomes are determined | Blinding of outcome assessment | 4 |
| | | Other potential threats to validity | 7 |
| Attrition bias | Systematic differences between groups in withdrawals from a study | Incomplete outcome data | 5 |
| Reporting bias | Systematic differences between reported and unreported findings | Selective outcome reporting | 6 |

Adapted from the *Cochrane Handbook for Systematic Reviews* [2]

The validity of a study has two dimensions: whether it is asking an appropriate research question for the targeted population of interest (external validity) and whether this question is being answered in the right way and free from systematic errors in results, inferences, or biases (internal validity). The Cochrane Collaboration tool for assessing bias in randomized trials gives a useful classification of the different forms of bias: selection, performance, detection, attrition, and reporting bias with corresponding domains that are assessed by the tool for individual studies within a systematic review, six of these domains are fully defined for assessment in every study; the seventh, final, domain is for "other sources of bias" that are relevant only in certain circumstances and across the different classes of bias (Table 9.1) [2, 10]. This final domain can be assessed as a single entry for studies as a whole or split into pre-specified entries to address specific other risks of bias.

The tool was developed to adhere to seven principles and so:

1. Does not use quality scales
2. Only focuses on internal validity
3. Assesses the risk of bias in trial results, not the quality of reporting or methodological problems
4. Requires reviewer judgment
5. Assesses domains based on a combination of theoretical and empirical consideration
6. Focuses on risk of bias in the data as represented in the review rather than other sources
7. Reports outcome-specific evaluations of risk of bias

For each of the risk of bias domains, assessment is in two parts. The first describes the relevant trial characteristics on which the risk of bias is based and the second an assigned judgment of "low risk," "high risk," or "unclear risk" of bias. Higgins and Altman give detailed criteria for each domain to aid this judgment [2]. Two figures are considered for presentation of risk of bias by study and domain across a systematic review. The first is a graph showing for each domain the percentage of studies with low, high, or unclear risk of bias and the second a summary (cross tabulation) of study by domain with each cell indicating the relevant risk of bias. Assessment of the overall risk of bias requires judgments of the relative importance of the different domains and should be informed by the source, direction, and likely magnitude of bias and might be considered at several levels: for a study across outcomes as it cannot be assumed that the risk of bias is the same across all outcomes in a study, for an outcome within a study across domains which is the recommended level to summarize the risk of bias within a study, or for an outcome across studies for meta-analysis.

As an extension to the Cochrane Collaboration tool, Higgins et al. [11] published a tool to assess the quality of a meta-analysis that provides greater granularity and response categories with an emphasis on the evaluation of statistical and interpretational issues. This tool consists of 43 items divided into four key categories: data sources, analysis of individual studies by the meta-analyst, general meta-analysis, and reporting and interpretation, with a summary judgment question appearing at the end of each category. Accompanying the tool is detailed guidance for completing the assessment form.

The Jadad scale is one of the most frequently utilized tools for assessing the quality of reporting of a randomized controlled trial (RCT), and many authors use it to evaluate risk of bias [12]. The scale consists of five questions that are divided into three different domains:

1. Randomization
2. Blinding
3. Description of withdrawals and drop-outs

The appropriateness of both randomization and blinding is also assessed. Each of the five questions is given a point if it was present in the RCT. Some authors designate a high-quality RCT when the Jadad score is greater than 3. Despite its widespread use, it has been criticized for evaluating the quality of RCT reporting rather than risk of bias and for the use of a summary scoring (although domain could be reported separately). As such, the Cochrane Collaboration explicitly discourages its use [2].

The Newcastle–Ottawa Scale (NOS) was designed to assess the quality of non-randomized studies, including case–control, cohort, and cross-sectional research designs [13]. It uses a "star system" to evaluate three areas of the study:

1. The selection of the study groups
2. The comparability of the groups
3. The ascertainment of either the exposure or outcome of interest

Separate scales geared towards case–control and cohort studies are available. Up to 10 stars can be awarded for case–control studies and up to 13 stars awarded for cohort studies, with a higher quantity of stars designating higher quality.

The Downs and Black checklist assesses the methodological quality of both RCT and nonrandomized studies [14]. This extensive questionnaire consists of 27 "yes" or "no" questions across five different sections:

1. Study quality
2. External validity
3. Study bias
4. Confounding and selection bias
5. Power of the study

A total maximum score of 31 is given, with higher values designating higher quality. Compared with other available tools, the Downs and Black can be confusing and time-consuming, ranging from 10 to 45 min per study.

## 9.3 Tools to Inform Reporting of Systematic Reviews

As previously mentioned, studies have shown that overall reporting of systematic reviews is variable and has been historically poor [3]. Interestingly, those published in higher impact factor journals were often judged to be of higher quality [15]. These results are not surprising given that many higher-quality journals require authors to comply with standard reporting guidance tools, which are discussed below. Given this heterogeneity, however, we felt it was appropriate to discuss the various reporting guidelines that are available, as it is relevant to performing of an umbrella review.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement consists of a 27-item checklist and a four-phase flow diagram [1]. The checklist includes items deemed essential for transparent reporting of a systematic review or meta-analysis across seven sections: Title, Abstract, Introduction, Methods, Results, Discussion, and Funding. In addition, the explanation and elaboration document is intended to enhance the use and understanding of the PRISMA statement, providing meaning and rationale for each checklist item through examples and explanations [16].

The research questions being addressed is one of the most critical parts of a systematic review. PRISMA recommends a structured approach to help frame explicit questions that uses five components to facilitate the process. Commonly known by the acronym "PICOS" this approach defines:

- The patient population and/or the disease being addressed (P)
- The interventions or exposure (I)
- The comparator group (C)
- The outcome or endpoint (O)
- The study designs (S)

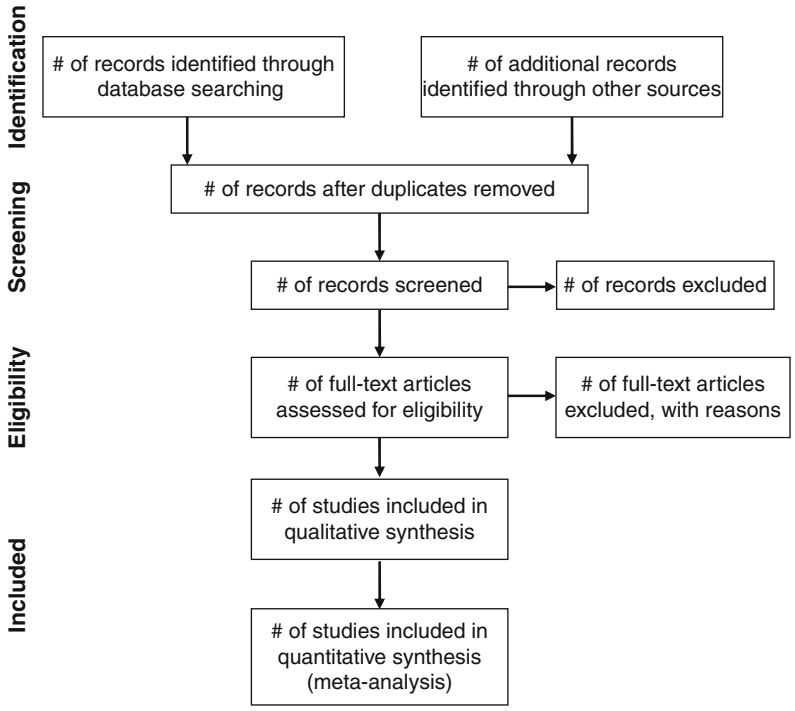



**Fig. 9.1** PRISMA flowchart (Adapted from Moher et al. [1])

The PRISMA four-phase flow diagram (Fig. 9.1) has become ubiquitous in meta-analysis papers and depicts the flow of information through the different phases of a systematic review. It maps out the number of records identified, included, and excluded and the reasons for exclusions.

The 27-item checklist is ordered across the seven sections according to the expected layout of a published report. In particular, the methods section includes items that should be in place and reported in planning and performing a systematic review; the result section specifies expected output (Table 9.2).

Although PRISMA focuses on randomized trials, it can be used as a basis for reporting other types of research designs that evaluate interventions. While the PRISMA statement assesses the quality of the reporting of a published systematic review or meta-analysis, it is not an instrument to gauge the quality of the underlying systematic review per se.

Systematic reviews should build on a detailed, well-described, protocol. To date, few protocols are published or are mentioned in published systematic reviews. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) 2015 statement consists of a 15-item checklist to help develop and report a systematic review protocol [17]. Again, an explanation and elaboration document provides understanding of the necessity of each item and a model example [18].

There are two further specialized extensions to the PRISMA guidelines. PRISMA-Abstracts provides a checklist to give authors a structured framework for

**Table 9.2** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist

| Item | Section/topic | Item | Section/topic | Item | Section/topic |
|---|---|---|---|---|---|
| | **Title** | | **Methods** | | **Results** |
| 1. | Title | 5. | Protocol and registration | 17. | Study selection |
| | **Abstract** | 6. | Eligibility criteria | 18. | Study characteristics |
| 2. | Structured summary | 7. | Information sources | 19. | Risk of bias within studies |
| | **Introduction** | 8. | Search | 20. | Results of individual studies |
| 3. | Rationale | 9. | Study selection | 21. | Synthesis of results |
| 4. | Objectives | 10. | Data collection process | 22. | Risk of bias across studies |
| | | 11. | Data items | 23. | Additional analysis |
| | | 12. | Risk of bias in individual studies | | **Discussion** |
| | | 13. | Summary measures | 24. | Summary of evidence |
| | | 14. | Synthesis of results | 25. | Limitations |
| | | 15. | Risk of bias across studies | 26. | Conclusion |
| | | 16. | Additional analyses | | **Funding** |
| | | | | 27. | Funding |

Adapted from Moher et al. [1]

condensing a systematic review or meta-analysis, into the essentials required for good reporting in journal and conference abstracts [19]. PRISMA-Equity takes the PRISMA 27-item checklist and extends the item descriptions to give reporting guidelines for systematic reviews of effects on inequities in health outcomes and health-care use across socioeconomic groups and other characteristics with an aim to improve global health equity [20].

The Meta-Analysis of Observational Studies in Epidemiology (MOOSE) developed a checklist for preferred reporting of meta-analyses of nonrandomized evidence [21]. Recommendations were the result of a workshop focused on meta-analyses in education and the social sciences in response to variable publishing practices in these areas. The checklist is organized into background, search strategy, methods, results, discussion, and conclusions. This model has been followed by other reporting documents, as discussed below.

## 9.4 Tools to Appraise the Quality of Systematic Reviews

As previously mentioned, appraising the quality of included reviews when performing an umbrella review is an integral step in the process and recommended by the Cochrane Collaboration [2]. While a variety of tools are available, no consensus exists on the most appropriate to use for umbrella reviews. We therefore highlight

**Table 9.3** Oxman and Guyatt index for appraisal of reviews

| Question | Details |
|---|---|
| 1 | Were the search methods used to find evidence stated? |
| 2 | Was the search for evidence reasonably comprehensive? |
| 3 | Were the criteria used for deciding which studies to include in the report included? |
| 4 | Was bias in the selection of studies avoided? |
| 5 | Were the criteria used for assessing the validity of the included studies reported? |
| 6 | Was the validity of all studies referred to in the text assessed using appropriate criteria? |
| 7 | Were the methods used to combine the findings of the relevant studies reported? |
| 8 | Were the findings of the relevant studies combined appropriately relative to the primary question the overview addresses? |
| 9 | Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview? |
| 10 | How would you rate the scientific quality of the overview? |

Adapted from Oxman and Guyatt [23]
The first p items of the scale are rated on a three-point scale (no, partially/can't tell, yes). The final (and tenth) question is based on an ordinal scale from 1 to 7, 1 to 3 representing extensive or major flaws and 5 to 7 representing minor or minimal flaws

explanations for each, including tables with information on the specific criterion used, and leave it to readers to consult the original publications for more details.

The first scale to provide a checklist of preferred items for assessing the quality of a meta-analysis was published by Sacks and colleagues in 1987 [22]. They proposed 23 items across six categories that were suggestive of a high-quality review. The categories included:

1. Study design
2. Combinability
3. Control of bias
4. Statistical analysis
5. Sensitivity analysis
6. Application of results

While many of the individual items are similar to the PRISMA statement for reporting of systematic reviews, a few differences remain. The Sacks scale additionally includes identifying various types of bias, including selection bias and data-extraction bias, which sets it apart from PRISMA. This scale has been a useful progenitor for subsequent scales that have evolved over the years commensurate with the rapid development of more detailed techniques on systematic reviews and meta-analysis.

Until recently, the most commonly used scale for assessing the quality of a review was provided by Oxman and Guyatt [23]. The team of Oxman and Guyatt developed the Overview Quality Assessment Questionnaire (OQAQ) with the aim of assessing the scientific quality of review articles (Table 9.3) [23]. The validity of the scale has been tested and validated using several different measures [23, 24].

**Table 9.4** AMSTAR tool for assessing methodological quality of systematic reviews

| Question | Details |
|---|---|
| 1 | Was an "a priori" design provided? |
| 2 | Was there duplicate study selection and data extraction? |
| 3 | Was a comprehensive literature search performed? |
| 4 | Was the status of publication (i.e., gray literature) used as an inclusion criterion? |
| 5 | Was a list of studies (included and excluded) provided? |
| 6 | Were the characteristics of the included studies provided? |
| 7 | Was the scientific quality of the included studies assessed and documented? |
| 8 | Was the scientific quality of the included studies used appropriately in formulating conclusions? |
| 9 | Were the methods used to combine the findings of studies appropriate? |
| 10 | Was the likelihood of publication bias assessed? |
| 11 | Was the conflict of interest stated? |

Adapted from Shea et al. [27]
Each item is answered as either "yes," "no," "can't answer," or "not applicable"

The OQAQ includes a total of ten items pertaining to individual aspects in the reporting of a systematic review, with a final question eliciting an overall scientific quality of the systematic review. Each of the first nine items is assessed using a three-point scale (no, partially/can't tell, yes). The final question (the tenth item) is based on a scale of 1–7, with a score of 1–3 representing extensive or major flaws and 5 or 7 representing minor or minimal flaws [25]. Studies using the OQAQ have found that lower-quality ratings are more likely to show treatment benefit [24, 26]. The potential concern with this tool is that it lacks several methodological domains that are now thought to be relevant and are included in more modern assessments, such as evaluations of heterogeneity and publication bias.

The most frequently recommended tool for use within an umbrella review is the Assessment of Multiple Systematic Reviews (AMSTAR) [27, 28]. This easy-to-use and validated [29] tool was developed using many of the criteria from the Sacks [22], as well as Oxman and Guyatt scales [23]. An initial list of 37 items was compiled and used to appraise 99 published systematic reviews and 52 Cochrane reviews. Each item was analyzed using these reviews and a final list of 11 components was proposed (Table 9.4) [27]. Each component is answered with responses of either "yes," "no," "can't answer," or "not applicable." This scale has been externally validated by a couple of groups [28, 29]. These studies showed it to have good interobserver agreement, reliability, and construct validity.

The average time to complete a single review using the AMSTAR tool was 15 min, further showing its ease of use [29]. The AMSTAR tool is now the most frequently utilized method for assessing the quality of a systematic review [5, 7]. A group of authors subsequently devised a revised version of AMSTAR (R-AMSTAR) with the aim of providing improved quantification of systematic review quality [30]. The R-AMSTAR assigns an overall quality score to the systematic review, a process, as previously mentioned, that is not recommended by the Cochrane

Collaboration [2]. While a number of studies have provided external validation of AMSTAR across medical subspecialties, such comparative evidence supporting R-AMSTAR is lacking [31].

The tools discussed to this point can be used to assess the quality of included systematic reviews of both RCTs and non-RCTs when performing an umbrella review. However, they give little or no consideration to the strength of the evidence supporting the conclusions of the systematic reviews. Grades of Recommendation, Assessment, Development and Evaluation (GRADE) provide a framework for rating quality of evidence and grading strength of recommendation in health care. In 2011 the *Journal of Clinical Epidemiology* published a series of 20 GRADE articles (http://www.gradeworkinggroup.org/publications/JCE_series.htm) to provide a step-by-step guide through the process of clarifying the questions, deciding on the importance of outcomes, rating the quality of evidence, summarizing the evidence and diagnostic tests, and making recommendations and special challenges in using observational studies [32, 33].

The GRADE approach is to rate the quality of a body of evidence for each main outcome of interest, not of individual studies, and within the context of a systematic review, GRADE reflects how confident we are that an effect estimate is close to the truth [34]. GRADE rates the quality of evidence for each outcome across studies on a four-level scale. Randomized trials start at the high-quality level, while observational studies start at the low quality level. Each outcome in each trial is then assessed against five reasons to downgrade and three reasons to upgrade as described in Table 9.5. A final rating of quality for each outcome is given across studies. As GRADE is "outcome centric," quality is likely to differ for outcomes both within and across studies. An impressive list of organizations, including the Cochrane Collaboration, has endorsed or is using the GRADE framework for grading the quality of evidence [2].

## 9.5 Tools to Appraise the Quality of Network Meta-analyses

Network meta-analysis (NMA) involves indirect treatment comparisons (ITC) or mixed treatment comparisons (MTC), which include both direct and indirect evidence. It provides quantitative information for evidence-based decision making in the absence of RCTs involving direct comparisons of all the treatments of interest within the studies [35, 36]. Systematic reviews of networks of multiple interventions that combine both direct and indirect comparisons are becoming increasingly prevalent and are highly informative for decision making. Thus, discussion of the current emerging tools for assessing the quality of NMA is appropriate should they be incorporated into an umbrella review.

The quality of an ITC depends on the methodology and the validity of underlying assumptions. In 2010, in the absence of published recommendations on how ITCs should be reported, Donegan and colleagues carried out a review of reporting and methodological quality [37]. Forty-three NMA reviews between 1992 and 2007 were assessed against developed quality criteria: the indirect comparison method, the

**Table 9.5** The GRADE approach to rating quality of evidence

| Study design | Quality of evidence | 5 Reasons to downgrade[a] | | 3 Reasons to upgrade[a] | |
|---|---|---|---|---|---|
| *Randomized trial* ➡ | *High* ⊕⊕⊕⊕ Further research is very unlikely to change our confidence in estimate of effect | *1. Risk of bias* | | *1. Large effect*[e] | |
| | | Serious | (−1) | +1 Large | (+1) |
| | | Very serious | (−2) | +2 Very large | (+2) |
| | | *2. Inconsistency*[b] | | *2. Dose response*[f] | |
| | | Serious | (−1) | +1 Evidence of gradient | (+1) |
| | | Very serious | (−2) | *3. All plausible confounding*[f] | |
| | *Moderate* ⊕⊕⊕◯ Further research likely to have impact on confidence in the estimate of effect and may change the estimate | *3. Indirectness*[c] | | Would reduce a demonstrated effect | (+1) |
| | | Serious | (−1) | | |
| | | Very serious | (−2) | Would suggest a spurious effect when results show no effect | (+1) |
| *Observational studies* ➡ | *Low* ⊕⊕◯◯ Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate | *4. Imprecision*[d] | | | |
| | | Serious | (−1) | | |
| | | Very serious | (−2) | | |
| | *Very low* ⊕◯◯◯ Any estimate of effect is very uncertain | *5. Publication bias* | | | |
| | | Serious | (−1) | | |
| | | Very serious | (−2) | | |

Adapted from Balshem et al. [34]

[a]−2, −1, +1, and +2 indicate the number of levels to downgrade or upgrade from the current quality of evidence rating for each reason assessed

[b]Inconsistency (heterogeneity) judgment based on similarity of point estimates, overlap of confidence intervals, and statistical tests of heterogeneity

[c]Imprecision criterion based on the 95 % confidence interval (CI)

[d]Indirectness may be classed in four ways: Patients may differ from the population of interest, interventions may differ from the intervention of interest, outcomes may differ from those of primary interest or comparisons of interventions not tested in head-to-head trials

[e]Rating up one or even two levels is possible when effects in observational studies are sufficiently large, particularly if they occur over short periods of time

[f]A dose response gradient, or a conclusion that plausible residual confounding would further support inferences regarding treatment effect, may also raise the quality of the evidence

assumption of similarity, homogeneity across trials involved in the indirect comparison, the consistency assumption, interpretation, and reporting. The review showed that the underlying assumptions were not routinely explored or reported in network meta-analyses. More recently, Sobieraj and colleagues identified 34 closed-loop Bayesian MTC between 2006 and 2011 and summarized their characteristics regarding performance and reporting [38]. They concluded that although publication of Bayesian closed-loop MTC is increasing in frequency, details regarding methodology are often poorly described and efforts to clarify appropriate methods and reporting should be a priority. In the answer to the suggestion that the quality of reporting of NMA is suboptimal, Hutton and colleagues performed an overview of existing evaluations of quality of NMA and ITC between 2004 and 2013 [39]. Although only eight reports were included, several deficiencies in the current reporting of NMA were observed which reinforced the need to develop reporting guidelines. The authors propose to use the findings from this review to guide the development of new guidelines on the reporting of NMA in the format of an extension to PRISMA.

The National Institute for Health Care Excellence (NICE) Decision Support Unit has commissioned seven technical support documents (TSD) on evidence synthesis for decision making, including a reviewer checklist [40]. The checklist is intended for use with traditional pairwise meta-analysis, ITC, and NMA, without distinction, as the TSD series views NMA as an extension of pairwise meta-analysis with the assumptions of trial similarity and consistency in NMA being just properties of the identical/exchangeability requirement across all studies contributing to the relative treatment effects between a pair of treatments. Specifically the reviewer checklist is designed to inform the actual decision-making process rather than guide academic paper submissions or the summarization of treatment comparisons for a body of literature. It focuses on a series of questions that allow the reviewer to comment on whether the assumptions are reasonable, the research question has been addressed, the interpretation of the evidence is justified, or there is a need for further analyses or sensitivity analyses. The checklist comes in four sections:

1. Definition of the decision problem, with an emphasis on effect modifiers
2. Methods of analysis and presentation of results
3. Issues specific to network synthesis
4. Embedding the synthesis in a probabilistic cost-effectiveness analysis

Similar to other systematic review guidelines, it does not generate a quality rating; it is not prescriptive and requires the reviewer to make a series of judgments. But unlike other guidelines, it provides a framework for open discussion and assumes reviewers can ask for clarification, alternative and sensitivity analyses, details of search algorithms, computer code, and so forth.

More recently, the Academy of Managed Care Pharmacy (AMCP), the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), and the National Pharmaceutical Council (NPC) collaborated to improve health outcomes through new comparative evidence tools. In 2014 they published an ISPOR–AMCP–NCP Good Practice Task Force Report on an Indirect Treatment

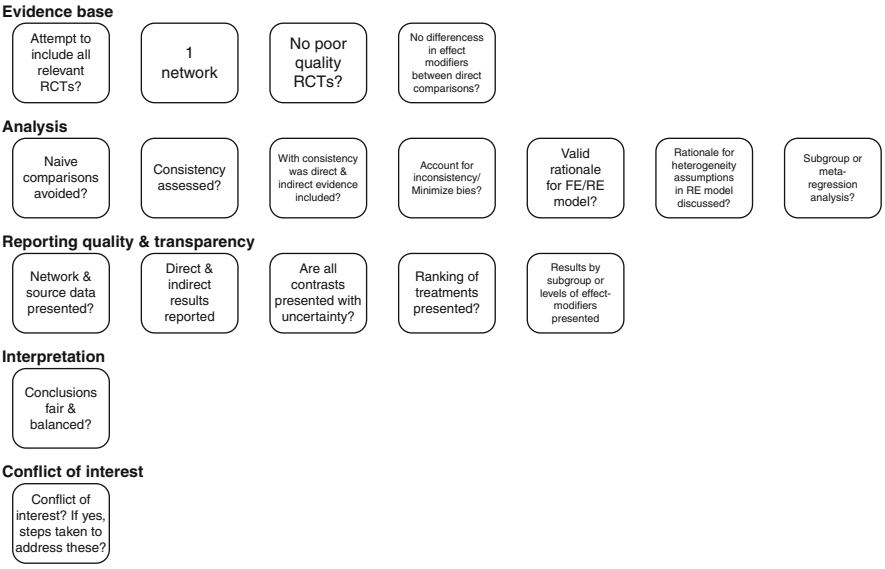



**Fig. 9.2** Assessment of credibility domains in a network meta-analysis. *FE* fixed effects, *RE* random effects, *RCTs* randomized controlled trials (Reproduced from Jansen et al. [41] with permission)

Comparison/Network Meta-Analysis Study Questionnaire to Assess Relevance and Credibility to Inform Health Care Decision Making [41].

The questionnaire is characterized by two principal concepts: relevance and credibility. Relevance reflects the extent to which the results of the NMA apply to the setting of interest and is captured by four questions:

1. Is the population relevant?
2. Are there any relevant interventions missing?
3. Are there any relevant outcomes missing?
4. Is the context (settings and circumstances) applicable?

Credibility is the extent to which the NMA provides accurate or valid answers to the questions it is designed to answer. It is assessed by 22 questions across five domains (Fig. 9.2), which address mainly quality and validity. Each question has detailed explanation and can be answered with "yes," "no," and "can't answer" (which is used to reflect insufficient report detail or assessor training). Based on these answers, a judgment about each credibility sub-domain can be made:

- Strength: NMA conducted well
- Neutral: potentially important concerns, unlikely to affect credibility
- Weakness: Likely to be biased and misleading
- Fatal flaw: Likely to be biased, conclusions misleading

A fatal flaw occurs when the method for ITC does not adjust for effect of study by taking the within-study treatment effect, but rather inappropriately combines results for a given treatment group across studies. Finally an overall judgment of the relevance and credibility of the NMA for decision making is opined as either sufficient or insufficient.

Salanti and colleagues proposed an approach to evaluate the quality of evidence from a NMA based on the GRADE working group methods for pairwise meta-analysis [42]. An NMA provides two principal types of findings for a specific outcome: (1) the relative treatment effect for all pairs of comparisons in the network and (2) the treatment ranking. The first one is derived from complex weighted averages of particular sources of direct and indirect evidence, whereas the second involves inferences about the network of evidence as a whole. The aim is to evaluate both of these outputs from an NMA with a framework that acknowledges the following: the importance of ITC, the contribution of each piece of direct evidence to the NMA, the transitivity assumption, consistency between direct and indirect evidence to obtain confidence in each (pairwise) effect size, and confidence in ranking.

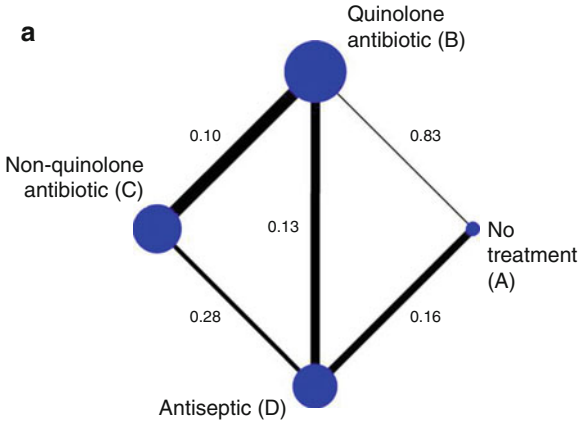
In extending GRADE to NMA, they take a two-staged approach. First, they consider each GRADE domain (reason to downgrade: study limitations [risk of bias], indirectness, inconsistency, imprecision, and publication bias) separately for the available direct comparisons in the network and for each pairwise network estimate the contribution of these direct estimates feeding into it. Second, they combine (not aggravate) the domain-specific judgments for each pairwise network comparison to obtain the overall quality of evidence. This approach requires assessment of the flow of evidence around the network and the contributions of each piece of direct evidence to the network effect estimates.

In the paper, a network of topical antibiotics is used as a working example and the contribution matrix is used to assess the information contribution of direct evidence to the NMA results (Fig. 9.3). In the contribution matrix the rows correspond to the NMA estimates and the columns correspond to direct meta-analysis estimates.

Each of the five GRADE domains is interpreted for assessment in NMA with a description of the procedure and instructions for downgrading. Table 9.6 gives a summary of this domain assessment for a specific pairwise effect and treatment ranking.

In the same timeframe a GRADE working group published a four-step approach for rating the quality of treatment effect estimates from NMA [43].

The first step is to present both direct and indirect treatment estimates for each comparison in the network. Direct comparisons can be obtained by classical pairwise meta-analysis techniques. There are several methods for estimating indirect comparisons including the Butcher method [44]. Node splitting is a method that will separate evidence for a particular comparison into both direct and indirect estimates [45].




**b**

| Network meta-analysis estimates | Direct comparisons in the network | | | | |
|---|---|---|---|---|---|
| | AB | AD | BC | BD | CD |
| Mixed estimates | | | | | |
| AB | 12.0 | 39.0 | 10.0 | 29.0 | 10.0 |
| AD | 12.3 | 72.3 | 3.1 | 9.1 | 3.2 |
| BC | 1.9 | 1.9 | 66.6 | 13.9 | 15.8 |
| BD | 6.8 | 6.8 | 17.6 | 51.2 | 17.6 |
| CD | 4.1 | 4.1 | 35.0 | 30.9 | 25.9 |
| Indirect setimates | | | | | |
| AC | 8.9 | 32.8 | 25.5 | 16.7 | 16.1 |
| **Entire network** | 8.0 | 26.6 | 25.3 | 25.0 | 15.1 |
| **Included studies** | 2 | 1 | 7 | 5 | 4 |

**Fig. 9.3** Network plot of topical antibiotics and contribution of direct comparisons. Panel (**a**): The size of the nodes represents the relative weight of the number of studies on each topical antibiotic. The *lines* (edges) connecting two nodes represent direct comparisons between the respective treatments with the thickness of the line weighted by the inverse of the variance (precision) of the summary statistic (here ln (odds ratio) presented along the edges). Panel (**b**): This shows the percentage contribution of each direct estimate to the network meta-analysis estimates. Rows correspond to network meta-analysis comparisons and columns correspond to direct meta-analysis comparisons. The "entire network" row shows the contribution of the direct comparisons to the total network of evidence that provides treatment ranking. The sizes of the boxes are proportional to the percentage contribution of each direct estimate to the relevant network meta-analysis row and the last row shows the number of studies included for each direct comparison (Reproduced from Salanti et al. [42])

**Table 9.6** Summary of domain assessment for evaluating the quality of evidence from a network meta-analysis

| GRADE domain | Domain assessment in NMA | Description of procedure | Instructions for downgrading |
|---|---|---|---|
| *Evaluate the confidence in a specific pairwise effect estimated in network meta-analysis* | | | |
| Study limitations | Study limitations | Determine which direct comparisons contribute to estimation of the NMA treatment effect[a] and integrate risk of bias assessments from these into a single judgment | Use standard GRADE considerations to inform judgment |
| Indirectness | Joint consideration of indirectness and intransitivity | Evaluate indirectness of populations, interventions, and outcomes as in standard GRADE. Evaluate transitivity by comparing the distribution of known effect modifiers across comparisons that contribute evidence to estimation of the NMA treatment effect[a] | If a priori assessment makes a transitivity assumption reasonable and suggests that effect modifiers are balanced, then do not downgrade. Otherwise, downgrade (either if a transitivity assumption does not look reasonable or if there is insufficient evidence to judge) |
| Inconsistency | Joint consideration of statistical heterogeneity and statistical inconsistency | (a) Judge the extent of heterogeneity, considering the comparison-specific heterogeneity variance, the NMA estimate of variance, a prediction interval and/or other relevant metrics such as $I^2$<br>(b) Evaluate the extent to which the comparison under evaluation is involved in inconsistent loops of evidence | (a) If important heterogeneity is found, downgrade. If heterogeneity is low, do not downgrade<br>(b) Power to detect inconsistency may be low; downgrade in absence of statistical evidence for inconsistency when direct and indirect estimates imply different clinical decisions |
| Imprecision | Imprecision | Focus on width of the confidence interval | Assess uncertainty around the pairwise estimate. Downgrade if confidence interval crosses null value or includes values favoring either treatment |

**Table 9.6** (continued)

| GRADE domain | Domain assessment in NMA | Description of procedure | Instructions for downgrading |
|---|---|---|---|
| Publication bias | Publication bias | Nonstatistical consideration of likelihood of non-publication of evidence that would inform the pairwise comparison. Plot pairwise estimates on contour-enhanced funnel plot | Use standard GRADE to inform judgment |
| *Evaluate the confidence in treatment ranking estimated in network meta-analysis* | | | |
| Study limitations | Study limitations | Integrate risk of bias assessments from each direct comparison to formulate a single overall confidence rating for treatment rankings[a] | Use standard GRADE considerations to inform judgment |
| Indirectness | Joint consideration of indirectness and intransitivity | Evaluate indirectness of populations, interventions, and outcomes as in standard GRADE. Evaluate transitivity across network by comparing the distribution of known effect modifiers across comparisons[a] | If a priori assessment of transitivity suggests effect modifiers are balanced across the network, do not downgrade. Otherwise, downgrade (either if a transitivity assumption does not look reasonable or if there is insufficient evidence to judge) |
| Inconsistency | Joint consideration of statistical heterogeneity and statistical inconsistency | (a) Judge the extent of heterogeneity considering primarily the NMA variance estimate(s) used and other network-wise metrics such as Q for heterogeneity in a network<br>(b) Evaluate inconsistency in network using statistical methods (such as global tests of inconsistency, or global inconsistency parameter) | (a) If important heterogeneity is found, downgrade. If heterogeneity is low do not downgrade.<br>(b) For overall treatment rankings, inconsistency should be given greater emphasis, since ranks are based on mean effects and the uncertainty they are estimated with. Downgrade in absence of statistical evidence for inconsistency when several direct and indirect estimates imply different clinical decisions |

(continued)

**Table 9.6** (continued)

| GRADE domain | Domain assessment in NMA | Description of procedure | Instructions for downgrading |
|---|---|---|---|
| Imprecision | Imprecision | Visually examine ranking probabilities (e.g., rankograms) for overlap to assess precision of treatment rankings | If probabilities are similarly distributed across the ranks, downgrade for imprecision |
| Publication bias | Publication bias | Nonstatistical consideration of likelihood of non-publication for each pairwise comparison. If appropriate, plot NMA estimates on a comparison adjusted funnel plot and assess asymmetry | As asymmetry does not provide concrete evidence of publication bias, downgrading should only be considered jointly with the nonstatistical assessment |

Reproduced from Salanti et al. [42]

[a]When integrating assessments about direct comparisons into a judgment about an NMA treatment effect or the ranking, more weight should be given to assessments from direct comparisons that contribute more information. We recommend use of the contributions matrix to quantify how much information each direct comparison contributes to the estimation of the NMA treatment effect under evaluation or the ranking

The second step is to rate the quality of each of these direct and indirect comparisons. The GRADE principles can be applied directly for outcomes where direct comparisons are available in head-to-head trials. Rating the quality of evidence for indirect comparisons can be difficult depending on the complexity of the network of evidence. To keep this indirect quality rating manageable, the working group suggest a focus on first-order loops–that is, where an indirect estimate of the relative treatment effect of two interventions can be made readily anchored on a third intervention (the common comparator). Rating the quality of evidence of the indirect comparison is based on applying the GRADE principles to the two contributing pairwise estimates with the anchor treatment serving as a common comparator between the two sets of comparisons.

In the third step the mixed treatment comparison estimates (both direct and indirect effects) are presented for the network.

The fourth and final step is to rate the quality of the NMA effect estimates. If only direct or only indirect information is available for a given comparison, the quality rating will be based on that information (determined in step 2). However for mixed treatment comparisons, which involve both direct and indirect evidence, the GRADE working group suggests using the higher of the 2 ratings (be it direct and indirect) for the comparison.

**Conclusions**

While formalized guidance on assessing systematic review quality when performing umbrella review is not currently available, standardized tools do exist for good methodological and clinical practice. Which specific tool to use depends on the circumstance of the umbrella review and the type of appraisal being performed. The PRISMA statement and AMSTAR tool remain the standards for reporting systematic reviews of RCTs and evaluating systematic review quality, respectively; the GRADE tool is also recommended to assess the strength of evidence within umbrella reviews. The expanding field of NMA continues to yield a number of tools to assess the quality of its systematic reviews. Future guidance documents on performance of umbrella reviews should give careful consideration to providing recommendations for formal risk of bias assessment as part of the evidence appraisal process.

## References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6: e1000097.
2. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org. Accessed October 28th, 2015.
3. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med. 2007;4:e78.
4. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. J Clin Epidemiol. 2012;65:1267–73.
5. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7:e49667.
6. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JAC. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
7. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282:1054–60.
8. Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, Santaguida PL, Shamliyan T, Singh K, Tsertsvadze A, Treadwell JR. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. Methods guide for effectiveness and comparative effectiveness reviews [Internet]. Rockville: Agency for Healthcare Research and Quality (US); 2008–2012.
9. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol. 2007;36:666–76.
10. Higgins JPT, Altman DG, Sterne, JAC, editors. Assessing risk of bias in included studies, chap. 8. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available from www.cochrane-handbook.org
11. Higgins JPT, Lane PW, Anagnostelis B, Anzures-Cabrera J, Baker NF, Cappelleri JC, Haughie S, Hollis S, Lewis SC, Moneuse P, Whitehead A. A tool to assess the quality of a meta-analysis. Res Syn Meth. 2013;4:351–66.

12. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolda DJM, Gavaughan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17:1–12.
13. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Last accessed 5 Mar 2015.
14. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52:377–84.
15. Fleming PS, Koletsi D, Seehra J, Pandis N. Systematic reviews published in higher impact clinical journals were of higher quality. J Clin Epidemiol. 2014;67:754–9.
16. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med. 2009;151:W-65–94.
17. Moher D, Shamseer L, Clarke M, Ghersi D, Petticrew M, Shekelle P, Steward LA. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1.
18. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;349:g7647.
19. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, Gøtzsche PC, Lasserson T, Tovey D, PRISMA for Abstracts Group. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. PLoS Med. 2013;10(4):e1001419. PMID: 23585737.
20. Welch V, Petticrew M, Tugwell P, Moher D, O'Neill J, Waters E, White H, PRISMA-Equity Bellagio group. PRISMA-equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity. PLoS Med. 2012;9(10):e1001333. PMID: 23222917.
21. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology. A proposal for reporting. JAMA. 2000;283:2008–12.
22. Sacks H, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med. 1987;316:450–5.
23. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44:1271–8.
24. Oxman AD. Checklists for review articles. BMJ. 1994;309:648–51.
25. Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. J Clin Epidemiol. 1996;49:235–43.
26. Assendelft WJ, Koes BW, Knipschild PG, et al. The relationship between methodological quality and conclusions in reviews of spinal manipulation. JAMA. 1995;274:1942–8.
27. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
28. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, Ramsay T, Bai A, Shukla VK, Grimshaw JM. External validation of a measurement tool to Assess Systematic Reviews (AMSTAR). PLoS One. 2007;2(12):e1350.
29. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009;62:1013–20.
30. Kung J, Chiapelli F, Cajulis OO, Avezova R, Kossan G, Chew L, Maida CA. From systematic reviews to clinical recommendations for evidence-based health care: Validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J. 2010;4:84–91.
31. Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. J Clin Epidemiol. 2015;68(5):574–83.

32. Guyatt GH, Oxman AD, Schunemann JH, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011;64:380–2.
33. Guyatt GH, Oxman AD, Kunz R, Vist GE, Brozek J, Norris S, et al. GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64:383–94.
34. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Yttr Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64;401–6. http://www.gradeworkinggroup.org/publications/JCE_series.htm.
35. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health. 2011;14:417–28.
36. Snedecor SJ, Patel DA, Cappelleri JC. From pairwise to network meta-analyses. In: Biando-Zoccai G, editor. Network meta-analysis: evidence synthesis with mixed treatment comparison. New York: Nova Science Publishers; 2014. p. 21–41.
37. Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect comparisons: a review of reporting and methodological quality. PLoS One. 2010;5(11):e11054.
38. Sobieraj DM, Cappelleri JC, Baker WL, Phung OJ, White CM, Coleman CI. Methods used to conduct and report Bayesian mixed treatment comparisons published in the medical literature: a systematic review. BMJ Open. 2013;3:e003111. doi:10.1136/bmjopen-2013-003111.
39. Hutton B, Salanti G, Chaimani A, Caldwell DM, Schmid C, Thorlund K, Mills E, Catalá-López F, Turner L, Altman DG, Moher D. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. Clifford T, ed. PLoS One. 2014;9(3):e92508. doi:10.1371/journal.pone.0092508.
40. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. Med Decis Making. 2013;33:679–91.
41. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andres S, Eldessouki R, Salanti G. Indirect treatment comparison/network meta-analysis study questionnaire to assess study relevance and credibility to inform healthcare decision-making: an ISPOR-AMCP-MPC good practice task force report. Value Health. 2014;17:157–73.
42. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014;9(7):e99682.
43. Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.
44. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50:683–91.
45. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29:932–44.

# Statistical Models for Overviews of Reviews

10

Tzu-Ting Chen and Yu-Kang Tu

**Abstract**

Since the rise of evidence-based medicine movement, systematic reviews and meta-analyses have been widely used for synthesis of evidence on beneficial and/or harmful effects of different treatments. Moreover, with the advances in medical science and knowledge, many new treatments and interventions become available, and identifying how best to compare multiple treatments is an important challenge to evidence-based medicine. Although network meta-analysis is a very powerful tool for comparing multiple treatments in terms of their benefits or harms, it requires a lot of resources and a substantial amount of time for literature search, data extraction and statistical analysis. If a decision based on currently available evidence needs to be made urgently, it might not be feasible to undertake network meta-analysis within a short period of time. Nevertheless, many traditional pairwise meta-analyses of a good quality may have been published, and there are great overlaps in literature search and data extractions between those pairwise meta-analyses and a network meta-analysis. If results of traditional meta-analysis can be used for an expedient comparison of multiple treatments, this would help researchers spend less time and resources in reaching a decision within a shorter period of time. Statistical models for an umbrella review are similar to those for a network meta-analysis, as they both aim to compare multiple treatments. Both Bayesian approach and generalized least-squares approach can be used to conduct statistical analysis for an umbrella review. This will be especially useful for policy makers or busy clinicians to obtain up-to-date evidence to make an informed decision.

T.-T. Chen • Y.-K. Tu (✉)
Department of Public Health, Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Room 539, No. 17, Xu-Zhou Road, Taipei 100, Taiwan
e-mail: yukangtu@ntu.edu.tw



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_10

## 10.1 Introduction

Since the rise of evidence-based medicine movement, systematic reviews and meta-analyses have been widely used for evaluation and assessment of evidence on beneficial and/or harmful effects of medical interventions. Whilst a systematic review may be broad in its scope by including multiple interventions in its comparisons of various interventions, traditional methodologies for meta-analyses can only make pairwise comparisons. With the advances in medical science and knowledge, many new treatments and interventions become available, and therefore how to compare multiple treatments simultaneously becomes an important challenge to evidence-based medicine to assist clinicians to make an informed decision on which treatment is the best for their patients. Traditional pairwise meta-analysis only uses direct evidence of head-to-head comparisons, but direct evidence is not always available. Even if direct comparisons can be made, results from different pairwise comparisons may not be consistent. For instance, one pairwise meta-analysis may show treatment *A* is better than treatment *B*, and another meta-analysis shows treatment *B* is better than treatment *C*; however, a third meta-analysis may fail to show *A* is better *C*. This scenario of inconsistency across pairwise meta-analyses can occur as different meta-analyses use different head-to-head trials when making the comparisons.

In the last decade, a new methodology for multiple treatments comparisons, known as network meta-analysis, has emerged and quickly gained great popularity. Compared to the traditional pairwise meta-analysis, the advantage of network meta-analysis is that it compares more than two treatments simultaneously in the same statistical model by using direct and indirect evidence.

Although the idea of indirect comparisons first emerged in the 1990s [1, 2], the concept of network meta-analysis was first proposed by Lumley in 2002 [3], and it provides a unified statistical framework for multiple treatment comparisons. A few years later, a Bayesian statistical approach to network meta-analysis was developed by Lu, Ades and their colleagues, and their approach is now known as mixed treatment comparison [4–9]. This Bayesian approach is now considered by many the standard approach [10–16]. Bayesian network meta-analysis has been received with great enthusiasm, and the number of studies using this method has increased dramatically in recent years [17]. Many systematic reviews with a network meta-analysis have been published in top general medical journals, e.g. *Lancet*, *Journal of the American Medical Association*, *British Medical Journal*, etc. We have also seen more and more papers with a network meta-analysis appear in specialist journals.

To compare multiple treatments by doing a network meta-analysis, researchers start with a comprehensive systematic review by conducting a literature search evaluation of studies and extraction of data. As more treatments are involved in the comparison, it takes more time and resources to complete the systematic review. However, for treatments involved in the network meta-analysis, it is very likely that many pairwise comparisons may have been undertaken and published, but those previous efforts generally contribute little to the network meta-analysis. Nevertheless,

it seems to be a waste of time and resources if many of the tasks such as literature search, studies screening and selection, quality assessment and data extraction are repeated. When researchers have limited resources and feel the time pressure to complete the multiple treatment comparisons, it can greatly save the resources and time, if results from previous systematic review with traditional pairwise meta-analyses could be used. This will be especially useful for policy makers or busy clinicians who can obtain results much quickly to make an informed decision.

For instance, let us imagine that a clinician wants to know which treatment is the best one for patients with a certain medical condition, and this clinician tries to obtain evidence from systematic reviews to make a decision. When there are several treatments to be compared and no network meta-analysis has been published, it would be unrealistic for this clinician to undertake a systematic review and network meta-analysis with colleagues, as the process may take a long time to complete. In contrast, if undertaking an overview of reviews could make the process of comparing multiple treatments much faster, and its results are similar to those from a complete network meta-analysis, the development of methods for undertaking a rapid review will become a powerful tool for decision making in evidence-based medicine.

Umbrella reviews have been developed in the past few years and initially focused on evaluating the quality of meta-analyses of related topics [18]. In 2009, Ioannidis discussed umbrella reviews which enable the quantitative analysis of trial networks using data from clinical trials on various interventions for the same disease or condition and described methods that synthesize evidence from multiple meta-analyses [19]. This paper also discussed how to apply the network meta-analysis method, especially with multiple treatments comparison, in an umbrella review. It also used a figure to illustrate the similarity between network meta-analyses and umbrella reviews. In addition, Ioannidis also discussed the potential biases when undertaking an umbrella reviews.

In 2010, Caldwell et al. gave an example for how to apply the network meta-analysis model for an umbrella review [20]. Firstly, search the literatures for pairwise comparison in specific condition and extract the results, including the results of fixed-effect model and random-effect model, from each pairwise meta-analysis. Secondly, assess the consistency between direct and indirect effect before proceeding to the network meta-analysis. It can be done by testing all loops together or testing each loop separately. Then, do the network meta-analysis twice: one uses the results from the fixed-effect model and the other uses results from the random-effect model reported by traditional pairwise comparisons. The meta-analysis in the umbrella review uses the point estimates and variances reported by pairwise meta-analyses, i.e. only one observation for each pairwise comparison, and therefore only the fixed-effect model for multiple treatment comparisons can be undertaken in umbrella review.

In the next sections, we first introduce the Bayesian method proposed by Caldwell et al. for evidence synthesis, and we then explain how this Bayesian model can be implemented in a frequentist statistical framework and be undertaken using standard software packages. We will use an example dataset to demonstrate these methods.

## 10.2 Basic Model

Statistical models for umbrella reviews are similar to those for the network meta-analysis, as they both aim to compare multiple treatments. Network meta-analysis synthesizes evidence from individual studies, whilst umbrella review results from traditional pairwise meta-analyses to undertake multiple treatment comparisons. In an umbrella review, each treatment contrast has only one data entry, i.e. the point estimate and its variance from a previous pairwise meta-analysis, and therefore only fixed-effect modelling is feasible, although the point estimate and variance may come from a previous fixed- or random-effect meta-analysis. Consequently, umbrella review has to use the trial-based/contrast-based model for statistical analysis. Network meta-analysis uses individual studies that compare at least two of all the treatments involved in the comparison, and therefore it can use fixed- or random-effect arm-based or trial-based model [21, 22]. As network meta-analysis involves more treatments and studies than traditional meta-analysis, substantial heterogeneity is expected and random-effect model is usually the preferred approach.

Nonetheless, umbrella review makes similar assumptions about multiple treatment comparisons as network meta-analysis does. The first is the homogeneity assumption that all trials for comparing treatment *X* to *Y* are similar for each possible treatment contrast *XY*. The second is the similarity assumption that factors that affect the response to any treatment contrast *XY* must be similarly distributed in other treatment contrasts in the network. These two assumptions yield consistency between direct and indirect comparisons within the network meta-analysis. For instance, the treatment effect *dBC* estimated by head-to-head trials comparing treatment *B* to *C* ($dBC^{\text{dir}}$) would be the same as the treatment effect indirectly estimated by the *A*–*B* and *A*–*C* trials ($dBC^{\text{ind}}$), i.e. $d_{BC}^{\text{dir}} = d_{BC}^{\text{ind}} = d_{AC}^{\text{dir}} - d_{AB}^{\text{dir}}$.

For a umbrella review that involves treatments *A*, *B*, *C*, …, *K*, the statistical model can be written as

$$\eta_{jbk} = d_{bk} + v_j, \ (\text{Model} - 1)$$
$$d_{bk} = d_{Ak} - d_{Ab}$$

where $\eta_j$ is the reported treatment contrast between baseline treatment *b* and test treatment *k* from the *j*th pairwise meta-analysis comparing *b* to *k*, $b = A, B, C, \ldots, J$, and $k = B, C, \ldots, K$, and $v_j$ is the standard error of $\eta_{jbk}$, $d_{bk}$ is the estimated treatment contrast between treatment *b* and the *b* in the pairwise meta-analysis, $d_{Ak}$ is the difference between the global baseline treatments *A* and *k* and $d_{Ab}$ is the difference between treatments *A* and *b*; $d_{bk}$ is equal to $d_{Ak} - d_{Ab}$.

## 10.3 Bayesian Approach to Statistical Analysis of Umbrella Review

The standard fixed-effect Bayesian network meta-analysis for multiple treatment comparisons proposed by Lu and Ades may be specified as

$$\text{Model}: \theta_{jk} = \begin{cases} \mu_{jb} & b = A,B,\ldots,J \quad \text{if } k = b \\ \mu_{jb} + d_{bk} = \mu_{jb} + d_{Ak} - d_{Ab} & k = B,C,\ldots,K, \quad \text{if } k \text{ is after } b \end{cases}$$

$$d_{AA} = 0$$

$$\text{Prior}: d_{AB} \sim N\left(0, \sigma^2\right), d_{AC} \sim N\left(0, \sigma^2\right), \ldots, d_{AK} \sim N\left(0, \sigma^2\right)$$

where $\theta_{jk}$ is the outcome for treatment $k$ in study $j$ and $\mu_{jb}$ is the baseline treatment effect in study $j$. The difference between treatment $k$ and treatment $b$ in the same trial is estimated by expressing them in terms of effects relative to treatment $A$, which is the global baseline treatment within the network. For reasons of identification and because of its interpretation as the effect of treatment $A$ compared to itself, $d_{AA}$ is fixed at 0. Lu and Ades called $d_{AB}$ to $d_{AK}$ the basic parameters. The advantage of expressing all treatment contrasts as relations between basic parameters is that the number of pairwise comparisons to be estimated for a network meta-analysis involving $K$ treatments is reduced to $K-1$ for the fixed effects. Usually, a noninformative prior is used in the network meta-analysis by specifying a large $\sigma^2$.

In the Bayesian approach, all treatments can be ranked by the probability of being the most effective for the outcome. The probability was calculated by the number of times being the best treatment divided by the number of iterations for each treatment. The ranking may be useful for making decisions on the recommendation of treatments, especially when the differences in the outcome among treatments are not statistically significant

For the umbrella review, the Bayesian statistical model for multiple treatment comparison can be written as

$$\text{Model}: \eta_{jbk} = d_{bk} + v_j = d_{Ak} - d_{Ab} + v_j$$

$$d_{AA} = 0$$

$$\text{Prior}: d_{AB} \sim N\left(0, \sigma^2\right), d_{AC} \sim N\left(0, \sigma^2\right), \ldots, d_{AK} \sim N\left(0, \sigma^2\right)$$

where $\eta_{jbk}$ is the results from the $j$th pairwise meta-analysis of studies that compares treatment $b$ to treatment $k$. The difference between treatment $k$ and treatment

$b$ in the same trial is expressed as a linear combination of basic parameters $d_{AB}$ to $d_{AK}$. For identification reason, $d_{AA}$ is fixed at 0. A noninformative prior is used in the analysis by specifying a large $\sigma^2$.

It is convenient to conduct a Bayesian approach to umbrella review using the free WinBUGS software package, and example WinBUGS code is available in the article by Caldwell et al. [20]. In addition, the umbrella review also can be analyzed by whichever software contained the Bayesian statistical methods, like SAS or R software.

## 10.4 Generalized Least-Squares Approach to Umbrella Review

As explained in the previous section, umbrella review can only undertake fixed-effect analysis for multiple treatment comparisons, because each treatment contrast has only one data entry. In the frequentist statistical framework, a fixed-effect analysis can be undertaken using generalized least-squares regression, and equation for Model-1 may be written as follows:

$$Y_{jbk} = b_1 t_{AB} + b_2 t_{AC} + \ldots + b_{K-1} t_{AK} + v_j \tag{10.1}$$

where $Y_{jbk}$ is the result, such as log odds ratio or log relative risk ratio, from the $j$th pairwise meta-analysis that compare treatment $b$ to $k$; $b_1$ to $b_{K-1}$ are regression coefficients for $K-1$ treatment contrasts $A$ versus $B$ ($t_{AB}$), $A$ versus $C$ ($t_{AC}$),…, and $A$ versus $K$ ($t_{AK}$) in the network; and $v_j$ is the standard error for $Y_{jbk}$. For instance, suppose that there are five treatments, treatments $A$ to $E$, included in an umbrella review, and all possible contrasts are presented in Table 10.1.

Equation 10.1 is not an ordinal least square in which the error terms $v_j$ are independent identically distributed. Instead, the variances of error terms are different in each included pairwise meta-analysis. As $v_j$ is different across different pairwise meta-analyses, this model is known as weighted least-squares regression which is a special case of the generalized least squares. If all the pairwise meta-analyses in an

**Table 10.1** All possible contrasts for five treatments $A$, $B$, $C$, $D$ and $E$

| Treatment | Control | $t_{AA}$ | $t_{AB}$ | $t_{AC}$ | $t_{AD}$ | $t_{AE}$ |
|---|---|---|---|---|---|---|
| $B$ | $A$ | −1 | 1 | 0 | 0 | 0 |
| $C$ | $A$ | −1 | 0 | 1 | 0 | 0 |
| $D$ | $A$ | −1 | 0 | 0 | 1 | 0 |
| $E$ | $A$ | −1 | 0 | 0 | 0 | 1 |
| $C$ | $B$ | 0 | −1 | 1 | 0 | 0 |
| $D$ | $B$ | 0 | −1 | 0 | 1 | 0 |
| $E$ | $B$ | 0 | −1 | 0 | 0 | 1 |
| $D$ | $C$ | 0 | 0 | −1 | 1 | 0 |
| $E$ | $C$ | 0 | 0 | −1 | 0 | 1 |
| $E$ | $D$ | 0 | 0 | 0 | −1 | 1 |

umbrella review included completely different sets of trials, $v_j$ are independent. However, if some trials compare more than two treatments, they may appear more than once in the pairwise meta-analyses. For instance, let us suppose a trial compare treatments *A*, *B* and *C*, and this trial will therefore be included in the pairwise meta-analysis on *A* versus *B*, on *A* versus *C* and on *B* versus *C*. Consequently, $Y_{jbk}$ is no longer strictly independent, and some $v_j$ are correlated. However, meta-analysts conducting an umbrella would not be able to know the extent of correlations, and this will give rise differences and bias in the results compared to a network meta-analysis based on individual studies.

The umbrella review analysis in Eq. 10.1 can be conducted using the software packages which implement the method of weighted least squares with the option to specify fixed values for the residual error terms. In the next section, we use Stata (version 13.1, StataCorp, College Station, TX, USA) to demonstrate an example of umbrella review. Stata commands vwls and gllamm can be used to undertake weighted least-squares regression for umbrella review [23, 24].

## 10.5 Practical Example: The Effects of Topical Antibiotics for Chronically Discharging Ears with Underlying Eardrum Perforations

In this example, we use data from a network meta-analysis on the effects of topical antibiotics or antiseptics for chronically discharging ears with underlying eardrum perforations by Macfadyen et al. [25]. The network plot is shown in Fig. 10.1. Each line in the network represents that at least one study reports the head-to-head comparison between the connected treatments, and the number next to the line was the number of trials reported the direct pairwise comparison. The following four treatments are compared in the analysis: no treatment, quinolone antibiotics, nonquinolone antibiotics and antiseptics, and coded 1–4 for convenience.

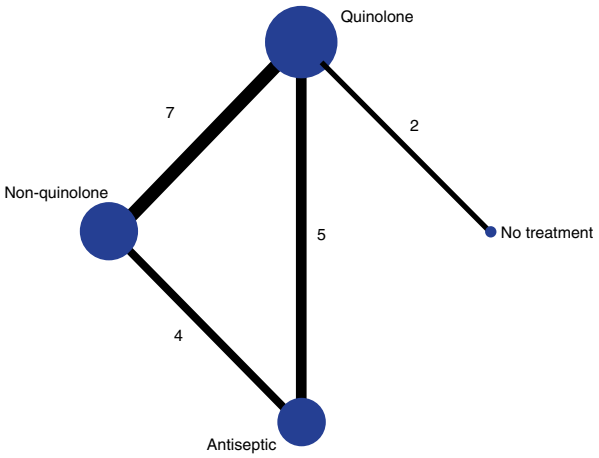



**Fig. 10.1** Network of evidence for the treatments of chronically discharging ears

**Table 10.2** Pooled treatment effects and heterogeneity for all pairwise comparisons reported in the umbrella review

| Treatment | Control | RCTs | LOR (S.E.) | | Heterogeneity | | $I^2$ (%) |
|---|---|---|---|---|---|---|---|
| | | | FE | RE | Cochran's $Q$ | $P$-value | |
| Quinolone [2] | No treatment [1] | 2 | −1.99 (0.30) | −2.53 (0.93) | 6.61 | <0.001 | 77.8 |
| Non-quinolone [3] | Quinolone [2] | 7 | 0.28 (0.12) | 0.41 (0.30) | 24.05 | 0.001 | 75.1 |
| Antiseptic [4] | Quinolone [2] | 5 | 1.14 (0.14) | 1.45 (0.45) | 24.86 | <0.001 | 83.9 |
| Antiseptic [4] | Non-quinolone [3] | 4 | 0.83 (0.24) | 1.01 (0.94) | 36.81 | <0.001 | 91.9 |

*Abbreviations*: *LOR* log odds ratio, *S.E.* standard error, *FE* fixed effect, *RE* random effects, *RCT* randomized control trial

The example dataset contains the results from 13 randomised clinical trials; three of those 13 trials are three-arm trial. We first undertook a traditional pairwise comparison meta-analysis as there is more than one randomised clinical trial for the four pairs of direct comparisons. The results of direct comparisons are reported in Table 10.2, showing the treatment and control arms, the number of studies, the log odds ratio and 95 % confidence interval of fixed-effect model and random-effect model, heterogeneity in Cochran's $Q$, $P$-value and $I^2$.

## 10.6 Inconsistency Within a Umbrella Review

Inconsistency between direct and indirect evidence within a network meta-analysis occurs, when three or more treatments formed a closed loop in the network. For instance, Fig. 10.1 shows one loop formed by quinolone, non-quinolone antibiotics and antiseptics. There is one potential inconsistency within this loop, such as the direct comparison between quinolone and non-quinolone antibiotics and the indirect comparison using antiseptics as the reference treatment. For our example data, we use the method by Bucher to evaluate the loop inconsistency [2]. Suppose three treatments $A$, $B$ and $C$ form a loop, and we choose to evaluate the inconsistency in treatment contrast $BC$. Note that results would not be affected by choosing other two treatment contrasts. The results from the direct evidence is expressed as $\hat{d}_{BC}^{\text{Dir}}$ and indirect evidence through the results of $AB$ and $AC$ which is represented by $\hat{d}_{BC}^{\text{Ind}}$. In theory, $\hat{d}_{BC}^{\text{Dir}}$ and $\hat{d}_{BC}^{\text{Ind}}$ should be equivalent, i.e. $\hat{d}_{BC}^{\text{Dir}} - \hat{d}_{BC}^{\text{Ind}} = 0$. In reality, there may be small differences between $\hat{d}_{BC}^{\text{Dir}}$ and $\hat{d}_{BC}^{\text{Ind}}$. The test of consistency is undertaken as follows:

1. Let $\omega_{BC} = \hat{d}_{BC}^{\text{Dir}} - \hat{d}_{BC}^{\text{Ind}}$.
2. $H_0$: $\omega_{BC} = 0$.
3. Calculate the variance of $\text{Var}\left(\omega_{BC}\right) = \text{Var}\left(\hat{d}_{BC}^{\text{Dir}}\right) + \text{Var}\left(\hat{d}_{BC}^{\text{Ind}}\right)$.

4. Then, the statistic is $Z_{BC} = \dfrac{\omega_{BC}}{\sqrt{\mathrm{Var}\left(\omega_{BC}\right)}}$, where ZBC follows the standard normal distribution.

If the statistic $Z_{BC}$ is large, the null hypothesis is then rejected, indicating that the inconsistency may exist.

In our example, the treatments quinolone antibiotic, non-quinolone antibiotic and antiseptic form a loop, and the results of inconsistency test were presented in Table 10.3 which showed that using summary statistics extracted from pairwise fixed-effect meta-analyses seems to find some evidence of inconsistency, because the *P*-value is smaller than 0.05. In contrast, using summary statistics extracted from pairwise random-effect meta-analyses does not find evidence of inconsistency. Although the difference in log odds ratios between direct and indirect comparisons is actually larger in the analysis using extracted information from pairwise random-effect meta-analyses, their variances are also much larger. This example demonstrates evaluating inconsistency is not always straightforward in a network meta-analysis, as many factors may affect the results.

## 10.7 Results from Bayesian Analysis

We use software package WinBUGS to undertake the analyses with pooled estimates and variances extracted from fixed- or random-effect pairwise meta-analysis and choose the appropriate model according to model deviance. For the Bayesian analysis, we set up 100,000 iterations with the first 50,000 iterations as burn-in. The posterior mean of summed deviance contributions is 2.99 and 2.98 for models with the use of the fixed-effect and random-effect summary estimates, respectively. The closer the posterior mean of summed model deviance is to the number of data, the better the model fit. There are four data points in our example on treatments of chronically discharging ear; the posterior means of summed model deviance therefore suggest a good model fit for both analyses. The results from the analyses for umbrella review are shown in Table 10.4. There are small differences in the results

**Table 10.3** Comparison of the direct and indirect estimates (LOR) of the effect of quinolone antibiotic relative to non-quinolone antibiotic

| Quinolone vs. non-quinolone | $\hat{d}$ | S.E. | Inconsistency $\omega_{BC} = \hat{d}_{BC}^{\mathrm{Dir}} - \hat{d}_{BC}^{\mathrm{Ind}}$ | Z-statistic | *P*-value |
|---|---|---|---|---|---|
| Based on FE summary estimates from pairwise comparison | | | | | |
| Direct | 0.28 | 0.24 | 0.59 | 1.94 | 0.0261 |
| Indirect | −0.31 | 0.08 | | | |
| Based on RE summary estimates from pairwise comparison | | | | | |
| Direct | 0.41 | 0.30 | 0.85 | 0.79 | 0.2155 |
| Indirect | −0.44 | 1.09 | | | |

*Abbreviations*: *LOR* log odds ratio, *S.E.* standard error, *FE* fixed effect, *RE* random effects

between models based on summary statistics extracted from fixed- or random-effect pairwise meta-analysis, although results from the model based on data extracted from random-effect pairwise meta-analyses show greater credible intervals. Both analyses suggest quinolone seems to be the most effect treatment. Table 10.5 shows the odds ratio of the three active treatments to observation only and the probability of each treatment being the most effective for persistent ear discharge. Quinolone has the greatest probability of being ranked the best treatment in both analyses with no differences. In addition, full ranking between these four treatments in the data extracted from fixed- and random-effect pairwise meta-analysis was presented in Figs. 10.2 and 10.3, separately.

**Table 10.4** Analysis of umbrella review in both fixed-effect and random-effect summary estimates in Bayesian approach method

| | | Fixed-effect summaries | | Random-effect summaries | |
|---|---|---|---|---|---|
| Treatment | Control | OR (95 % CrI) | Posterior mean $Dev_{XY}$ | OR (95 % CrI) | Posterior mean $Dev_{XY}$ |
| Quinolone [2] | No treatment [1] | 0.14 (0.08, 0.24) | 0.97 | 0.08 (0.01, 0.47) | 0.98 |
| Non-quinolone [3] | No treatment [1] | 0.18 (0.10, 0.34) | – | 0.12 (0.02, 0.78) | – |
| Antiseptic [4] | No treatment [1] | 0.42 (0.23, 0.79) | – | 0.33 (0.05, 2.34) | – |
| Non-quinolone [3] | Quinolone [2] | 1.33 (1.07, 1.66) | 0.83 | 1.52 (0.87, 2.64) | 0.92 |
| Antiseptic [4] | Quinolone [2] | 3.10 (2.44, 3.96) | 0.80 | 4.24 (1.92, 9.46) | 0.83 |
| Antiseptic [4] | Non-quinolone [3] | 2.34 (1.75, 3.12) | 0.38 | 2.80 (1.13, 7.00) | 0.24 |
| $\bar{D}$ | | | 2.99 | | 2.98 |

*Abbreviations*: $\bar{D}$ posterior mean of summed deviance contributions, *Dev* deviance, *CrI* credible interval

**Table 10.5** Posterior median OR (95 % CrI) of each treatment relative to no treatment and probability that each treatment is the most effective for outcome persistent discharge at the end of observation in Bayesian approach method

| | Fixed effect | | Random effect | |
|---|---|---|---|---|
| Treatment | Probability (best) | OR relative to no treatment (95 % CrI) | Probability (best) | OR relative to no treatment (95 % CrI) |
| No treatment [1] | 0 | 1 | 0.0021 | 1 |
| Quinolone [2] | 0.9942 | 0.14 (0.08, 0.24) | 0.9257 | 0.08 (0.01, 0.47) |
| Non-quinolone [3] | 0.0058 | 0.18 (0.10, 0.34) | 0.0721 | 0.12 (0.02, 0.78) |
| Antiseptic [4] | 0 | 0.42 (0.23, 0.79) | 0.0001 | 0.33 (0.05, 2.34) |

*Abbreviations*: *OR* odds ratio, *CrI* credible intervals

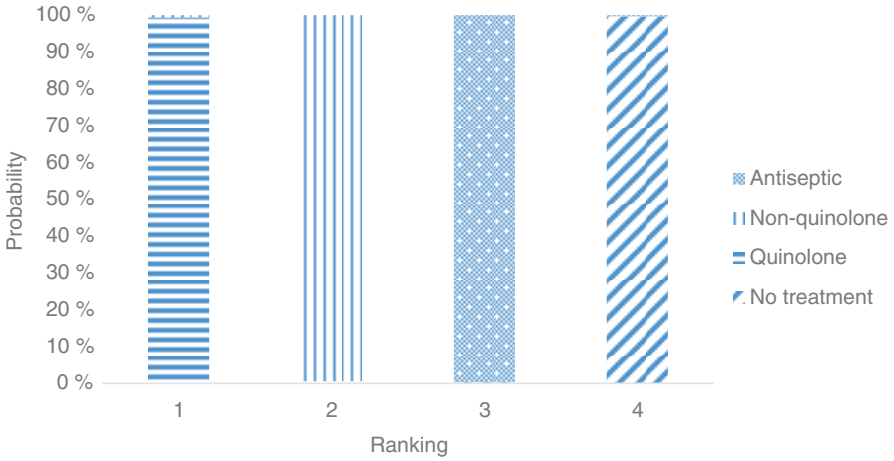



**Fig. 10.2** Full ranking of treatments in the data extracted from fixed-effect pairwise meta-analysis

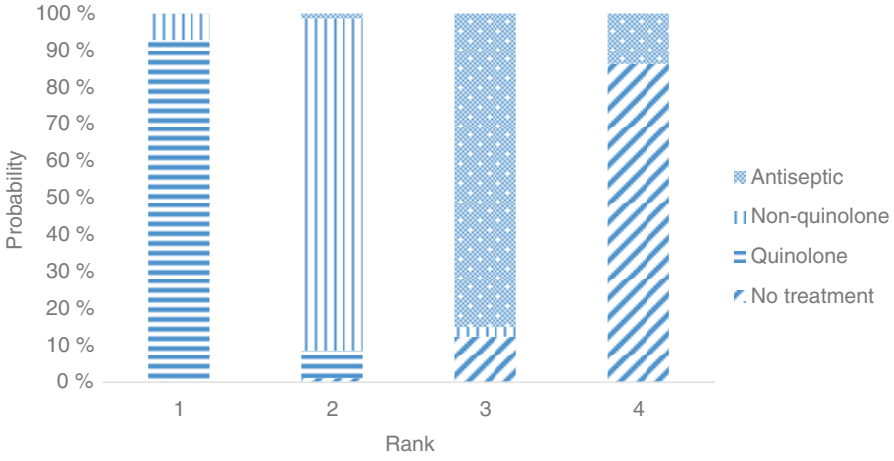



**Fig. 10.3** Full ranking of treatments in the data extracted from random-effect pairwise meta-analysis

## 10.8 Results from Generalized Least-Squares Analysis

The regression model for our example can be written as

$$Y_{jbk} = b_1 t_{12} + b_2 t_{13} + b_3 t_{14} + v_{ij},$$

where $Y_{jbk}$ is the log odds ratio for treatment $b$ versus $k$ from the $j$th pairwise meta-analysis in the umbrella review; $t_{12}$ to $t_{14}$ are the dummy variables with contrast coding for treatment 1 versus 2, treatment 1 versus 3 and treatment 1 versus 4 in the

network; $b_1$ to $b_3$ are their regression coefficients; and *vj* is the standard error for $Y_{jbk}$. Analysing continuous outcomes for network meta-analysis by means of generalized least-squares methods or linear mixed models can be a challenge, as it requires statistical software packages to specification of special patterns of model error variance and covariance structure. In our previous study, we show how the arm-based and contrast-based models can be implemented and analysed by means of commercial software packages such as Stata [22]. As umbrella review uses the summary data from pairwise meta-analysis, so the trial-based model is specified and the Stata commands vwls and gllamm can be used to undertake the analysis.

The contrast coding for our example is presented in Table 10.6, and results from the umbrella review using the generalized least-squares regression is presented in Table 10.7. The results of generalized least squares method are very similar to those from Bayesian analysis, and quinolone appears to be the most effective treatment followed by non-quinolone antibiotics, antiseptics and no treatment.

## 10.9 Network Meta-analysis of Example Data

To explore the potential bias in umbrella review, we also undertake the Bayesian network meta-analysis of example data with the use of individual studies. Some differences are to be expected, since three studies have three arms. The results of fixed-effect and random-effect network meta-analysis with the original 13 trials are shown in Table 10.8. The treatment 2, quinolone antibiotics, is still the most effective treatment among the four treatments, although the odds ratios and the

**Table 10.6** Contrast coding for four pairwise comparisons of the effects of topical antibiotics for chronically discharging ears with underlying eardrum perforations

| Treatment | Control | $t_{11}$ | $t_{12}$ | $t_{13}$ | $t_{14}$ |
|---|---|---|---|---|---|
| Quinolone [2] | No treatment [1] | −1 | 1 | 0 | 0 |
| Non-quinolone [3] | Quinolone [2] | 0 | −1 | 1 | 0 |
| Antiseptic [4] | Quinolone [2] | 0 | −1 | 0 | 1 |
| Antiseptic [4] | Non-quinolone [3] | 0 | 0 | −1 | 1 |

**Table 10.7** OR (95 % CI) of each active treatment relative to no treatment for outcome persistent discharge at the end of observation in non-Bayesian approach method

| | Fixed-effect summaries | Random-effect summaries |
|---|---|---|
| Treatment | OR relative to no treatment (95 % CI) | OR relative to no treatment (95 % CI) |
| No treatment [1] | 1 | 1 |
| Quinolone [2] | 0.14 (0.08, 0.25) | 0.08 (0.01, 0.49) |
| Non-quinolone [3] | 0.18 (0.10, 0.34) | 0.12 (0.02, 0.81) |
| Antiseptic [4] | 0.42 (0.22, 0.80) | 0.34 (0.05, 2.47) |

*Abbreviations*: *OR* odds ratio, *CrI* credible intervals

**Table 10.8** Posterior median OR (95 % CrI) of each treatment relative to no treatment and probability that each treatment is the most effective for outcome persistent discharge at the end of observation in network meta-analysis

| | Fixed effect | | Random effect | |
|---|---|---|---|---|
| Treatment | Probability (best) | OR relative to no treatment (95 % CrI) | Probability (best) | OR relative to no treatment (95 % CrI) |
| No treatment [1] | 0 | 1 | 0.0085 | 1 |
| Quinolone [2] | 0.9808 | 0.18 (0.09–0.32) | 0.8841 | 0.11 (0.01–0.66) |
| Non-quinolone [3] | 0.0192 | 0.26 (0.13–0.52) | 0.1040 | 0.19 (0.02–1.45) |
| Antiseptic [4] | 0 | 0.64 (0.34–1.20) | 0.0034 | 0.48 (0.06–3.38) |

*Abbreviations*: *OR* odds ratio, *CrI* credible intervals

probability of being the most effective are slightly different from those of umbrella review analysis. Those differences are likely due to dependency within the data extracted from the pairwise meta-analysis caused by the multi-arm trials that is ignored in the statistical analysis for the umbrella review. For example, a three-arm trial comparing treatments *A*, *B* and *C* will be included in three pairwise comparisons: *A* versus *B*, *A* versus *C* and *B* versus *C*. Hence, the correlations between the pooled estimates of these pairwise comparisons might exist. In our example data, three of thirteen trials included in ear discharge study are multi-arm trials, so this may explain the small differences in the results between umbrella review and network meta-analysis.

To explore the impact of multi-arm trials in umbrella review, we then exclude the three multi-arm trials from the analysis and compare the differences in the results between umbrella review and network meta-analysis. The results of umbrella review and network meta-analysis based on two-arm trials only are shown in Table 10.9, and their differences are relative smaller compared to those in which the multi-arm trials are included in the analysis. This indicates that ignoring the data dependency within an umbrella review could be a major source of bias and adjusting for this dependency needs to be considered.

All the WinBUGS and Stata codes for our analyses can be found in the appendix at the end of this chapter.

## 10.10 Discussion

When we have limited resources and are under the time pressure to make a decision, umbrella review may be an efficient tool to provide an overview of current evidence. For statistical analysis in umbrella review, Bayesian approach provides a flexible modelling framework, but it is less accessible to non-statisticians. Generalized least-squares method can be undertaken using standard software packages such as Stata and SAS that are more accessible and gives rise to similar results to those from Bayesian analysis.

**Table 10.9** OR (95 % CI) of each treatment relative to no treatment of umbrella review and network meta-analysis in study without multi-arm trials

| | Umbrella review | | Network meta-analysis |
|---|---|---|---|
| Treatment | Bayesian approach | Non-Bayesian approach | Bayesian approach |
| No treatment [1] | 1 | 1 | 1 |
| Quinolone [2] | 0.03 (0.01, 0.12) | 0.03 (0.01, 0.13) | 0.02 (0.00, 0.14) |
| Non-quinolone [3] | 0.04 (0.01, 0.19) | 0.04 (0.01, 0.21) | 0.04 (0.00, 0.30) |
| Antiseptic [4] | 0.05 (0.01, 0.25) | 0.05 (0.01, 0.27) | 0.04 (0.00, 0.31) |

*Abbreviations*: *OR* odds ratio, *CrI* credible intervals

For umbrella review to provide unbiased evaluation of multiple treatment comparisons, a few important issues need to be resolved. In theory, results from umbrella review of pairwise meta-analyses should be the same or very similar to those from a standard network meta-analysis of same individual studies that have been used to conduct those pairwise meta-analyses. Nevertheless, as explained in the previous section, results from these two approaches can be different, when multi-arm trials are involved. The greater the number of multi-arm trials involved in the network, the greater the differences in the results between umbrella review and network meta-analysis.

To overcome this limitation, one solution is to undertake statistical analysis using individual trials of those pairwise meta-analyses rather than their pooled estimates. This is feasible, if all the pairwise meta-analyses published sufficient information of the included trials. This approach is equivalent to a standard network meta-analysis but without undertaking a new, comprehensive literature search. Multi-arm trials would therefore be account for properly, and the potential bias will now arise in that the literature search in those pairwise meta-analyses has not been updated. If information of individual trials is not available for all studies, researchers have to retrieve the full text of those studies and carry out data extraction; hopefully, the number of those studies would be small. Nevertheless, we can image that in some scenarios where the number of treatments and trials involved in the comparisons is large and many pairwise meta-analyses do not provide sufficient information, to retrieve full text and carry out data extraction could take a lot of time and efforts. Alternatively, we may try to adjust for the data dependency caused by multi-arm trials by imputing a small correlation between the pooled estimates extracted from pairwise meta-analyses that involve those multi-arm trials. A sensitivity analysis can also be undertaken by imputing different correlation coefficients to obtain the range for the possible biases. This would however use statistical software packages that provide flexible ways to specify the variance-covariance structure, and at the moment, advanced software programming skills would be required to accomplish this task.

In summary, umbrella review provides a simple and efficient tool to undertake evidence synthesis for multiple treatment comparisons. Both Bayesian approach and non-Bayesian approaches have been developed and yield similar results. However, researchers need to be aware of the potential biases caused by multi-arm treatments, when the statistical analysis is based on the use of pooled summary data extracted from pairwise meta-analysis.

## Appendix 10.1: WinBUGS Code of Umbrella Review

```
#MODEL
model{
for (i in 1:N){
LOR[i]~dnorm(lor[comp[i],treat[i]],prec[i])
prec[i]<-1/var[i]
var[i]<-sd[i]*sd[i]
sd[i]<-(LUCI[i]-LLCI[i])/3.92
dev[i]<-(LOR[i]-lor[comp[i],treat[i]])*(LOR[i]-
lor[comp[i],treat[i]])*prec[i]
}
resdev<-sum(dev[]) #summed residual deviance
d[1]<-0
for (k in 2:NT){
d[k]~dnorm(0,0.0001) #vague prior for basic parameters
}
for (k in 1:NT){
best[k]<-equals(rank(d[],k),1)
}
for (c in 1:NT-1){
for (k in (c+1):NT){
lor[c,k]<-d[k]-d[c]
or[c,k]<-exp(lor[c,k])
}
}
}
#DATA for FE
list(N=4, NT=4,
comp=c(1,2,2,3),
treat=c(2,3,4,4),
LOR=c(-1.993,0.279,1.14,0.83),
LLCI=c(-2.583,0.038,0.868,0.359),
LUCI=c(-1.402,0.521,1.412,1.3)
)
#DATA for RE
list(N=4, NT=4,
comp=c(1,2,2,3),
treat=c(2,3,4,4),
LOR=c(-2.528,0.412,1.453,1.011),
LLCI=c(-4.345,-0.166,0.58,-0.839),
LUCI=c(-0.711,0.991,2.327,2.86)
)
#INITIAL VALUES
list(
d=c(NA, 0,0,0)
)
```

## Appendix 10.2: Stata Codes for Statistical Analysis of Umbrella Reviews

```
Using weighted least squares (WLS) method
vwls d t12-t14, noconstant sd(se)
Using Stata command GLLAMM
generate lns = ln(se)
* set up lns as the lower level residual variance
eq het: lns
* set constraint 1: the coefficient for lns is 1
constraint define 1 [lns1]lns=1
* generate a new variable cons which is a vector of 1
gene cons=1
* set up random intercept model for treatment effect
eq int: cons
* constrain the variance of random effects to 0, so it becomes a
fixed effect analysis constraint define 2 [stud1]_cons=0
* run the fixed effect analysis using gllamm
gllamm d t12-t14, noconstant i(study) constraint(1 2) adapt s(het)
nip(5)
```

## References

1. Hasselblad V. Meta-analysis of multitreatment studies. Med Decis Making. 1998;18(1):37–43.
2. Bucher HC, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50(6):683–91.
3. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med. 2002;21(16):2313–24.
4. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004;23(20):3105–24.
5. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, Watanabe N, Nakagawa A, Omori IM, McGuire H, Tansella M, Barbui C. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet. 2009;373:746–58.
6. Cipriani A, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. Lancet. 2011;378(9799):1306–15.
7. Tu YK, Woolston A, Faggion Jr CM. Do bone grafts or barrier membranes provide additional treatment effects for infrabony lesions treated with enamel matrix derivatives? A network meta-analysis of randomized-controlled trials. J Clin Periodontol. 2010;37(1):59–79.
8. Wong MC, et al. Cochrane reviews on the benefits/risks of fluoride toothpastes. J Dent Res. 2011;90(5):573–9.
9. Tu Y-K, et al. A bayesian network meta-analysis on comparisons of enamel matrix derivatives, guided tissue regeneration and their combination therapies. J Clin Periodontol. 2012;39(3):303–14.
10. Dias S, et al. Evidence synthesis for decision making 1: introduction. Med Decis Making. 2013;33(5):597–606.

11. Dias S, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making. 2013;33(5):607–17.
12. Dias S, et al. Evidence synthesis for decision making 3: heterogeneity – subgroups, meta-regression, bias, and bias-adjustment. Med Decis Making. 2013;33(5):618–40.
13. Dias S, et al. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making. 2013;33(5):641–56.
14. Dias S, et al. Evidence synthesis for decision making 5: the baseline natural history model. Med Decis Making. 2013;33(5):657–70.
15. Dias S, et al. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. Med Decis Making. 2013;33(5):671–8.
16. Ades AE, et al. Evidence synthesis for decision making 7: a reviewer's checklist. Med Decis Making. 2013;33(5):679–91.
17. Andrea Cipriani JPTH, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med. 2013;159:130–7.
18. Moe RH, et al. Effectiveness of nonpharmacological and nonsurgical interventions for hip osteoarthritis: an umbrella review of high-quality systematic reviews. Phys Ther. 2007;87(12):1716–27.
19. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ. 2009;181(8):488–93.
20. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol. 2010;63(8):875–82.
21. Tu Y-K. Use of generalized linear mixed models for network meta-analysis. Med Decis Making. 2014;34(7):911–8.
22. Tu Y-K. Linear mixed model approach to network meta-analysis for continuous outcomes in periodontal research. J Clin Periodontol. 2015;42(2):204–12.
23. Rabe-Hesketh S, Skrondal A, Pickles A. Generalized multilevel structural equation modelling. Psychometrika. 2004;69(2):167–90.
24. Rabe-Hesketh S, Skrondal A, Pickles A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. J Econom. 2005;128(2):301–23.
25. Macfadyen CA, Acuin JM, Gamble CL. Topical antibiotics without steroids for chronically discharging ears with underlying eardrum perforations (review). Cochrane Database Syst Rev. 2005;(4):CD004618.

# Bayesian Approach to Evidence Synthesis

# 11

Francisco José Vázquez Polo, Miguel A. Negrín, and María Martel

**Abstract**

We briefly present the advantages and opportunities available to umbrella reviews from the use of Bayesian techniques while taking into account that the concerns commonly arising in Bayesian meta-analysis procedures are also present in umbrella reviews. This is the case, for example, of sparse data, for which the hierarchical logit-normal model can give very poor results. An additional concern in this context is that of the choice of noninformative priors, which can lead to a significant variation in the final conclusions drawn. Accordingly, this chapter highlights the potential for Bayesian approaches in umbrella reviews, overviews of reviews, and meta-epidemiologic studies while acknowledging their limitations and complexities.

## 11.1 Introduction

Umbrella reviews are currently used in many practical situations in which various systematic reviews and/or multiple meta-analyses on the same disease or condition are available and we need a method to compile data from these diverse sources. An umbrella review is a compilation of reviews of many treatments, an approach that provides a broader outlook than can a single meta-analysis. However, these reviews are limited by the quantity and quality of information available in the previous studies included, and so an umbrella review is best conducted prospectively, defining in advance the number of interventions and outcomes to be considered. This calls for

F.J.V. Polo, PhD (✉) • M.A. Negrín • M. Martel
Department of Quantitative Methods and TiDES Institute, University of Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas de Gran Canaria, Canary Islands, Spain
e-mail: francisco.vazquezpolo@ulpgc.es



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_11

greater effort than performing a separate analysis of each review, but it is more efficient than a piecemeal, uncoordinated approach [1]. As Ioannidis pointed out, "the synthesis of such complex information requires rigorous and systematic methods" [1].

In general, the Bayesian approach is very suitable and has been widely used to consider different ways of incorporating and encapsulating evidence in health technology assessment, ranging from a simple meta-analysis of randomised head-to-head trials to complex multiple treatment comparison meta-analyses, in which sophisticated procedures are used to obtain indirect comparisons of different treatments. One such application of meta-analytic techniques is possible within an umbrella review.

The most common methods based on the normal approximation are Bayesian random-effect methods, in which the observed effect (or a transformation of this effect) is assumed to follow a normal distribution with parameters $\theta i$ and $\sigma i^2$, where $\theta i$ represents the effect of the treatment in the study $i = 1,\ldots,k$, among the $k$ studies considered. Furthermore, it is assumed that the distribution of $\theta i$, given $\theta$, is also normal. This allows us to use an unobservable variable $x$ with a normal distribution of parameters $\theta$ and $\sigma^2$, where $\theta$ represents the (unconditional) effectiveness of the treatment and $\sigma^2$ represents the heterogeneity among studies. As a hierarchical model, it would have the following structure:

$$\begin{aligned}
&\left(X_i \mid \theta_i\right) \sim N\left(\theta_i, \sigma_i^2\right) \quad \text{independent} \\
&\left(\theta_i \mid \theta, \sigma^2\right) \sim N\left(\theta, \sigma^2\right) \quad \left(\text{link distribution}\right) \\
&\left(\theta, \sigma^2\right) \sim \pi\left(\theta\right)\pi\left(\sigma^2\right)
\end{aligned}$$

where the a priori densities $\pi(z)$ are elicited previously.

As is well known, Bayesian methods make it possible to incorporate several sources of information, such as observational studies and/or expert judgement. Broadly speaking, Bayesian procedures assume that the true effect of each treatment is the same across different trials and/or reviews and also assume that the logit transformation of rates follows a multivariate normal distribution. Accordingly, all the advantages obtained from using a Bayesian approach, as presented in [2], are shared by the above procedures. However, they also suffer various weaknesses, indicated in [3] as follows, and these should be investigated:

1. The use of the subjective prior beliefs destroys any element, or illusion, of objectivity.
2. Eliciting prior beliefs is a non-trivial exercise, and at present there are few guidelines to help the Bayesian analyst.
3. Different prior distributions can be used, generating varying results, and therefore there is no definitive analysis and a sensitivity analysis is always required.
4. Bayesian methods can be computationally complex and thus is time consuming to perform.

To date, only point 4 has been addressed in depth, and this is basically due to the development of Markov chain Monte Carlo (MCMC) techniques. Points 1, 2, and 3 are related to the choice of priors to be used in any Bayesian analysis. With respect to points 1 and 2, in practice, prior densities are chosen to be noninformative or weakly informative. In these cases the results from the hierarchical logit-normal model can be very poor when binary data are considered, i.e. Bayesian random-effect normal models should be used with caution for a meta-analysis of sparse binary data.

An additional concern in this context is related to the choice of noninformative priors, which can lead to significant variations in the final conclusions drawn. Therefore, it is very important to conduct robustness studies and sensitivity analyses for classes of plausible prior distributions. A particular type of Bayesian sensitivity study is developed in [3]. Clearly, it would be very useful to perform a sensitivity analysis with respect to the prior distributions in a meta-analysis [4], and we examine different ways in which this can be done. Observation of a wide-ranging body of research in this area would provide a broad impression of the magnitude of the results to be obtained from such an analysis.

## 11.2 The Case of Binary Data

For the sake of simplicity, we focus on the case of binary data, an area in which many systematic reviews have been conducted. One of the most controversial points in Bayesian random-effect models is the choice of the linking distribution, that is, the probability conditional distribution which "links" the treatment effect on the ith study with the (unconditional) treatment effect.

Moreno et al. [5] showed that by choosing a suitable class of linking distribution, the continuity corrections required in normal random-effect models are not necessary and that a sensitivity analysis of the quantities of interest can be performed straightforwardly, by selecting a particular family of compatible priors.

It is increasingly accepted that objective Bayesian methods are valid instruments that can be used in many areas. The meta-analysis of unusual events is a particular case in which these techniques could be useful. Most previous studies of meta-analysis for treatment comparison have sought to estimate metaparameters, rather than addressing this as a problem in the field of Bayesian model choice. In terms of their computation, objective Bayesian methods are increasingly feasible, and so significantly greater use is to be expected.

### 11.2.1 Bayesian Model Selection Problems

Following [5], we focus the problem as one in the Bayesian model selection context. For the *i*th study, we consider the model for the observable variable $x_i$:

$$M_i : \left\{ \Pr\left(x_i \mid n_i, \theta_i\right) = \mathrm{Bin}\left(x_i \mid n_i, \theta_i\right), \pi\left(\theta_i\right) \right\}, \quad i = 1, \ldots, k \tag{11.1}$$

and the corresponding induced meta-model for the 0–1 unobservable effectiveness variable $x$ is given by

$$M_0 : \left\{ \Pr\left(x|\theta\right) = \text{Ber}\left(x \mid \theta\right), \pi\left(\theta\right) \right\}, \tag{11.2}$$

where Bin and Ber denote the binomial and Bernoulli distributions, respectively, and $\pi\left(\cdot\right)$ is the prior distribution on the parameters of interest. An objective prior is then selected, for instance, a uniform or the Jeffreys prior.

In order to obtain inferences on the parameter $\theta$ (i.e. to obtain its posterior distribution), a link distribution is required, and this is denoted by $\pi(\theta i|\theta)$. Obviously, this link distribution is a conditional one and it must be compatible with the marginals in [1] and [2], i.e. it should satisfy the following integral equations:

$$\int_0^1 \pi\left(\theta_i \mid \theta\right)\pi\left(\theta\right)\mathrm{d}\theta_i = \pi\left(\theta\right), \quad \int_0^1 \pi\left(\theta_i \mid \theta\right)\pi\left(\theta\right)\mathrm{d}\theta = \pi\left(\theta_i\right). \tag{11.3}$$

Moreno et al. [5] proposed some well-suited classes of linking distributions, namely, Farlie-Gumbel-Morgenstern (FGM) distributions, the family of the Sarmanov distributions, and the class of the intrinsic distributions, and these present interesting properties. Among other aspects, they allow us to model between-study heterogeneity in a simple way. For instance, for the FGM family of distributions, the conditional variance is easily obtained by

$$V\left(\theta_i \mid \theta\right) = \frac{1}{3} + \frac{\rho}{2}\left(2\theta - 1\right) - \left(\frac{1}{2} + \frac{\rho}{2}\left(2\theta - 1\right)\right)^2,$$

where $\rho$ is the Pearson's correlation coefficient between $\theta_i$ and $\theta$

Now, assuming the data $\left(x_i, n_i\right), \quad i = 1, \ldots, k$ are independent, conditionally on $\theta$, and that $\theta i$ are conditionally independent given $\theta$, we can obtain the likelihood associated with the effectiveness $\theta$:

$$\Pr\left(x \mid n, \theta\right) = \prod_{i=1}^{k} \int_0^1 \Pr\left(x_i \mid n_i, \theta_i\right)\pi\left(\theta_i \mid \theta\right)\mathrm{d}\theta_i,$$

where $x$ and $n$ are the entire data set including all data from the studies considered. The posterior distribution of $\theta$ containing all the information supplied by the data is given by the Bayesian theorem by

$$\pi\left(\theta \mid x, n\right) = \frac{\Pr\left(x \mid n, \theta\right)\pi\left(\theta\right)}{\int_0^1 \Pr\left(x \mid n, \theta\right)\pi\left(\theta\right)\mathrm{d}\theta}. \tag{11.4}$$

The posterior distribution in [4] can now be used, for instance, to test the equality of treatment effects. Suppose that two treatments T1 and T2 are applied to patients in $k$ studies, in which we observe the effectiveness samples $(z_1, m_1)$ and $(z_2, m_2)$

$$z_t = \left(x_1^t, x_2^t, \ldots, x_k^t\right), \quad m_t = \left(n_1^t, n_2^t, \ldots, n_k^t\right).$$

Let us consider the likelihoods and priors of the meta-effectiveness of the treatments

$$\{\Pr(z_1 \mid m_1,\varsigma),\pi(\varsigma)\},\{\Pr(z_2 \mid m_2,\xi),\pi(\xi)\}.$$

We are interested in testing the null hypothesis that the treatment effectiveness is equal in each case, i.e. $H_0 : \varsigma = \xi$ . As in [4] this problem is equivalent to the Bayesian model selection problem:

$$M_0 : \{\Pr(z_1 \mid m_1,\theta)\Pr(z_2 \mid m_2,\theta),\pi(\theta)\},$$

and

$$M_1 : \{\Pr(z_1 \mid m_1,\varsigma)\Pr(z_2 \mid m_2,\xi),\pi(\varsigma)\pi(\xi)\}.$$

The posterior probability of the null coincides with the posterior probability of $M_0$ and, in the usual case of noninformative prior probabilities, is given by $B_{10}(z_1, m_1, z_2, m_2)$:

$$\Pr(M_0 \mid z_1,m_1,z_2,m_2) = \frac{1}{1+B_{10}(z_1,m_1,z_2,m_2)}, \tag{11.5}$$

where $B_{10}(z_1, m_1, z_2, m_2)$ is the Bayesian factor for comparing $M_1$ and $M_0$, and where

$$B_{10}(z_1,m_1,z_2,m_2) = \frac{\int_0^1\int_0^1 \Pr(z_1 \mid m_1,\varsigma)\Pr(z_2 \mid m_2,\xi)\pi(\varsigma)\pi(\xi)\,\mathrm{d}\varsigma\,\mathrm{d}\xi}{\int_0^1 \Pr(z_1 \mid m_1,\theta)\Pr(z_2 \mid m_2,\theta)\pi(\theta)\,\mathrm{d}\theta}. \tag{11.6}$$

The optimal decision under a 0–1 loss function is to accept the null when $B_{10} < 1$ .

**Conclusion**

The Bayesian meta-analysis described depends strongly on the linking distribution chosen. The particular dependence structure between $\theta_i$ and the $\theta$ induced by the FGM, Sarmanov, or intrinsic family [3] means that a range of heterogeneity can be modelled straightforwardly, and so this approach is very useful in practical applications. If any of the studies analysed present significant homogeneity, the likelihoods obtained for the models should be corrected accordingly.

Examining a sensitivity analysis with respect to the prior distributions in a meta-analysis [2] is a question of obvious practical interest and we believe it is interesting to explore different means of doing so. By establishing a wide-ranging body of research into this question, a broad view of the magnitude of the outcome of such an analysis could be obtained, and the present study should be considered a step in this direction.

The umbrella review is a recent and sophisticated technique for the compilation of results on treatment effects, one that complements the techniques of meta-analysis commonly applied. Certain problems remain to be overcome in

meta-analysis and these must be studied with procedures other than those routinely adopted. In this paper, we present, by way of example, the case of 0–1 binary data, where the standard techniques of random-effect models run into serious difficulties and where a formulation in terms of Bayesian model comparison can solve the problem. Objective Bayesian techniques provide good solutions to these problems and umbrella review procedures, both present and future, should clearly take them into consideration.

**Acknowledgements** This work was partially supported by grant ECO2013-47092 (Ministerio de Economía y Competitividad, Spain).

## References

1. Ioannidis JPA. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. Can Med Assoc J. 2009;181(8):488–93.
2. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res. 2001;10:277–303.
3. Vázquez-Polo FJ, Moreno E, Negrín MA, Martel M. A Bayesian sensitivity study of risk difference in the meta-analysis of binary outcomes from sparse data. Expert Rev Pharmacoecon Outcomes Res. 2015;15(2):317–22.
4. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med. 2005;24:2401–28.
5. Moreno E, Vázquez-Polo FJ, Negrín MA. Objective Bayesian meta-analysis for sparse discrete data. Stat Med. 2014;33:3676–92.

# Appraising Between-Study Homogeneity, Small-Study Effects, Moderators, and Confounders

12

Areti Angeliki Veroniki, Tania B. Huedo-Medina, and Kostas N. Fountoulakis

**Abstract**

Meta-analysis is the statistical synthesis of results from two or more clinical studies that address the same issue and compare two different interventions. Although the combination of results of several studies in a meta-analysis can increase power and improve precision, caution is needed in the presence of between-study heterogeneity and selection bias. These two factors can importantly impact meta-analysis conclusions and hence influence decision-making. Several methods have been developed to appraise the between-study variation and the tendency of small studies to yield larger intervention effects compared to larger studies. This chapter presents an overall review of methods presented in the meta-analysis literature along with their properties.

## 12.1 Introduction

Systematic reviews and meta-analyses of well-conducted randomized controlled trials (RCTs) that address the same clinical question(s) can provide the highest level of evidence for decision-making on interventions and are vital in the practice of evidence-based medicine. Although meta-analysis constitutes a valuable tool to

A.A. Veroniki, PhD (✉)
Li Ka Shing Knowledge Institute, Faculty of Medicine, St. Michael's Hospital and University of Toronto, 30 Bond St, Toronto, ON M5B 1W8, Canada
e-mail: veronikia@smh.ca

T.B. Huedo-Medina
Department of Allied Health Sciences, University of Connecticut, Storrs, CT, US

K.N. Fountoulakis
Department of Psychiatry, Division of Neurosciences School of Medicine, Aristotle University of Thessaloniki, Pournari Pylaia, Thessaloniki, Greece



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_12

summarize study-specific results and may reduce both bias and uncertainty from individual studies, it widely depends on the quality, homogeneity, and freedom from bias of the available studies. The main two threats of the meta-analysis validity are:

1. The between-study variability beyond random error, termed heterogeneity
2. The phenomenon that small RCTs suggest different, often larger, intervention effects than large RCTs, termed "small-study effects" [1–3]

A certain degree of variability in study-specific intervention effects is almost always present due to chance, but additional variability might occur due to many reasons. These might include differences in the way studies are conducted and how the intervention effect estimates are measured. There are three different types of heterogeneity:

1. Clinical heterogeneity, which is referred to as the variability in the participants, interventions, and outcomes
2. Methodological heterogeneity, which reflects the variability in study design and risk of bias
3. Statistical heterogeneity, which is referred to as the variability in the intervention effects

Statistical heterogeneity is usually a consequence of clinical or methodological variability, or both, among trials, and is often called "heterogeneity" omitting the term "statistical." The estimation of heterogeneity is an additional aim in meta-analysis as it improves interpretation of results and can provide insights on the summary intervention effect predictions. One of the most widely statistical methods used in meta-analysis is the inverse-variance method; it uses the reciprocal of the within-study variances as study weights. The presence of heterogeneity affects the estimation of study weights and hence the estimated uncertainty of the summary intervention effect.

A commonly encountered association in meta-analysis is the one between the estimated study-specific intervention effects and the size of studies; it can be caused by several reasons. One possible explanation is that small studies with non-significant results are less likely to be published, because journals and authors may tend to publish and submit small studies with significant results. Other explanations may include selective outcome reporting (e.g., reporting outcomes with statistically significant results), heterogeneity between small and large studies (e.g., small studies recruit patients of high baseline risk that would largely benefit from the intervention), mathematical artifact between the two factors, or simply coincidence.

Several approaches have been proposed to estimate the between-study heterogeneity and small-study effects as a result of selection bias (including publication bias, language bias, citation bias, and reporting bias) [4–6]. This chapter includes a review of the graphical methods, statistical tests, and statistical measures used in pairwise meta-analysis to evaluate homogeneity and selection bias.

## 12.2 Approaches for Assessing the Between-Study Heterogeneity

A key aim in meta-analysis is to make inferences about the between-study heterogeneity as its presence can have a considerable impact on the meta-analysis conclusions. There are multiple approaches available to evaluate heterogeneity in meta-analysis, including graphical methods and statistical tests to assess its presence, statistical measures to quantify heterogeneity, and methods to estimate its magnitude. This section discusses several alternatives to appraise between-study heterogeneity in meta-analysis.

### 12.2.1 Graphical Representation of the Between-Study Heterogeneity

A visual inspection of graphical representations is commonly the first approach researchers select to assess the variation between study-specific effects due to heterogeneity, beyond what is expected by chance. This is an informal approach but a very useful way to indicate outlier studies, as well as those that might be responsible for the between-study heterogeneity. In the next subsections, we present the graphical displays that have most commonly been used in the meta-analysis literature [7, 8].

#### 12.2.1.1 Forest Plot

Forest plots (Fig. 12.1) are the most popular plots in meta-analysis; they display the study-specific effect estimates along with their confidence intervals, and at the bottom of the plot, the meta-analysis result is provided [10–12]. The effect measure (e.g., odds ratio) is usually presented on the horizontal axis allowing detailed study data to be plotted alongside the results, such as the number of events and sample size for each study arm. However, some authors argue that the effect measure should be presented on the vertical axis as dependent variables are commonly plotted in statistics [13]. The size of the plotting symbol used to represent the intervention effect is usually selected to be proportional to the inverse of the variance of the

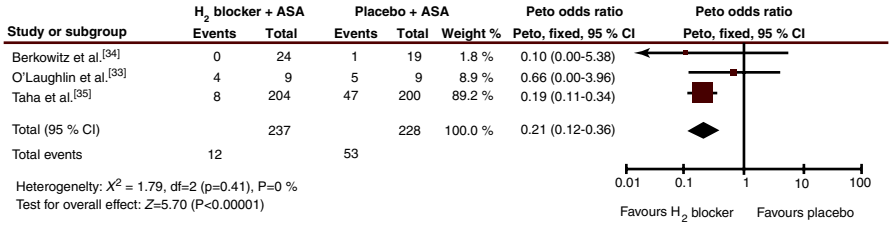



**Fig. 12.1** Forest plot. Meta-analysis of three randomized controlled trials of histamine $H_2$ receptor antagonists ($H_2$ blockers) in conjunction with acetylsalicylic acid (*ASA*) therapy for outcome of peptic ulcer (Reproduced with permission [9])

study effect estimate. Therefore, more precise estimates (i.e., with smaller variance) are represented by larger plot symbols, highlighting also the amount of information that they contribute to the meta-analysis.

A greater variation in the study-specific intervention effects, more than it would be expected by chance alone, suggests there is evidence for between-study heterogeneity. In a forest plot, this is usually inspected by the poor overlap of the intervention effects' confidence intervals.

#### 12.2.1.2 Galbraith Plots

Galbraith (or radial) plots (Fig. 12.2) are often used to present the results of studies in a meta-analysis and to informally assess between-study heterogeneity [15, 16]. The plot is a scatter plot of the standardized study-specific intervention effects, i.e., the estimated effect measures (e.g., log-odds ratio) divided by their standard errors (SE) (or equivalently the z-score) on the *y*-axis, against their inversed SEs on the *x*-axis. Each study is represented by a single point, and a regression line is drawn corresponding to the pooled fixed-effect meta-analysis estimate. Therefore, the slope of the regression is as an estimate of the intervention effect, when there are no small-study effects. In addition, the 95 % confidence region of the through-the-origin regression line is depicted by the area between the two lines drawn at a vertical distance of $\pm 2$ above and below the regression line. Under the assumption that all studies estimate a common (fixed) intervention effect, we expect that the majority (95 %) of study points lie within this confidence region.

Using this graphical representation, studies outside this region contribute to between-study heterogeneity, and the imprecise (small 1/SE, or large SE, or small

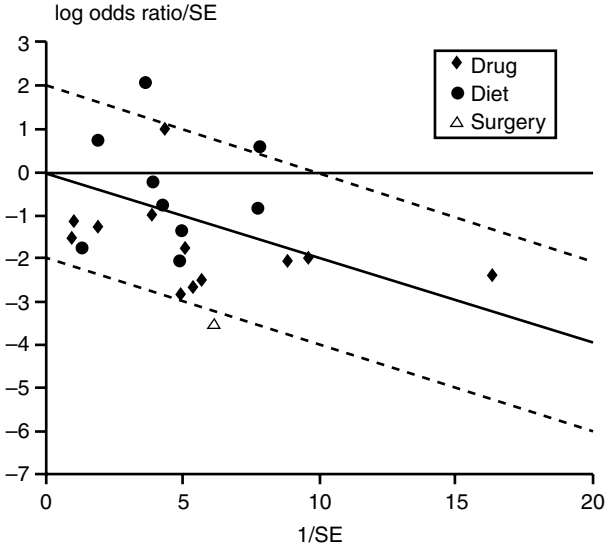



**Fig. 12.2** Galbraith plot. Log-odds ratio for ischemic heart disease in trials of serum cholesterol reduction by type of intervention (Reproduced with permission [14])

studies) intervention effects lie close to the *y*-axis, whereas precise intervention effects will be situated further away.

#### 12.2.1.3 L'Abbé Plot

L'Abbé plots (Fig. 12.3) facilitate the examination of whether the intervention effects across studies are homogeneous, but they can be used for dichotomous outcome data only [18]. This type of plot presents the risks (or odds) in the intervention group on the *y*-axis against those of the control group on the *x*-axis and often includes the diagonal line of equality and a regression line. The diagonal line of equality indicates that the risks in the control and intervention groups are equal within trials, and the regression line represents the risk ratio (or odds ratio), which is estimated by pooling the results in the meta-analysis. It is advisable that the study points are presented according to the precision of the intervention effect estimates (or study size) to make the plot more informative [7].

The plot can be used to infer the presence of heterogeneity, specifically where trials are widely spread around the regression line. In the absence of heterogeneity, study points should lie closely around the regression line.

#### 12.2.1.4 Baujat Plot

Baujat plots (Fig. 12.4) are used to identify studies that influence the overall intervention effect and impact on the magnitude of the heterogeneity [19]. The rationale is that excluding an influential study will affect the meta-analytic estimate, and hence this plot assesses which studies cause the between-study heterogeneity and the greatest shifts in the overall intervention effect. The plot presents the contribution of each study to the Cochran *Q*-statistic (see Sect. 12.2.2.1) on the *x*-axis against

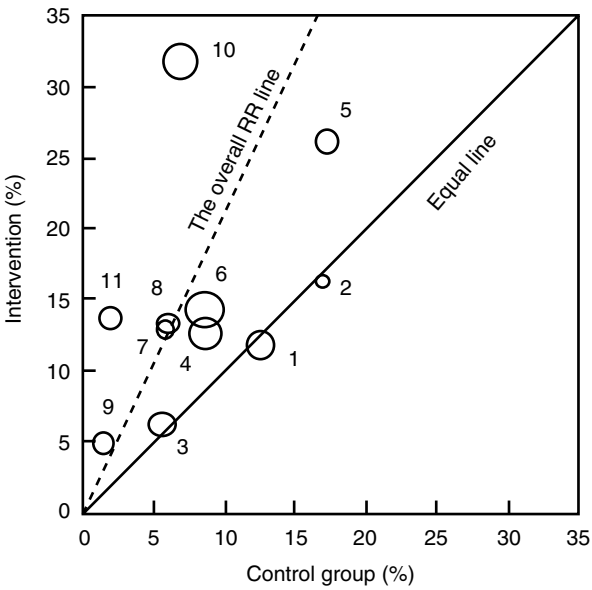



**Fig. 12.3** L'Abbé plot. Rates of smoking cessation in the intervention and control group (Reproduced with permission [17])

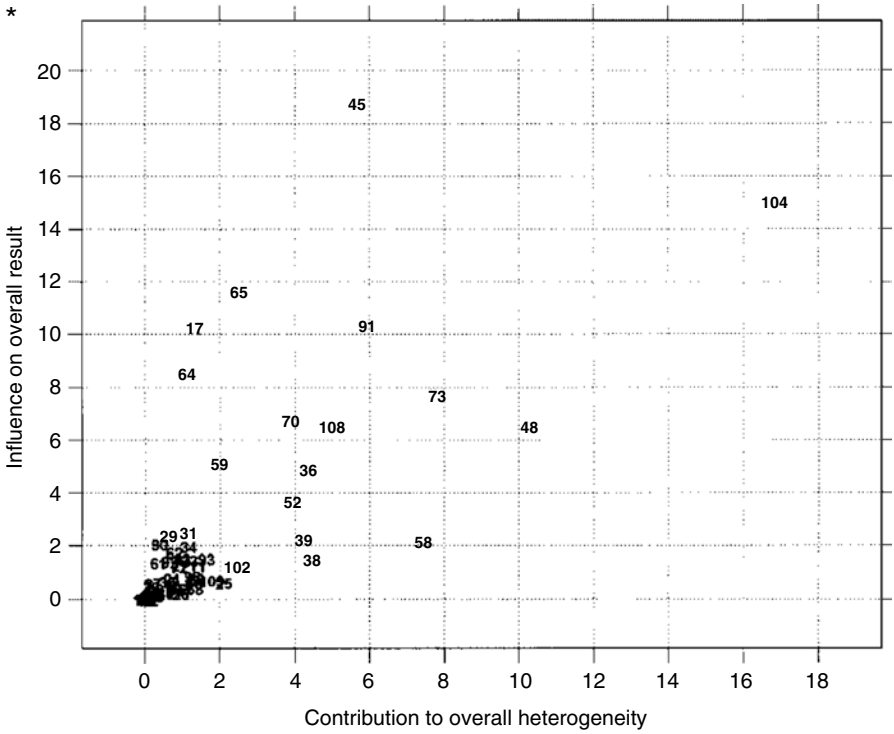



**Fig. 12.4** Baujat plot for a meta-analysis of chemotherapy in head and neck cancer (Reproduced with permission [19])

the influence of each study. The influence of each study is defined as the standardized squared difference between the overall intervention effects with and without the *i*th study under the fixed-effect model, on the *y*-axis. Studies lying on the upper right corner of the plot are the most influential with the highest contribution to the total heterogeneity.

### 12.2.2 Statistical Tests for the Evaluation of the Between-Study Variance

The most commonly used method to assess the homogeneity assumption in meta-analysis is to carry out a statistical test. Several tests for this evaluation have been suggested in the literature, including the "generalized Cochran *Q*," Wald, likelihood ratio, and score tests [20, 21]. A popular choice for the between-study homogeneity assessment in meta-analyses is the Cochran *Q*-statistic (see Sect. 12.2.2.1) [22]. It has been suggested that among the aforementioned homogeneity tests, the Cochran *Q*-statistic performs best in terms of type I error for meta-analyses with large studies

(e.g., with arm size greater than 640) [21]. The Cochran $Q$-statistic belongs to the "generalized Cochran between-study variance statistics" family [23], with

$$Q_a = \sum a_i \left( y_i - \mu_a \right)^2,$$

where $y_i$ is the observed intervention effect (e.g., log-odds ratio), index $i$ refers to the $i$th study with $i = 1,\ldots k$, $a_i$ the weight assigned to each study, and $\mu_a = \left(\sum a_i y_i\right) / \sum a_i$ the overall intervention effect. Jackson showed that $Q_a$ has a $\chi^2_{k-1}$ distribution as a linear combination of independent central $\chi_1^2$ random variables [24].

#### 12.2.2.1 Cochran *Q*-Statistic

The standard test widely used in meta-analysis, is the Cochran $Q$-statistic testing the hypothesis that all studies share a common true effect ($\mu$) or equivalently that the between-study variance ($\tau^2$) is zero [22]. The Cochran $Q$-statistic is a special form of the "generalized Cochran between-study variance statistic" for $a_i = 1/v_i$, with $vi$ the within-study variance in study $i = 1,\ldots k$. Hence, the $Q$-statistic is the weighted sum of squared differences between the observed study-specific effects and the overall effect across studies derived under the fixed-effect model. Under the null hypothesis, $H_0: \tau^2 = 0$, the $Q$-statistic follows approximately a $\chi^2$-distribution with $k-1$ degrees of freedom and a critical region $Q > \chi^2_{k-1,\,1-(a/2)}$, where $a$ is the confidence level. Several efforts have been done to define the distribution of the $Q$-statistic, including Biggerstaff and Tweedie approximating $Q$ with a gamma distribution, and Biggerstaff and Jackson deriving the exact distribution, when $\tau^2 \neq 0$ [25, 26].

It has been shown that the power of the test to detect heterogeneity depends on the number and size of studies, as well as the magnitude of the true between-study variance [21]. Simulation studies suggest that the test has low power when the total information available in the meta-analysis is low (e.g., sparse data, small size and number of studies), and hence a nonsignificant result might erroneously be interpreted as absence of between-study heterogeneity [21, 27]. It is therefore recommended that reviewers use 0.10 as a cutoff level of significance instead of the usual 0.05 [28, 29]. However, a higher cutoff value increases type I error and the risk of drawing false-positive results. The $Q$-statistic may suggest significant heterogeneity when many studies are included in the meta-analysis and particularly when their sample sizes are very large (see, e.g., Barbui et al. that included over 15,000 participants from 135 studies) [30]. The power of the test may also be limited when the study sizes differ substantially or a single study is a lot larger when compared with the others in the analysis [27].

#### 12.2.2.2 Generalized *Q*-Statistic

Similarly to Cochran $Q$, the generalized $Q$-statistic ($Q_{\text{gen}}$) is a special form of the "generalized Cochran" between-study variance statistic for $a_i = 1/\left(v_i + \tau^2\right)$. The $Q_{\text{gen}}$-statistic is the weighted sum of squared differences between the observed study-specific effects and the overall effect derived under the random-effects model. Under the null hypothesis that the true between-study variance is equal to a certain

amount ($\tau_0^2 \geq 0$), $Q_{\text{gen}}$ follows a $\chi^2$-distribution with $k-1$ degrees of freedom and a critical region: $Q_{\text{gen}} > \chi^2_{k-1,\,1-(a/2)}$.

To the best of our knowledge, the properties of the test have not been examined, providing an avenue for further work.

#### 12.2.2.3 Cochran *Q*-Statistic Adjusted for Small-Study Effects

Rücker et al. extended Cochran $Q$-statistic by adjusting for small-study effects [31]. We call "small-study effects" the tendency of small studies to show larger intervention effects compared to the larger studies (see also Sect. 12.4). This can be derived by

$$Q_a^{\text{Adj}} = \sum a_i \left( y_i - \mu_a^{\text{Adj}} - \frac{\hat{s}_a}{\sqrt{a_i}} \right)^2,$$

where $\mu_a^{\text{Adj}}$ is the summary intervention effect adjusted for small-study effects with $a_i = 1/v_i$ and $\hat{s}_a$ represents a potential small-study effect. The $Q_a{}^{\text{Adj}}$ measures the residual variation with respect to a fixed-effect model allowing for small-study effects, and compared to the Cochran's $Q$-statistic, it holds that $Q_a^{\text{Adj}} \leq Q$. Under the null hypothesis of no between-study heterogeneity, $Q_a{}^{\text{Adj}}$ follows a $\chi^2$-distribution with $k-2$ degrees of freedom and a critical region: $Q_a^{\text{Adj}} > \chi^2_{k-2,\,1-(a/2)}$.

In the presence of small-study effects, it is suggested to use $Q_a{}^{\text{Adj}}$ to assess the remaining between-study heterogeneity [31]. The main limitation of the Cochran's $Q$-statistic adjusted for small-study effects is that it depends on the choice of the estimation method for $\tau^2$ (see Sect. 12.2.4).

### 12.2.3 Statistical Measures to Quantify Between-Study Variance

The statistical tests discussed in Sect. 12.2.2 are only useful for testing the existence of heterogeneity, but do not quantify the extent of heterogeneity. To date, several statistical measures have been suggested for the quantification of the degree of variability in a meta-analysis that is explained by between-study differences rather than by random error [32–34]. As for every point estimate, apart from quantifying between-study heterogeneity using a statistical measure, it is important to quantify its corresponding uncertainty too. Confidence intervals provide information on the precision and the range of values that reflect the statistical measure for heterogeneity. Methods for constructing the confidence intervals include the variance estimates recovery method [35, 36], the method using the distribution of $Q_a$-statistic [24–26, 32], the method based on the statistical significance of $Q$ [32], the method based on the between-study variance estimator (see Sect. 12.2.4) [5, 32, 37], and the method using a nonparametric bootstrap approach [32].

#### 12.2.3.1 $H^2$ Index

$H^2$ index (also known as Birge ratio) [38] has been presented by Higgins and Thompson [32] and shows the excess of the observed $Q$ over its expected value,

$E(Q) = k - 1$. The measure reflects the relationship of between- and within-study variance and can be obtained by

$$H^2 = \frac{Q}{k-1} = \frac{\hat{\tau}^2_{\mathrm{DL}} + \sigma^2}{\sigma^2}$$

where $\hat{\tau}^2_{\mathrm{DL}}$ is the estimated between-study variance using the DerSimonian and Laird [39] estimator and $\sigma^2$ is the "typical" within-study variance:

$$\sigma^2 = \frac{\sum \frac{1}{v_i}(k-1)}{\left(\sum \frac{1}{v_i}\right)^2 - \sum\left(\frac{1}{v_i}\right)^2}.$$

The statistic takes values within the range (1, ∞), and in the absence of between-study heterogeneity, it equals 1. Higgins and Thompson [32] suggest that there is no universal rule to define thresholds for 'low,' 'moderate,' and 'high' heterogeneity for $H^2$. However, they suggest that values greater than 1.5 may show considerable heterogeneity, and values lower than 1.2 may show moderate to low heterogeneity.

#### 12.2.3.2 $I^2$ Index

The $I^2$ index reflects the percentage of the total variability in a set of effect measures that is due to between-study variability beyond what is expected by within-study error. The "generalized $I^2$ statistics" family [37] can be expressed as

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \sigma^2}$$

where $\hat{\tau}^2$ is the estimated between-study variance using one of the methods suggested in the literature (see Sect. 12.2.4) [5]. The $I^2$ index can be expressed as a percentage ranging from 0 to 100 %, where a value of 0 % indicates no observed heterogeneity. The Cochrane Handbook advises avoiding the use of specific thresholds for the interpretation of the $I^2$ statistic as they may be misleading. A general guideline to its interpretation is the following [3]:

- From 0 to 40 %, may not be important.
- From 30 to 60 %, may represent moderate heterogeneity.
- From 50 to 90 % may represent substantial heterogeneity.
- From 75 to 100 %, may represent considerable heterogeneity.
- Note that should these guidelines be used with caution, and always interpret the $I^2$ index along with its confidence interval.

##### $I^2$ Index Based on Cochran *Q*-Statistic

The $I^2$ based on Cochran $Q$-statistic is the most popular statistic and is usually the default method to quantify heterogeneity in most meta-analysis software.

The method is a special form of the "generalized $I^2$ statistics" using the DerSimonian and Laird approach [39] (see Sect. 12.2.4.1):

$$I^2_{\mathrm{DL}} = \frac{\hat{\tau}^2_{\mathrm{DL}}}{\hat{\tau}^2_{\mathrm{DL}} + \sigma^2}.$$

Alternatively, the method can be presented as

$$I^2_{\mathrm{DL}} = \frac{H^2 - 1}{H^2} = \frac{Q - (k-1)}{Q}$$

in terms of either $H^2$ or Cochran's $Q$-statistic and its degrees of freedom $(k-1)$. The $I^2$ statistic should be interpreted with caution when the number and size of studies in the meta-analysis are small (e.g., for fewer than ten studies in the meta-analysis and studies with fewer than 100 participants) [34, 40, 41]. Simulation studies have shown that $I^2_{\mathrm{DL}}$ increases with increasing study size [40, 41] and that it is associated with low power when a small number of studies are included in the meta-analysis [34]. Empirical evidence suggests care is also needed with the interpretation of $I^2_{\mathrm{DL}}$ when a meta-analysis includes roughly fewer than 500 events and that 95 % confidence intervals for $I^2_{\mathrm{DL}}$ have on average a good coverage [42].

#### $I^2$ Index Based on Generalized $Q$-Statistic

The $I^2$ based on generalized $Q$-statistic is a special form of the "generalized $I^2$ statistics" expressed as [37]

$$I^2_{\mathrm{PM}} = \frac{\hat{\tau}^2_{\mathrm{PM}}}{\hat{\tau}^2_{\mathrm{PM}} + \sigma^2}$$

where $\hat{\tau}^2_{\mathrm{PM}}$ is the estimated between-study variance using the Paule and Mandel estimator (see Sect. 12.2.4.1) [5, 43]. A simulation study suggested that the confidence interval for $I^2_{\mathrm{PM}}$ is wider compared to those of $I^2_{\mathrm{DL}}$ and that $I^2_{\mathrm{PM}}$ maintains coverage close to the nominal level in contrast to $I^2_{\mathrm{DL}}$ method [37].

### 12.2.3.3 $R^2$ Index

An alternative to $H^2$ and $I^2$ measures is the $R^2$ statistic that describes the quadratic inflation in the confidence interval for the summary intervention effect under the random-effects model compared to that from the fixed-effect model

$$R^2 = \frac{\mathrm{Var}(\mu_{\mathrm{RE}})}{\mathrm{Var}(\mu_{\mathrm{RE}})}$$

where $\mu_{\mathrm{RE}}$ is the overall intervention effect under the random-effects model with weights $a_i = 1/(v_i + \hat{\tau}^2)$ and $\mu_{\mathrm{FE}}$ the overall intervention effect under the fixed-effect model with weights $a_i = 1/v_i$. The statistic takes values within the range $(1, \infty)$, and 1 suggests identical inferences under the two meta-analysis models and homogeneity across the study-specific effects. It should be noted that $R^2$ and $H^2$ are equal when all study-specific estimates have equal precision. Since $R^2$ is a function

of $\hat{\tau}^2$ alone (the weights are assumed to be known), one approach to estimate the confidence interval for $R^2$ is via the calculation of the confidence interval for $\tau^2$. However, note that approaches based on the Cochran's $Q$-statistic may not be applicable for constructing confidence intervals for $R^2$.

#### 12.2.3.4 $D^2$ Index

Wetterslev et al. proposed the $D^2$ statistic to quantify the relative variance when we change from the random-effects model to the fixed-effect model [33]. The statistic is interpreted as the proportion of the between-study heterogeneity in meta-analysis relative to the total model variance of the included studies and is given by

$$D^2 = \frac{\mathrm{Var}\left(\mu_{\mathrm{RE}}\right) - \mathrm{Var}\left(\mu_{\mathrm{FE}}\right)}{\mathrm{Var}\left(\mu_{\mathrm{RE}}\right)} = 1 - \frac{1}{R^2}$$

or equivalently

$$D^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \sigma_D^2},$$

where

$$\sigma_D^2 = \frac{\hat{\tau}^2 \left(\mathrm{Var}\left(\mu_{\mathrm{FE}}\right)\right)}{\mathrm{Var}\left(\mu_{\mathrm{RE}}\right) - \mathrm{Var}\left(\mu_{\mathrm{FE}}\right)}$$

is the sampling error. Although $D^2$, similar to $I^2$, is interpreted as a percentage (taking values between 0 and 1), a simulation study suggested that $D^2$ is equal to or greater than $I^2$, irrespective of the chosen effect measure and number of studies in the meta-analysis [33].

#### 12.2.3.5 $G^2$ Index

Rücker et al. proposed an alternative statistic, called $G^2$, to measure between-study heterogeneity while adjusting for small-study effects (see also Sect. 12.4) [31]. The statistic can be obtained by

$$G^2 = 1 - \frac{\left[\sum a_i y_i^{\mathrm{Adj}} - \frac{1}{k}\left(\sum \sqrt{a_i}\right)\left(\sum \sqrt{a_i}\, y_i^{\mathrm{Adj}}\right)\right]^2}{\left[\sum a_i - \frac{1}{k}\left(\sum \sqrt{a_i}\right)^2\right]\left[\sum a_i \left(y_i^{\mathrm{Adj}}\right)^2 - \frac{1}{k}\left(\sum \sqrt{a_i}\, y_i^{\mathrm{Adj}}\right)^2\right]},$$

where $y_i^{\mathrm{Adj}}$ are the study-specific intervention effect estimates adjusted for small-study effects, $y_i^{\mathrm{Adj}} = \mu_{\mathrm{RE}}^{\mathrm{Adj}} + \sqrt{\hat{\tau}^2 / \left(v_i + \hat{\tau}^2\right)}\left(y_i - \mu_{\mathrm{RE}}^{\mathrm{Adj}}\right)$, with $\mu_{\mathrm{RE}}^{\mathrm{Adj}}$ the summary intervention effect under the random-effects model and adjusted for small-study effects, and $a_i = 1/v_i$.

The $G^2$ statistic is closely related to the $Q$-statistic adjusted for small-study effects (see Sect. 12.2.2.3), and it is suggested to quantify heterogeneity in the presence of small-study effects [31]. Similarly to $I^2$ and $D^2$, $G^2$ is interpreted as a percentage (taking values between 0 and 1) and reflects the proportion of the variability in the intervention effect that is not explained under the fixed-effect model that allows for the presence of small-study effects.

### 12.2.4 Estimating the Between-Study Variance

An important aspect in meta-analysis is to quantify the extent of between-study heterogeneity. The DerSimonian and Laird (DL) between-study variance estimator is the most commonly implemented approach and is the default approach in many statistical software (e.g., RevMan) [39, 44]. However, its use has often been criticized because the method may underestimate the true between-study variance, thereby producing narrow confidence intervals (CIs) for the overall intervention effect, especially for a small number of studies (e.g., $k < 10$) [45]. Hence, several alternative methods have been proposed that vary in popularity and complexity. The estimators for $\tau^2$ are categorized as closed form and iterative methods, and their families presented in the literature to date are:

1. The method of moments estimators (e.g., DL and Paule and Mantel (PM)) [39, 43]
2. The maximum likelihood estimators (e.g., maximum likelihood (ML) [20, 46] and restricted maximum likelihood (REML) [46])
3. The model error variance estimators (e.g., Sidik and Jonkman method) [47]
4. The Bayes estimators (e.g., Rukhin Bayes, full Bayes) [48, 49]
5. The bootstrap estimators [50]

It has been shown that estimating the between-study variance in meta-analyses including only a few studies is particularly inaccurate [50–52]. Therefore, it is recommended to quantify the uncertainty around the point estimates to avoid misleading results. Again, several options exist to quantify the uncertainty in the estimated amount of the between-study variance [20, 24, 53].

In this chapter, we briefly describe the most popular estimators for the between-study variance, as well those recommended for the most frequently encountered meta-analysis. For a comprehensive overview of methods used for estimating the between-study variance and its uncertainty, see Veroniki et al. [5].

#### 12.2.4.1 Approaches for the Between-Study Variance Point Estimate

##### Method of Moments Estimators

The generalized method of moments (GMM) estimator [23] can be derived by equating $Q_a$ (see Sect. 12.2.2) and its expected value:

$$E(Q_a) = \left( \sum a_i v_i - \frac{\sum a_i^2 v_i}{\sum a_i} \right) + \tau^2 \left( \sum a_i - \frac{\sum a_i^2}{\sum a_i} \right)$$

Then, solving for $\tau^2$, we obtain

$$\hat{\tau}^2_{\text{GMM}} = \max \left\{ 0, \frac{Q_a - \left( \sum a_i v_i - \frac{\sum a_i^2 v_i}{\sum a_i} \right)}{\sum a_i - \frac{\sum a_i^2}{\sum a_i}} \right\}$$

The method of moments estimators presented in the following subsections is a special case of the GMM estimator with varying weights $a_i$.

DerSimonian and Laird (DL)

This method is the most frequently used approach for the estimation of the between-study variance, and many software programs have DL as the default method. The DL estimator is a non-iterative method and is a special case of the GMM estimators with study weights $a_i = 1/v_i$.

Simulation studies have suggested that the DL method performs well when the true between-study variance is small or close to zero and the number of studies in the meta-analysis is large, whereas when $\tau^2$ is large, DL produces estimates with significant negative bias [37, 47, 52, 54–56]. The negative bias that has been reported with respect to the DL estimator seems to be something related to using effect size measures based on $2 \times 2$ table data (e.g., odds ratios, risk ratios), where problems arise when using very large $\tau^2$ values in simulation studies. In particular, very large $\tau^2$ can lead to extreme values of the effect size measure, at which point many tables will include zero cells and the accuracy and applicability of the inverse-variance method becomes questionable. Jackson et al. evaluated the efficiency of the DL estimator asymptotically and showed that DL is inefficient when the studies included in the meta-analysis are of different sizes and particularly when $\tau^2$ is large [57]. However, they suggested that the DL estimator performs well and can be efficient for inference on the summary effect when the number of studies included in the meta-analysis is large. The confidence interval for the between-study variance when using the DL method can be ideally estimated using the Jackson's method [24], as they are based on the same statistical principle and are naturally paired.

Paule and Mandel (PM)

Paule and Mandel [43] proposed to profile the generalized $Q$-statistic (see Sect. 12.2.2.2) until $Q_{\text{gen}}$ equals its expected value (i.e., $E(Q_{\text{gen}}) = k - 1$). The PM estimator is an iterative method and a special case of the GMM estimator with $a_i = 1/(\tau^2 + v_i)$.

Rukhin et al. showed that when assumptions underlying the method do not hold, the method is more robust than the DL estimator, which depends on large sample

sizes [58]. It has been shown that the PM method has upward bias for a small number of studies and heterogeneity and downward bias for large number of studies and heterogeneity [52], but generally the method is less biased than its alternatives. One simulation study suggested that PM outperforms the DL and REML (see below) estimators in terms of bias [59]. Panityakul et al. [59] showed that the PM estimator is approximately unbiased for large sample sizes, and Bowden et al. [37] in their empirical study showed that as heterogeneity increases, $\hat{\tau}^2_{\text{PM}}$ becomes greater than $\hat{\tau}^2_{\text{DL}}$. The uncertainty around the between-study variance using the PM method can be ideally estimated using the *Q*-profile method [53], as they are based on the same statistical principle and are naturally paired.

### Maximum Likelihood Estimators

The maximum likelihood estimators are iterative methods and are derived after maximizing the (restricted) log-likelihood function [20, 60]. A limitation of the methods is that their success to converge to a solution depends on the selection of the maximization technique (e.g., Newton-Raphson, expectation-maximization algorithm).

#### Maximum Likelihood (ML)

The method is asymptotically efficient and can be obtained by iterating

$$\hat{\tau}^2_{\text{ML}} = \max\left\{0, \frac{\sum w^2_{i,\text{RE}}\left(\left(y_i - \mu_{\text{RE}}\left(\hat{\tau}^2_{\text{ML}}\right)\right)^2 - v_i\right)}{\sum w^2_{i,\text{RE}}}\right\}$$

and

$$\mu_{\text{RE}}\left(\hat{\tau}^2_{\text{ML}}\right) = \frac{\sum w_{i,\text{RE}} y_i}{\sum w_{i,\text{RE}}}$$

until they converge and do not change from one iteration to the next. The study weights are derived under the random-effects model, $w_{i,\text{RE}} = 1/\left(v_i + \hat{\tau}^2_{\text{ML}}\right)$. An initial estimate of $\hat{\tau}^2_{\text{ML}}$ can be decided a priori as a plausible value of the heterogeneity variance, or it can be estimated with any other non-iterative estimation method. Each iteration step requires nonnegativity.

Simulation studies have suggested that although the ML estimator is efficient, it exhibits large negative bias for large $\tau^2$ when the number and size of studies are small (e.g., for fewer than 10 studies and fewer than 80 participants in each study) [50–52, 56, 59]. It has been shown that the ML method is more efficient than PM, and REML methods, but exhibits the largest amount of bias [51, 52, 60, 61]. However, because of the large amount of bias, it is recommended avoiding the ML estimator [56, 59]. The confidence interval for the between-study variance when using the ML method can be ideally computed using the profile likelihood method [1], as they are based on the same statistical principle and are naturally paired.

Restricted Maximum Likelihood (REML)

The REML method is often used to correct for the negative bias produced by the ML method and can be obtained by

$$\hat{\tau}^2_{\text{REML}} = \max\left\{0, \frac{\sum w^2_{i,\text{RE}}\left(\left(y_i - \mu_{\text{RE}}(\hat{\tau}^2_{\text{REML}})\right)^2 - v_i\right)}{\sum w^2_{i,\text{RE}}} + \frac{1}{\sum w_{i,\text{RE}}}\right\},$$

with study weights derived under the random-effects model, $w_{i,\text{RE}} = 1/(v_i + \hat{\tau}^2_{\text{REML}})$ [39, 52]. The estimator is calculated by an iterative process with a nonnegative initial estimate. Again, each iteration step requires nonnegativity.

Simulation studies suggested that the REML method underestimates the true between-study variance, especially when the data are sparse [47, 52, 54, 56, 62]. For dichotomous outcome data, it was shown that the REML estimator is less biased, but less efficient than the DL estimator [51, 52]. For continuous data, it has been suggested that the REML estimator is less efficient than the ML estimator and comparable to DL estimator [56]. An empirical study [63] with dichotomous outcome data showed that the REML estimator can be smaller or larger in magnitude than the DL method. REML is recommended when large studies are included in the meta-analysis [56]. The uncertainty around the between-study variance when using the REML estimator can be ideally estimated using the profile likelihood method [20].

### Bayes Estimators

Full Bayes (FB)

The FB approach takes into account the uncertainty of all parameters (including $\tau^2$) in the results. Several investigators claim that in practice the differences between frequentist and Bayesian approaches appear to be small [60, 64]. The FB method uses non-informative priors to approximate a likelihood-based analysis. When the number of studies is large, the choice of the prior does not have a major influence on the results since they are data driven. The choice of prior is particularly important though when the number of studies is small, as it may impact on the estimated between-study variance and hence on the overall intervention effect [65, 66].

A simulation study compared 13 different prior distributions for the heterogeneity variance and suggested that the results might vary substantially when the number of studies is small [65]. The study showed that, in terms of bias, none of the distributions considered performed best for all meta-analysis scenarios. More specifically, inverse-gamma, uniform, and Wishart distributions for the between-study variance all perform poorly when the number of studies is small (<10) and produce estimates with substantial bias. An inverse-gamma prior with small hyper-parameters is often considered to be an approximately non-informative prior, but it was shown that inferences can be sensitive to the choice of hyper-parameters [67, 68]. Informative priors were recently proposed for the between-study variance using the log-odds ratio and standardized mean difference effect measures, and these might

considerably improve estimation when few studies are included in the meta-analysis [69–71]. The uncertainty around the between-study variance when using the FB estimator can be ideally estimated using Bayesian credible intervals.

## 12.3 Possible Causes and Approaches to Deal with Heterogeneity

Despite the best efforts of investigators to construct a dataset of carefully selected studies where the homogeneity assumption would hold, an imbalance in the distribution of effect modifiers might arise resulting in between-study heterogeneity. The identification of the causes of heterogeneity may help to account for such variation in the results thereby aiding in the interpretation of existing data, as well the planning of future studies. Between-study heterogeneity may be due to clinical and/or methodological heterogeneity, biases, and chance [3, 72]. Clinical heterogeneity suggests that a possible variability in intervention or patient-level characteristics, or in outcomes studied, can influence the intervention effect. Methodological heterogeneity refers to the variability across studies due to study design or quality (e.g., inadequate randomization or allocation concealment, high dropout rates, intention-to-treat versus per-protocol analyses). In addition to biases captured by methodological heterogeneity, there are other biases that might cause between-study heterogeneity, including selection or funding biases. It is also possible that outlier studies show extreme results due to chance (e.g., studies with small sizes and/or event rates).

Quantifying the amount of between-study heterogeneity and exploring its sources are among the most important aspects of meta-analysis. When heterogeneity is identified, the first step researchers should follow is to check the data included in the meta-analysis for potential data abstraction errors. If no errors are found and between-study variability beyond chance is still evident, a different choice in effect measure may improve homogeneity. Empirical studies have shown that relative measures (e.g., odds ratio, risk ratio) are associated with less heterogeneity than absolute measures (e.g., risk difference) [73–75]. Heterogeneity might also be due to intervention effect modifiers. This exploration might include applying subgroup or meta-regression analyses adjusting the estimated intervention effects accordingly. It should be noted that the use of individual patient data in meta-analysis allows for a thorough investigation of potential sources of heterogeneity and a better evaluation of both within- and between-study heterogeneity, avoiding the assumption that a relationship between groups holds between individuals as well [76, 77]. For small to moderate amount of heterogeneity (for a general guideline, see Sect. 12.2.3.2), one can apply the random-effects model assuming that the true study-specific effects are not identical but come from the same distribution. Under the random-effects model, the between-study variation is taken into account in the meta-analysis results, but this is not a remedy for heterogeneity as it still exists.

To facilitate the interpretation of the meta-analysis' result capturing both between-study variance and variance of summary intervention effect, a prediction interval of the possible intervention effect in an individual setting can be calculated [78–80].

A prediction interval indicates the range of values for the true intervention effect when a future study is conducted and can be obtained by

$$\mu_{\mathrm{RE}} \pm t_{1-\frac{a}{2},k-2}\sqrt{\hat{\tau}^2 + \mathrm{var}\left(\mu_{\mathrm{RE}}\right)}$$

where $t_{1-a/2,\,k-2}$ is the $100(1-a/2)\,\%$ quantile of the $t_{k-2}$ distribution. A prediction interval can be calculated when at least three studies are included in the meta-analysis.

## 12.4 Methods to Appraise Small-Study Effects

The association between size and effect of the studies included in a meta-analysis should be explored, as the presence of selection bias and small-study effects may lead to meaningless conclusions. Funnel plots and statistical tests based on funnel plot asymmetry are popular in meta-analysis for assessing small-study effects. Several methods have been suggested to adjust for small-study effects, including the trim-and-fill method, the Copas selection model, and various regression-based approaches (for a review, see Mavridis and Salanti) [6].

### 12.4.1 Graphical Representation of Small-Study Effects

Funnel plots facilitate the visual examination for detecting bias or heterogeneity, and often it is not possible to distinguish between the two. A funnel plot (see Fig. 12.5) is a scatter plot of the study-specific intervention effect estimates against a measure of precision or study size. In agreement with forest plots (see Sect. 12.2.1.1) and in contrast to conventional scatter plots, the intervention effect estimates are usually plotted on the *x*-axis, whereas the study size or precision is plotted on the *y*-axis [82–84]. It is recommended to plot the SE (or 1/SE) of the intervention effect

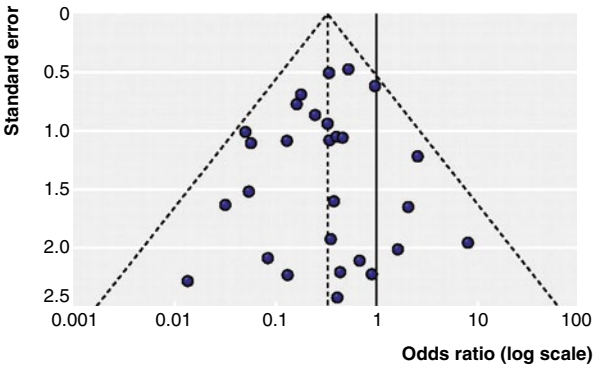



**Fig. 12.5** Funnel plot. Example of symmetrical funnel plot (Reproduced with permission [81])

on the vertical axis, rather than study size, as study power is based on several other factors apart from sample size alone (e.g., number of events, standard deviation) [84], and these are summarized by SE. The plot usually includes a triangular 95 % confidence region and a vertical line corresponding to summary intervention effect under the fixed-effect model. In the absence of bias and heterogeneity, 95 % of the studies are expected to lie within the triangular region and be scattered symmetrically around the summary intervention effect. In such a case, the plot resembles a symmetrical and inverted funnel. Small studies are expected to lie at the bottom of the graph and widely spread around the summary intervention effect compared to larger studies. It is advisable to draw funnel plots when ten or more studies are available in the meta-analysis [7].

An asymmetric funnel plot suggests there is a relationship between the study-specific effect measure and precision, which might be due to selection bias (including publication bias, language bias, citation bias, and reporting bias), small-study effects, heterogeneity, sampling variation, or chance [10]. An inappropriate choice of effect measure might also result in an asymmetrical funnel plot. It should be noted that some effect measures (e.g. log-odds ratios and standardized mean differences) are correlated with their SEs, and this may produce artificial funnel plot asymmetry. In the presence of small-study effects, the funnel plot will be asymmetrical with small studies missing at the bottom right corner (for an efficacy outcome, and at the left corner for a safety outcome) suggesting an unfavorable effect. Some argue that the visual interpretation of a funnel plot is a subjective issue, and sometimes it is difficult to distinguish between symmetry and asymmetry [85, 86].

Peters et al. proposed a modified version of the conventional funnel plot, in which extra contours representing the statistical significance of each study are added (see Fig. 12.6) [87]. This may aid visual interpretation by suggesting that if the missing studies come from a "nonsignificance area," then asymmetry may be due to selection bias. However, if the missing studies come from a "significance-area" or

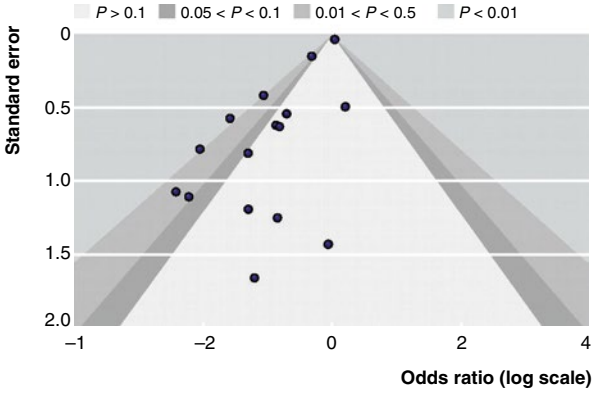



**Fig. 12.6** Contour-enhanced funnel plot for trials of the effect of intravenous magnesium on mortality after myocardial infarction. Example of asymmetrical funnel plot (Reproduced with permission [81])

there is a certain direction of the intervention effect, then asymmetry is probably due to factors other than selection bias [81].

### 12.4.2 Tests for Small-Study Effects and Selection Bias

#### 12.4.2.1 Funnel Plot-Based Tests

Apart from assessing for small-study effects using a visual inspection of funnel plots, several tests have been suggested to statistically assess funnel plot asymmetry. The tests are categorized as (1) rank-correlation tests or (2) linear regression tests. Begg and Mazumdar used a nonparametric rank-correlation method for the examination of the association between the standardized intervention effect estimates and their SEs [88]. When small studies (with large SEs) tend to have larger intervention effect estimates compared to the larger studies, the test identifies a correlation between the two factors. However, the test is associated with low power, and Begg suggests using a very liberal significance level (such as 0.10) [89]. Gjerdevik and Heuch suggested modification of Begg test based on Spearman rho and Kendall tau, to improve type I and II error rates; they suggested that the test based on Spearman rho is preferred for small datasets [90]. Egger et al. proposed a more powerful test compared to Begg test to assess the funnel plot asymmetry based on a regression analysis of Galbraith plot (see also Sect. 12.2.1.2) [83]. The test is based on the weighted linear regression of the standardized intervention effect (z-score) against study precision, with weights equal to the inverse of the variance. The intercept of the regression is used to measure asymmetry; specifically if it is estimated to be statistically significantly different from 0, then there is evidence of selection bias, and a negative intercept would suggest small-study effects are present. Tang and Liu suggested an alternative test using a linear regression of intervention effect estimate on $1/\sqrt{n}$, with weights $n$ the study size [91].

Several modifications of the tests have been presented in the literature, which apply to dichotomous outcome data only. More specifically, for group correlation, the test by Schwarzer et al. could be used [92]. For linear regression, several modifications have been proposed including those by Macaskill et al. [93], Harbord et al. [94], Peters et al. [95], and the "arcsine" test by Rücker et al. [96]. For all aforementioned tests, the cutoff $P$-value 0.10 is considered to infer asymmetry in the funnel plot.

More specifically, the test proposed by Macaskill et al. is a linear regression of the intervention effect estimate on $n$, with weights $m_E m_{NE}/n$, where $m_E$ and $m_{NE}$ represent the total number of events and nonevents, respectively [93]. Harbord et al. [94] presented a modified version of the test proposed by Egger et al. [83], based on the efficient score ( $Z = a - m_E n_E / n$ ) and its variance ( $V = n_E n_C m_E m_{NE} / n^2 (n-1)$ ) of the log-odds ratio, where $n_E$ and $n_C$ are the sample sizes of the experimental and control groups, respectively. Peters et al. [95] suggested a slightly modified test compared to Macaskill et al. [93] test using the log-odds ratio effect measure and a linear regression of intervention effect estimate on $1/n$, with weights $m_E m_{NE}/n$, for a better control of type I error. Schwarzer et al. [92] suggested a rank-correlation test

for sparse data, using mean and variance of the noncentral hypergeometric distribution and avoiding correlation between log-odds ratio and its SE. However, for large between-study heterogeneity, the test has low power compared to the other tests [92]. Although the tests by Harbord et al. [94], Peters et al. [95], and Schwarzer et al. [92] have been presented using the odds ratio effect measure, they can be applied for other effect measures too. However, for a dichotomous outcome and the log-odds ratio or log-risk ratio, the intervention effect is statistically dependent on its variance, and hence tests based on these two factors might erroneously suggest the small-study effects' presence. Rücker et al. [96] suggested a test based on arcsine transformation of observed risks avoiding false-positive results when a large intervention effect or substantial between-study heterogeneity is present. In contrast to the other tests, the one suggested by Rücker et al. [96] can model studies with zero events in both arms.

Sterne et al. [81] advise using regression tests to address selection bias and small-study effects as they have larger power compared to rank tests as well as avoiding tests for funnel plot asymmetry if all studies are of similar sizes and hence of similar SEs. The Egger test has greater power for continuous outcomes than for dichotomous outcomes and is suggested for testing for funnel plot asymmetry. For dichotomous outcomes, the Harbord, Peters, and Rücker tests are suggested, as they have greater power compared to the other tests and avoid the mathematical association between log-odds ratio and its SE (this is also known as "regression to the mean"). It should be noted though that the performance of the tests deteriorates as the between-study heterogeneity increases. A general recommendation is to select one of the Harbord, Peters, and Rücker tests for small heterogeneity ($\tau^2 < 0.1$) and to use Rücker test for large heterogeneity ($\tau^2 > 0.1$) [3, 81].

### 12.4.3 Adjusting Intervention Effect Estimates for Small-Study Effects

#### 12.4.3.1 Trim-and-Fill Method

The trim-and-fill method is a nonparametric method and aims to correct for funnel plot asymmetry due to small-study effects. The method is a four-step process:

1. The smaller studies are "trimmed" (i.e., removed) so that a symmetrical funnel plot is produced.
2. The summary intervention effect from the "trimmed" funnel plot is estimated.
3. The omitted studies are returned to the funnel plot and their "missing counterparts" are imputed or "filled" as their mirror images.
4. An adjusted overall intervention effect with its corresponding confidence interval is estimated using the complete set of studies [97, 98].

This is a nonparametric method and provides an estimate of both the number of missing studies and of the summary intervention effect adjusted for selection bias.

Although no assumptions are required about the mechanism leading to selection bias, the trim-and-fill method assumes that the small-study effect is solely caused by selection bias and that in truth there should be a symmetric funnel plot. However, the adjusted intervention effect should be interpreted with caution as it is not necessarily the intervention effect that would have been observed in the absence of selection bias.

Simulation studies have shown that the method performs well in the presence of selection bias, but it underestimates the intervention effect when there is large between-study heterogeneity and no selection bias [99, 100].

#### 12.4.3.2 Selection Models

To evaluate the potential impact of missing studies on the results of a meta-analysis, selection models have been suggested that account for the mechanism by which studies are published. Selection models assume that missing studies are not missing at random, and the observed studies are due to certain characteristics (e.g., sample size, quality of design) that increase their propensity for publication. These models associate each observed study with an a priori probability to be published, and then estimate the summary intervention effect from the distribution of the observed sample.

A popular selection model in meta-analysis is the one developed by Copas [101], in which the probability that a study is observed depends on its SE. Although selection models correct effect estimates for selection bias, they have not been widely used probably because of their complexity, the large number of studies needed and the strong modeling assumptions about the severity of selection bias (i.e., that the factor causing small-study effects is selection bias). Copas [101] suggested applying a sensitivity analysis so that the researcher has the full picture of the estimated values of the intervention effect (and its uncertainty) under a range of assumptions about the severity of selection bias. It has been alternatively suggested to use expert opinion to inform the probabilities of publication [102]. A Copas selection model accounts for the correlation between the observed intervention effect and the probability that a study is published, which is:

1. Zero in the absence of selection bias.
2. Positive for a large intervention effect and large propensity for publication (e.g., for safety outcomes).
3. Negative for a large intervention effect and small propensity for publication (e.g., for efficacy outcomes; harms are less likely to be studied in trials and hence less likely to be published) [101, 103, 104].

Empirical studies using large collections of meta-analyses with dichotomous data suggest that the Copas selection model is preferable than the trim-and-fill method, as the latter produces systematically larger SEs and *P*-values [105, 106].

#### 12.4.3.3 Extrapolation Methods

Extrapolation approaches model the relationship between the observed intervention effects and a measure of their uncertainty (e.g., SE). Stanley [107] and Copas and Malley [108] are early proponents of the regression-based approaches, with Stanley [107] adjusting the estimated intervention effect and Copas and Malley [108] adjusting the *P*-values for small-study effects. The approach suggested by Moreno et al. [109, 110] regresses the study-specific effects against their precision and computes the "unbiased" intervention effect as the extrapolation of the regression line to predict the intervention effect in a study with infinite sample size (or zero SE). The slope of the meta-regression is used to test for funnel plot asymmetry (see also Sect. 12.4.2.1), and the intercept is interpreted as the estimated intervention effect of a study with infinite sample size and hence infinite precision, adjusted for selection bias.

A key concern in these methods, as already stated in Sect. 12.4.2.1, is the mathematical association between some effect measures (e.g., log-odds ratio) and its SE, which might erroneously suggest the presence of small-study effects. Also, the performance of these methods depends on the variability of the meta-analysis' study sizes; if, for example, all studies are small, then the methods will not perform well. The regression-based methods, as any meta-regression model, suffer from lack of power to detect existing associations when few studies are available and in the presence of substantial heterogeneity. Simulation studies suggest that extrapolation within funnel plots outperform the trim-and-fill method, but still the adjusted effect estimates should be interpreted with caution [4, 109].

## 12.5 Moderators and Confounders

The impact of moderators and confounders is best viewed in light of the prior sections on heterogeneity issues and small-study effects, as any meaningfully important moderator or confounder is likely going to have an impact on homogeneity and symmetry of effects. The typical approaches to moderator and confounders include subgroup analyses and regression methods, which can be undertaken in the context of meta-analysis as well as more comprehensive overviews of reviews. As always, it remains important to recognize the presence of clustering and to minimize, especially in umbrella reviews, the risk of duplicate entry of trials with multiple arms as this may have a biasing effect on the accuracy and precision of the overall estimates.

## 12.6 Discussion

This chapter illustrates a vast range of approaches to evaluate the presence and estimate the magnitude of between-study heterogeneity as well as a wide variety of methods to test and adjust for small-study effects, which can easily be extrapolated to the analysis of key moderators and confounders. Heterogeneity and selection bias

are two of the greatest threats in meta-analysis and may lead to meaningless and/or overoptimistic intervention effect estimates. Researchers should routinely address and explore reasons for their presence and assess the extent to which these may influence the meta-analysis results.

Recent methodological research supports use of the random-effects model when completing a meta-analysis because it accounts for the between-study heterogeneity [3, 111, 112]. The random-effects model is considered more realistic than the fixed-effect model in most contexts. The new methodologies in meta-analysis help us incorporate heterogeneity and adjust for small-study effects and general funnel plot asymmetries as parts of the modeling that can also be reflected in the results. As presented in this chapter, both heterogeneity and selection bias can be examined using graphical methods, statistical tests, subgroup, and meta-regression analyses.

When selection bias is present, it is advisable that researchers make efforts to reduce (or if possible to eliminate) it, such as identifying unpublished or difficult to locate material from the "gray" literature for potential inclusion in the meta-analysis [113]. Also, exploration of heterogeneity should always take place when conducting a meta-analysis but should be interpreted with caution if individual participant data is not used in the statistical modeling. When few studies are included in a meta-analysis, we suggest conducting a sensitivity analysis using a variety of methods for addressing heterogeneity and small-study effects, before reaching definitive conclusions.

**Acknowledgements** AAV is funded by the CIHR Banting Postdoctoral Fellowship Program.

We would also like to thank Dr. Sharon E. Straus for her comments on a previous draft of this chapter.

## References

1. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol. 2000;53(11):1119–29.
2. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Random-effects model. Introduction to meta-analysis [Internet]. John Wiley & Sons, Ltd; 2009 [cited 2014]. p. 69–75. Available from: http://onlinelibrary.wiley.com/doi/10.1002/9780470743386.ch12/summary.
3. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions [Internet]. Version 5.1.0. The Cochrane Collaboration; 2011. Available from: www.cochrane-handbook.org.
4. Rücker G, Carpenter JR, Schwarzer G. Detecting and adjusting for small-study effects in meta-analysis. Biom J. 2011;53(2):351–68.
5. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, Kuss O, Higgins JPT, Langan D, Salanti G. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Method. 2015. doi: 10.1002/jrsm.1164.
6. Mavridis D, Salanti G. Exploring and accounting for publication bias in mental health: a brief overview of methods. Evid Based Ment Health. 2014;17(1):11–5.
7. Anzures-Cabrera J, Higgins JPT. Graphical displays for meta-analysis: an overview with suggestions for practice. Res Synth Methods. 2010;1(1):66–80.

8. Bax L. MIX 2.0 – professional software for meta-analysis in Excel [Internet]. BiostatXL; 2011. Available from: http://www.meta-analysis-made-easy.com.
9. Tricco AC, Alateeq A, Tashkandi M, Mamdani M, Al-Omran M, Straus SE. Histamine H2 receptor antagonists for decreasing gastrointestinal harms in adults using acetylsalicylic acid: systematic review and meta-analysis. Open Med. 2012;6(3):e109–17.
10. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ. 1997;315(7121):1533–7.
11. Ried K. Interpreting and understanding meta-analysis graphs – a practical guide. Aust Fam Physician. 2006;35(8):635–8.
12. Moja L, Moschetti I, Liberati A, Gensini GF, Gusinu R. Understanding systematic reviews: the meta-analysis graph (also called "forest plot"). Intern Emerg Med. 2007;2(2):140–2.
13. DuMouchel W. Predictive cross-validation of Bayesian meta-analyses. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. Bayesian statistics 5. Oxford: Oxford University Press; 1996. p. 107–27.
14. Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. Stat Methods Med Res. 1993;2(2):173–92.
15. Galbraith RF. Some applications of radial plots. J Am Stat Assoc. 1994;89(428):1232–42.
16. Galbraith RF. A note on graphical presentation of estimated odds ratios from several clinical trials. Stat Med. 1988;7(8):889–94.
17. Song F. Exploring heterogeneity in meta-analysis: is the L'Abbé plot useful? J Clin Epidemiol. 1999;52(8):725–30.
18. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med. 1987;107(2):224–33.
19. Baujat B, Mahé C, Pignon J-P, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med. 2002;21(18):2641–52.
20. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. Stat Med. 1996;15(6):619–29.
21. Viechtbauer W. Hypothesis tests for population heterogeneity in meta-analysis. Br J Math Stat Psychol. 2007;60(1):29–60.
22. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954;10(1):101–29.
23. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. Contemp Clin Trials. 2007;28(2):105–14.
24. Jackson D. Confidence intervals for the between-study variance in random effects meta-analysis using generalised Cochran heterogeneity statistics. Res Synth Methods. 2013;4(3):220–9.
25. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. Stat Med. 1997;16(7):753–68.
26. Biggerstaff BJ, Jackson D. The exact distribution of Cochran's heterogeneity statistic in one-way random effects meta-analysis. Stat Med. 2008;27(29):6093–110.
27. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. Stat Med. 1998;17(8):841–56.
28. Fleiss JL. Analysis of data from multiclinic trials. Control Clin Trials. 1986;7(4):267–75.
29. Petitti DB. Approaches to heterogeneity in meta-analysis. Stat Med. 2001;20(23):3625–33.
30. Barbui C, Hotopf M, Freemantle N, Boynton J, Churchill R, Eccles MP, et al. WITHDRAWN: treatment discontinuation with selective serotonin reuptake inhibitors (SSRIs) versus tricyclic antidepressants (TCAs). Cochrane Database Syst Rev. 2006;3:CD002791.
31. Rücker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. Biostatistics. 2011;12(1):122–42.
32. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–58.
33. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol. 2009;9(1):86.
34. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 2006;11(2):193–206.

35. Zou GY. On the estimation of additive interaction by use of the four-by-two table and beyond. Am J Epidemiol. 2008;168(2):212–24.
36. Donner A, Zou GY. Closed-form confidence intervals for functions of the normal mean and standard deviation. Stat Methods Med Res. 2012;21(4):347–59.
37. Bowden J, Tierney JF, Copas AJ, Burdett S. Quantifying, displaying and accounting for heterogeneity in the meta-analysis of RCTs using standard and generalised Q statistics. BMC Med Res Methodol. 2011;11(1):41.
38. Birge RT. The calculation of errors by the method of least squares. Phys Rev. 1932;40(2):207–27.
39. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177–88.
40. Mittlböck M, Heinzl H. A simulation study comparing properties of heterogeneity measures in meta-analyses. Stat Med. 2006;25(24):4321–33.
41. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol. 2008;8:79.
42. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, et al. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLoS One. 2012;7(7):e39471.
43. Paule RC, Mandel J. Consensus values and weighting factors [Internet]. National Institute of Standards and Technology; 1982 [cited 25 Mar 2014]. Available from: http://archive.org/details/jresv87n5p377_A1b.
44. The Nordic Cochrane Centre. Review manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Cochrane Collaboration; 2014.
45. Cornell JE, Mulrow CD, Localio R, Stack CB, Meibohm AR, Guallar E, et al. Random-effects meta-analysis of inconsistent effects: a time for change. Ann Intern Med. 2014;160(4):267–70.
46. Raudenbush SW. Analyzing effect sizes: Random-effects models. In: Cooper H, Hedges LV, Valentine JC, (Eds.). The handbook of research synthesis and meta-analysis (2nd ed). New York: Russell Sage Foundation; 2009.
47. Sidik K, Jonkman JN. Simple heterogeneity variance estimation for meta-analysis. J R Stat Soc Ser C Appl Stat. 2005;54(2):367–84.
48. Rukhin AL. Estimating heterogeneity variance in meta-analysis. J R Stat Soc Ser B Stat Methodol. 2013;75(3):451–69.
49. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. Stat Med. 1995;14(24):2685–99.
50. Kontopantelis E, Springate DA, Reeves D. A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. PLoS One. 2013;8(7):e69930.
51. Chung Y, Rabe-Hesketh S, Choi I-H. Avoiding zero between-study variance estimates in random-effects meta-analysis. Stat Med. 2013;32(23):4071–89.
52. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. Stat Med. 2007;26(9):1964–81.
53. Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. Stat Med. 2007;26(1):37–52.
54. Novianti PW, Roes KCB, van der Tweel I. Estimation of between-trial variance in sequential meta-analyses: a simulation study. Contemp Clin Trials. 2014;37(1):129–38.
55. Sidik K, Jonkman JN. A note on variance estimation in random effects meta-regression. J Biopharm Stat. 2005;15(5):823–38.
56. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. J Edu Behav Stat. 2005;30(3):261–93.
57. Jackson D, Bowden J, Baker R. How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? J Stat Plan Inference. 2010;140(4):961–70.
58. Rukhin AL, Biggerstaff BJ, Vangel MG. Restricted maximum likelihood estimation of a common mean and the Mandel–Paule algorithm. J Stat Plan Inference. 2000;83(2):319–30.

59. Panityakul T, Bumrungsup C, Knapp G. On estimating residual heterogeneity in random-effects meta-regression: a comparative study. J Stat Theory Appl. 2013;12(3):253.
60. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med. 1999;18(20):2693–708.
61. William H, Swallow JFM. Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. Technometrics. 1984;26(1):47–57.
62. Goldstein H, Rasbash J. Improved approximations for multilevel models with binary responses. J R Stat Soc Ser A Stat Soc. 1996;159(3):505.
63. Thorlund K, Wetterslev J, Awad T, Thabane L, Gluud C. Comparison of statistical inferences from the DerSimonian–Laird and alternative random-effects model meta-analyses – an empirical assessment of 920 Cochrane primary outcome meta-analyses. Res Synth Methods. 2011;2(4):238–53.
64. Morris CN. Parametric empirical Bayes inference: theory and applications. J Am Stat Assoc. 1983;78(381):47.
65. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med. 2005;24(15):2401–28.
66. Senn S. Trying to be precise about vagueness. Stat Med. 2007;26(7):1417–30.
67. Chung Y, Rabe-Hesketh S, Dorie V, Gelman A, Liu J. A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. Psychometrika. 2013;78(4):685–709.
68. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 2006;1(3):515–34.
69. Rhodes KM, Turner RM, Higgins JPT. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. J Clin Epidemiol. 2015;68(1):52–60.
70. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JPT. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol. 2012;41(3):818–27. doi: 10.1093/ije/dys041.
71. Pullenayegum EM. An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. Stat Med. 2011;30(26):3082–94.
72. Bailey KR. Inter-study differences: how should they influence the interpretation and analysis of results? Stat Med. 1987;6(3):351–60.
73. Friedrich JO, Adhikari NKJ, Beyene J. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. J Clin Epidemiol. 2011;64(5):556–64.
74. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Stat Med. 2000;19(13):1707–28.
75. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med. 2002;21(11):1575–600.
76. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol. 2002;55(1):86–94.
77. Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Stat Med. 2005;24(9):1307–19.
78. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549.
79. Guddat C, Grouven U, Bender R, Skipka G. A note on the graphical presentation of prediction intervals in random-effects meta-analyses. Syst Rev. 2012;1:34.
80. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009;172(1):137–59.
81. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.

82. Light RJ, Pillemer DB. Summing up: the science of reviewing research. Highlightingth ed. Cambridge: Harvard University Press; 1984. 191 p.
83. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629–34.
84. Sterne JAC, Egger M, Smith GD. Investigating and dealing with publication and other biases in meta-analysis. BMJ. 2001;323(7304):101–5.
85. Ioannidis JPA, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. CMAJ. 2007;176(8):1091–6.
86. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol. 2005;58(9):894–901.
87. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol. 2008;61(10):991–6.
88. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994;50(4):1088–101.
89. Cooper H, Hedges LV, editors. The handbook of research synthesis. 1st ed. New York: Russell Sage; 1994. 573 p.
90. Gjerdevik M, Heuch I. Improving the error rates of the Begg and Mazumdar test for publication bias in fixed effects meta-analysis. BMC Med Res Methodol. 2014;14:109.
91. Tang J-L, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. J Clin Epidemiol. 2000;53(5):477–84.
92. Schwarzer G, Antes G, Schumacher M. A test for publication bias in meta-analysis with sparse binary data. Stat Med. 2007;26(4):721–33.
93. Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med. 2001;20(4):641–54.
94. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med. 2006;25(20):3443–57.
95. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA. 2006;295(6):676–80.
96. Rücker G, Schwarzer G, Carpenter J. Arcsine test for publication bias in meta-analyses with binary outcomes. Stat Med. 2008;27(5):746–63.
97. Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J Am Stat Assoc. 2000;95(449):89–98.
98. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455–63.
99. Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. Stat Med. 2003;22(13):2113–26.
100. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. Stat Med. 2007;26(25):4544–62.
101. Copas J. What works?: selectivity models and meta-analysis. J R Stat Soc Ser A Stat Soc. 1999;162(1):95–109.
102. Mavridis D, Sutton A, Cipriani A, Salanti G. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. Stat Med. 2013;32(1):51–66.
103. Copas JB, Shi JQ. A sensitivity analysis for publication bias in systematic reviews. Stat Methods Med Res. 2001;10(4):251–65.
104. Copas J, Shi JQ. Meta-analysis, funnel plots and sensitivity analysis. Biostat Oxf Engl. 2000;1(3):247–62.
105. Schwarzer G, Carpenter J, Rücker G. Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. J Clin Epidemiol. 2010;63(3):282–8.

106. Carpenter JR, Schwarzer G, Rücker G, Künstler R. Empirical evaluation showed that the Copas selection model provided a useful summary in 80% of meta-analyses. J Clin Epidemiol. 2009;62(6):624–31.e4.
107. Stanley TD. Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. Oxf Bull Econ Stat. 2008;70(1):103–27.
108. Copas JB, Malley PF. A robust P-value for treatment effect in meta-analysis with publication bias. Stat Med. 2008;27(21):4267–78.
109. Moreno SG, Sutton AJ, Ades AE, Stanley TD, Abrams KR, Peters JL, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Med Res Methodol. 2009;9(1):2.
110. Moreno SG, Sutton AJ, Ades AE, Cooper NJ, Abrams KR. Adjusting for publication biases across similar interventions performed well when compared with gold standard data. J Clin Epidemiol. 2011;64(11):1230–41.
111. Schmidt FL, Oh I-S, Hayes TL. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. Br J Math Stat Psychol. 2009;62(1):97–128.
112. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. Stat Med. 2001;20(6):825–40.
113. Searching grey literature: grey matters [Internet]. [cited 29 Mar 2015]. Available from: http://www.nccmt.ca/registry/view/eng/130.html.

# State-of-the-Art Reporting

# 13

Akira Onishi and Toshi A. Furukawa

**Abstract**

Reporting overviews of systematic reviews, umbrella reviews, or meta-epidemiologic studies should be transparent, while several empirical evidences showed many overviews of systematic reviews had poor reporting and methodological quality. Reporting overviews of systematic reviews is somewhat different from reporting systematic reviews. In this chapter, we provided a reporting checklist for overviews of systematic reviews and explanations for each checklist item that contains specific considerations on reporting of overviews with the ultimate goal of improving use and application for effective and efficient decision making.

## 13.1 Introduction

Overviews of systematic reviews, umbrella reviews, or meta-epidemiologic studies are relatively new level of researches that synthesize reviews and meta-analysis [1]. This new type of research design assesses different participants, multiple interventions or diagnostic tests, various outcomes, and methodological or nonclinical features (e.g., funding issues, sample size). Overviews of systematic reviews have become increasingly popular and their numbers are increasing [2]. These overviews can offer a broad summary of evidence to decision-makers, help to develop clinical

A. Onishi, MD, MPH, PhD (✉)
Department of Rheumatology, Kobe University Hospital, Kobe, Japan
e-mail: aonishi@med.kobe-u.ac.jp

T.A. Furukawa
Department of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_13

practice guideline or decide health policy, and provide the important clinical issues that remain unresolved for future research.

Therefore, overviews of systematic reviews should be reported fully and transparently as well as systematic reviews. However, there is no reporting and methodological standards for overviews of systematic reviews, while the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-analysis) and the MOOSE (Meta-analysis Of Observational Studies in Epidemiology) proposed a checklist that contains essential items for transparent reporting of systematic reviews of interventions and observational studies, respectively [3, 4]. In addition, there is considerable evidence that important information is often poorly reported in overviews of systematic reviews [2, 5, 6]. This chapter offers a reporting checklist for overviews of systematic reviews which was developed based on the epidemiologic principles and characteristic design of overviews of systematic reviews by reference to recommendations from the Cochrane Collaboration [1], PRISMA statement [3], and AMSTAR (A MeaSurement Toll to Assess Systematic Reviews) [7] (Table 13.1). We also provide differences between reporting of systematic reviews and overviews of reviews with their rationales (Table 13.2). The definitions, purposes, and classifications of overviews of reviews, umbrella reviews, and meta-epidemiologic studies can be seen in the other chapter.

## 13.2 Reporting Checklist

### 13.2.1 Title and Abstract

#### 13.2.1.1 Title

Identify the reports as an overview of (systematic) reviews, an umbrella review, or a meta-epidemiologic study.

**Explanation.** Overview authors should state an overview of (systematic) reviews, an umbrella review, or a meta-epidemiologic study in the title because readers cannot identify whether the review was an overview of review, a systematic review, or a narrative review if only the term such as “review” or “overview” was used.

#### 13.2.1.2 Structured Summary

Provide a structured abstract, consisting of background, objective, data sources, selection criteria, data extraction, review appraisal, data synthesis methods, results, limitations, conclusions, and implication.

**Explanation.** A structure abstract is recommended because readers can get systematic information more easily than an unstructured abstract. A structured abstract includes background, objective, data sources, selection criteria, data extraction, review appraisal, data synthesis methods, results, limitations, conclusions, and implication. Because abstracts should be targeted not only to researchers but also consumers of healthcare, overview authors provide plain language summary that enables readers to understand easily.

**Table 13.1** Reporting checklist

| Section/topic | # | Checklist item |
|---|---|---|
| **Title and Abstract** | | |
| Title | 1 | Identify the reports as an overview of (systematic) reviews, an umbrella review, or a meta-epidemiologic study |
| Structured summary | 2 | Provide a structured abstract |
| **Introduction** | | |
| Rationale | 3 | Specify the rationale for the overview of reviews in the context of an already-formed body of knowledge on the topic |
| Objectives | 4 | Describe a precise statement of questions |
| **Methods** | | |
| Protocol and registration | 5 | Report if an overview protocol was developed and if and where it can be obtained and provide registration information |
| Criteria for considering reviews for the overview | 6 | Describe review characteristics and report characteristics for eligibility criteria |
| Information sources | 7 | State all information sources in the search and date last searched |
| Search strategy | 8 | Specify full electronic search strategy including any limits used, such as language restriction |
| Review selection | 9 | Provide the process for selecting reviews and its relevant details |
| Additional searches to identify other relevant primary studies | 10 | Report whether and why additional searches were conducted to identify other eligible primary studies |
| Data extraction and management | 11 | State the processes of data extraction from included reviews and their relevant details |
| Data items | 12 | Specify all items overview authors sought (e.g., PICOS, methods, results, funding source) |
| Assessment of methodological quality of included reviews | 13 | Describe methods used for assessing methodological quality and quality of evidence and how this information was used for analyses |
| Data synthesis | 14 | Specify the methods of handling data and their details |
| **Results** | | |
| Review selection | 15 | Provide the details of review selection or a flow diagram of the overview process |
| Review characteristics | 16 | Describe characteristics of each review (e.g., title, PICOS, number of studies and participants included, assessment of methodological quality of reviews, results of individual reviews) |
| Assessment of methodological quality of included reviews | 17 | Report the results of assessment of methodological quality and quality of evidence of each included review |
| Syntheses of results | 18 | Summarize the main findings of the overview. If overview authors undertook data synthesis, present each summary measure with a confidence interval or a credible interval and measures of heterogeneity or inconsistency |

(continued)

**Table 13.1** (continued)

| Section/topic | # | Checklist item |
|---|---|---|
| **Discussion** | | |
| Summary of evidence | 19 | Provide a concise summary of the main findings with the strength and shortcomings of evidence for each main outcome |
| Limitations | 20 | Discuss limitations of the overview of review |
| Conclusions | 21 | Present implications for practice and future research |
| Funding | 22 | Describe sources of funding for the overview of reviews |

*Abbreviation*: *PICOS* participants interventions comparisons outcomes study design

## 13.2.2 Introduction

### 13.2.2.1 Rationale

Specify the rationale for the overview of reviews in the context of an already-formed body of knowledge on the topic.

**Explanation.** Overview authors should provide the rationale for the overview of reviews because readers need to understand why it is important to do the overview of reviews. Therefore, the background should specify what is already known, what is still unknown, and what the overview of reviews may add to current knowledge.

### 13.2.2.2 Objectives

Describe a precise statement of questions that specifies the types of population (participants), the types of interventions or exposures (comparisons), and the types of outcomes.

**Explanation.** The primary research questions should be presented precisely and explicitly, which enables readers to understand the scope of the overview of reviews quickly. This section should include the types of population (participants), the types of interventions or exposures (comparisons), and the types of outcomes that are of interest.

## 13.2.3 Methods

### 13.2.3.1 Protocol and Registration

Report if an overview protocol was developed and if and where it can be obtained and provide registration information.

**Explanation.** A protocol for an overview should be developed in advance before a research was undertaken in order to minimize potential biases as well as a protocol of a systematic review. Post hoc decisions or analyses based on knowledge of the available reviews are highly susceptible to bias, such as selective outcome reporting. Although protocols are sometimes modified on rational grounds, such as extending the period of searches to include older or newer reviews, changes in the protocol should be described with their reasons.

**Table 13.2** Differences between reporting of systematic reviews and overviews of systematic reviews

| | Systematic reviews | Overviews of systematic reviews |
|---|---|---|
| Eligibility criteria | Specify design of primary studies that should be potentially included | Specify design of primary studies that have already included in systematic reviews. Methodological characteristics of systematic reviews can be used as eligibility criteria |
| Information sources | Thorough search in all relevant sources should be conducted | Restricted search by databases and/or dates of coverage can be used |
| Search strategy | Language restriction and publication restriction should be avoided | Limited reporting characteristics might be reasonable |
| Selection process | All relevant primary studies on the same topic should be included | Only some of all relevant reviews on the same topic can be used or merged |
| Additional searches for other primary studies | This item is not necessary for systematic reviews | This item is specific to overviews of systematic reviews |
| Data extraction | The unit of interest is primary study, not report although there are some multiple reports of the same study | Specify how accuracy of data abstracted from primary studies is ensured and how overlapping and discordant reviews were managed |
| Data items | Only information from primary studies is extracted | Both information at the primary study level and at the systematic review level is extracted |
| Assessment of methodological quality | For included primary studies | For both included primary studies and included systematic reviews |
| Data synthesis | Pairwise or network meta-analyses can be conducted | Pairwise or network meta-analyses can be conducted in umbrella reviews, whereas meta-epidemiologic studies compare the results of primary studies with and without the characteristic of interest |
| Limitations | At the primary study level and the review level | At the overview level in addition to the primary study level and the review level |
| Implications | Implications from systematic reviews are usually for practice | Implications from umbrella reviews are usually for practice, whereas those from meta-epidemiologic studies are for research |

Registration of a protocol for an overview is desirable because it may reduce the possibility of a number of overviews assessing the same question, reduce selective outcome reporting biases or publication biases, promote transparency of overview process, and enable peer review of the protocol. The same databases can be used for registration of a protocol for both an overview and a systematic review (e.g., the PROSPERO http://www.crd.york.ac.uk/PROSPERO/).

#### 13.2.3.2 Criteria for Considering Reviews for the Overview

Describe review characteristics and report characteristics for eligibility criteria, including questions (participants, interventions or exposures, comparisons, outcomes, study design (PICOS)), and report characteristics (language, publication status, publication year).

**Explanation.** Overview authors should report eligibility criteria clearly to guide selection of reviews for inclusion. Because there might be a few differences in eligibility criteria of existing reviews that handled the same scope, overview authors should develop spreadsheets or other data tools to organize detailed criteria for considering reviews characteristics such as types of participants, types of interventions or exposures and comparisons, types of outcomes, and types of design of included primary studies (e.g., randomized controlled trails, observational studies, or both). These components are generally appropriate in umbrella reviews, as well as systematic reviews, which focused on a specific clinical topic such as a specific condition or drug, whereas in meta-epidemiologic studies that focused on a nonclinical topic at the level of the primary studies (e.g., funding issues and sample size), or at the level of the systematic reviews (e.g., publication bias and duplicate study selection and data extraction), it may be appropriate that only types of included study design and other information (e.g., review assessing primary outcome measure for a binary outcome, meta-analysis involving five or more primary trials) are included.

Reporting characteristics should also be specified as eligibility criteria. Reporting characteristics include language restriction, publication status, and publication year. In addition, overview authors can use the methodological quality of systematic reviews as eligibility criteria in order to include high-quality systematic reviews (item 13). Because high-quality systematic reviews generally include all relevant primary studies regardless of language, publication status, and publication year, it might be appropriate that overview authors restrict only recent reviews published in English to identify high-quality reviews effectively to conduct effective searches. Conversely, systematic reviews should avoid language restriction and publication restriction.

#### 13.2.3.3 Information Sources

State all information sources in the search and date last searched.

**Explanation.** Information sources include bibliographic databases with dates of coverage, handsearching, conference abstracts and other gray literature, reference lists in included reviews, related reviews and guidelines, and protocol registries.

Because recent, relevant, high-quality systematic reviews are needed for an overview of systematic reviews, although multiple existing reviews exist on the same topics, restricted databases such as MEDLINE's Top 120 Index Medicus Journals, the Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Health Technology Assessment, the Evidence-Based Practice Center Program, and other subject-specific sources may be appropriate to minimize time and resources [8]. In addition, limiting dates of the coverage in searching databases may also be reasonable because very few systematic reviews were published before 1990 [9]. On the other hand, comprehensive search for relevant primary studies in

all relevant sources without limited start date is desirable in systematic reviews to minimize selection bias.

However, when overview authors conducted additional searches to identify other relevant primary studies for incomprehensive searches or out-of-dateness of existing reviews, databases should at least include MEDLINE, EMBASE, the Cochrane Central Register of Controlled Trials, and other subject-specific sources.

#### 13.2.3.4 Search Strategy

Specify full electronic search strategy including any limits used, such as language restriction.

**Explanation.** Documenting the full search strategy could ensure that all the searches of all the databases are reproducible. It is also recommended that overview authors state whether search strategies were peer reviewed as a part of the overview process.

Because it is recommended overview authors select newer and more comprehensive reviews effectively among multiple reviews on the same topics, search strategy might limit language, publication status, and publication year for selecting more appropriate review (see item 6).

#### 13.2.3.5 Review Selection

Provide the process for selecting reviews and its relevant details.

**Explanation.** These includes how overview authors screened the retrieved records, how they decided which reviews were included, and if any types of reviews were excluded. A typical process for selecting reviews is, like systematic reviews, to merge search results and remove duplicates, to screen the retrieved records by titles and abstracts, to retrieve the full texts of the potentially relevant reports, and to screen the full texts and make decisions on review inclusion.

There might be some differences in inclusion and exclusion criteria of included reviews and the current overview criteria, and therefore, overview authors should assess inclusion and exclusion criteria for important domains (e.g., PICOS) and state its differences. If only some of primary studies in potentially relevant reviews meet the scope of overview, overview authors should specify it and include only them, rather than all primary studies in the reviews.

Moreover, if two or more reviews that met inclusion criteria exist on the same topic but each review included some different relevant primary studies or had different methodological quality (see item 13), overview authors should specify how one review was selected among them or how those reviews were merged. In umbrella reviews, overview authors may merge all relevant reviews to include additional relevant primary studies. In meta-epidemiologic studies, authors may include the most relevant review with sufficient quality among multiple reviews on the same topics when reviews focused on the level of the primary studies whreas authors may include all relevant reviews when reviews focused on the level of the systematic reviews. On the other hand, all relevant primary studies should be included in systematic reviews.

Authors also should report how many people and who carried out each stage and how disagreements were handled. It is usually recommended that at least more than one author assess each selection process. The process for resolving disagreements includes discussion among review authors, consultation with another person, and contact with the authors of the original reviews.

#### 13.2.3.6 Additional Searches to Identify Other Relevant Primary Studies

Report whether and why additional searches were conducted to identify other eligible primary studies that were not included in the identified reviews.

**Explanation.** Overview authors may have to conduct additional searches on other databases that were not included in original reviews, including MEDLINE, EMBASE, the Cochrane Central Register of Controlled Trials, and other subject-specific sources because existing reviews sometimes did not conduct comprehensive searches [10]. Overview authors might also have to conduct updating searches to identify new primary studies because existing reviews can become out-of-date within 3–5 years [11]. In addition, if there was no relevant systematic review on a certain condition or intervention, overview authors could also conduct a systematic review of primary studies by themselves. However, this is generally another research and out of overview's scope. When authors conducted additional searches to identify other relevant primary studies, they should report information sources, full electronic search strategy, criteria for considering primary studies for the overview, and the process for selecting primary studies. In addition, authors should provide their reasons whether or not additional searches were conducted.

#### 13.2.3.7 Data Extraction and Management

State the processes of data extraction from included reviews and their relevant details.

**Explanation.** The description includes whether a data collection form was used, if the form was piloted, who extracted data, whether any extractions were completed in duplicate and, if so, whether overview authors extracted data independently, and how disagreements were handled.

Overview authors should describe whether they contacted authors of included reviews when they were not able to obtain all of the relevant information they sought from the original reviews. If overview authors could not get such information from authors of included reviews, they should state whether they obtained data from primary studies of included reviews. The overview authors should also state whether information from contacted original investigators included individual patient data and provided in which analyses of such data were used.

Existing reviews might have data abstraction errors, and therefore, overview authors should provide whether they abstracted data from primary studies by themselves or evaluated a random sample of primary research to make sure abstracted data were accurate and reproducible. However, extracting all relevant data directly from primary studies might be out of overviews' scope for its time-consuming process.

Because some topics are frequently published in more than one systematic review, overview authors should report strategies to deal with discordance between reviews. The reasons for discordance include different strategies, methods, populations, study designs, outcomes, and interpretation of the same data [2]. There is a guide to assess discordant systematic reviews [12].

#### 13.2.3.8 Data Items

Specify all items overview authors sought (e.g., methods, participants, setting, intervention, outcomes, results, funding source).

**Explanation.** Overview authors should plan in advance what data will be collected for their overview of review. If overview authors selected only variables that were reported in the included reviews, rather than those that were important but could not be identified, analysis of those variables might lead to bias. Overview authors should also report variables that were added after data collection and the reasons for adding them.

#### 13.2.3.9 Assessment of Methodological Quality of Included Reviews

Describe methods used for assessing methodological quality and quality of evidence and how this information was used for analyses.

**Explanation.** Overview authors must assess methodological quality of included reviews. Important items for evaluating methodological quality include assessment of processes of literature search, study selection, data extraction, quality assessment of included studies, data synthesis, and methods used to assess for publication bias. Several methods for assessing quality can be used, such as the AMSTAR and the OQAQ (Overview Quality Assessment Questionnaire), although there are no superior quality assessment tools available [7, 13]. However, measurement tools that contain all of the important methodological domains and have good agreement, reliability, construct validity, and feasibility are recommended.

The methods for assessing quality of evidence in included reviews should also be reported. Although decision regarding which method is used depends on design and specific circumstances of the included primary studies, strength of empirical data, and theoretical rationale, one of the representative tools is the Grades of Recommendation, Assessment, Development and Evaluation Working Group (GRADE) system [14–16]. The GRADE approach uses quality of evidence within and across the included primary studies. For example, on the subject of systematic reviews of intervention, common items of risk of bias within the included primary studies for randomized trials include random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective outcome reporting [17]. Quality of evidence across the included primary studies involves indirectness of evidence, unexplained heterogeneity, precision of effect estimates, and risk of publication bias. This information may be extracted directly from systematic reviews or might be assessed by overview authors.

Overview authors should state whether any assessments were completed in duplicate and, if so, whether overview authors assessed quality independently and

how any disagreements were resolved. Because instruments for quality assessment are relatively subjective, it is recommended that at least two overview authors should assess quality independently and methods for resolving any discrepancies should be reported.

Existing reviews might provide inappropriate judgments of risk of bias within the included primary studies because of subjectivity of quality assessment tools, and therefore, overview authors should specify whether they judge risk of bias within the primary studies by themselves or evaluated a random sample of primary research to make sure judgment of risk of bias was accurate and reproducible. Overview authors should also state strategies to deal with the problem that review authors used different tools for assessing risk of bias. In this situation, overview authors might reassess risk of bias within primary studies with the same tool.

Finally, overview authors have to state how they used the results of methodological quality and quality of evidence. These include subsequent analyses only based on high-quality reviews, planned sensitivity analyses, and subgroup analyses related to quality assessment.

#### 13.2.3.10 Data Synthesis

Specify the methods of handling data and their details.

**Explanation.** Overview authors should prespecify approach to handle data. These approaches used in overviews include simply exacting data from reviews and reformatting them, indirect comparisons or network meta-analyses, comparing intervention effects in trials with and without the characteristic of interest (e.g., ratio of odds ratio), and qualitative or narrative approaches. Overview authors should also report the principal summary measures (e.g., odds ratio, mean difference) and how to assess heterogeneity if overview authors undertook data synthesis and the ways of assessing inconsistency if indirect comparisons or network meta-analyses were performed. Overview authors may assess small study effects, publication bias, or selective reporting bias and provide the method to assess them. When overview authors included only some of the primary studies in reviews or added other relevant primary studies to the related reviews, approaches to handle this information should also be described. If additional analyses, such as sensitivity analyses, subgroup analyses, or meta-regression, were conducted, overview authors should describe their ways.

### 13.2.4 Results

#### 13.2.4.1 Review Selection

Provide the details of review selection (e.g., numbers of reviews screened, retrieved, and included and excluded in the overview) or a flow diagram of the overview process.

**Explanation.** Overview authors should state at least the total number of reviews identified from all sources or ideally report each number of reviews identified from

different sources, such as bibliographic databases, handsearching, conference abstracts and other gray literature, reference lists in included reviews, experts, related reviews and guidelines, and trial registers.

The process for selecting reviews for inclusion and exclusion in an overview should be clearly described. Overview authors should state the numbers of reviews identified in searches, duplicate reviews removed, reviews screened by titles and abstracts, potentially relevant full-text reports retrieved for detailed evaluation with potentially eligible reviews that were not retrievable, and reviews and primary studies included in an overview. The number of excluded reviews at each step should also be noted with their reasons. When overview authors included only some of the primary studies in certain reviews, this information and its reasons should be described. Moreover, the number of primary studies identified in additional searches should be reported with their reasons. It is recommended this information should be described in both a text and a flow diagram.

#### 13.2.4.2 Review Characteristics

Describe characteristics of each review (e.g., title, PICOS, number of studies and participants included, follow-up period, assessment of methodological quality of included reviews, search date, results of individual reviews).

**Explanation.** Overview authors should concisely describe sufficient detail of characteristics of included reviews because readers need to know them to assess their relevance, validity, and applicability. This information includes title, PICOS, number of studies and participants included, follow-up period, assessment of methodological quality of included reviews, search date, and results of individual reviews. These items are also presented as a table in general. When any differences between the eligibility criteria of the included reviews and that of an overview of reviews were found, overview authors should specify them.

If overview authors were not able to obtain all of the relevant information from the original reviews, primary studies, or original investigators, the overview authors should note this information. If overview authors imputed this information, they should describe which items were done.

When only some primary studies in reviews were included or additional relevant primary studies were included to the related reviews, overview authors should specify how and which items were incorporated. Overview authors should also specify how and why there are some differences among systematic reviews that handled the same topic.

#### 13.2.4.3 Assessment of Methodological Quality of Included Reviews

Report the results of assessment of methodological quality and quality of evidence of each included review.

**Explanation.** It is recommended that overview authors report both summary assessment of included reviews and the methodological features evaluated for each review. When overview authors described only summary assessment (e.g., 40 % of included reviews conducted duplicate study selection and data extraction), readers cannot tell which reviews had the specific methodological flaws.

#### 13.2.4.4 Syntheses of Results

Summarize the main findings of the overview. If overview authors undertook data synthesis, present each summary measure with a confidence interval or a credible interval and measures of heterogeneity or inconsistency.

**Explanation.** The main findings of overviews of systematic reviews should be summarized according to clinically meaning categories, such as types of studies, participants, interventions, or outcomes. The use of summary tables and figures is helpful in presenting results of overviews of reviews as well as that of systematic reviews.

If any data syntheses were conducted, each summary measure with a confidence interval or a credible interval should be reported. A measure of heterogeneity or inconsistency should also be described. If no data synthesis was performed, qualitative evidence should be provided with a reason why meta-analysis was not conducted (e.g., suitable numerical data were not available, meta-analysis was not considered appropriate). Overview authors should also report the outcome measures that they considered important but for which could not find evidence.

### 13.2.5 Discussion

#### 13.2.5.1 Summary of Evidence

Provide a concise summary of the main findings with the strengths and shortcomings of evidence for each main outcome.

**Explanation.** Overview authors should summarize the main findings including the strengths and weaknesses of evidence. Reporting outcome for which little or no data were found is important for policy decision and future research. In overviews of reviews of interventions, the balance between important benefits and harms should also be specified. Applicability of evidence to different relevant participants, settings, or characteristics should be mentioned.

#### 13.2.5.2 Limitations

Discuss limitations of the overview of review.

Explanation.Overview authors should address limitations at the primary study level, the review level, and the overview level. Limitations at the primary study level are risk of bias within the included primary studies, and limitation at the review level includes quality and quality of evidence in included reviews, whereas examples of limitations at the overview level are overall completeness and applicability of evidence, potential biases in the overview process, and disagreements among included reviews.

Limitations of quality and quality of evidence in original reviews (item 12) might include whether duplicate study selection and data extraction was performed, whether comprehensive literature search was conducted (e.g., status of publication, language), whether reviews were up-to-date, and whether review authors could obtain all relevant data, design of primary study, indirectness of evidence, heterogeneity, and publication bias.

Overview authors should specify whether overview addressed all relevant types of participants, interventions, and outcomes as generalizability of the overview. If not, they should report what key issues should be addressed. Potential bias in overview process may be limitations of searching reviews, additional searches to identify all relevant primary studies and data, study selection, data collection, and analysis. If disagreements among reviews are present, overview authors should report it.

#### 13.2.5.3 Conclusions

Present implications for practice and future research.

**Explanation.** It is recommended that conclusion of the overview authors should be divided into implication for practice and implication for future research. The results of umbrella reviews are usually important for decision-makers, whereas those of meta-epidemiologic studies are for researchers.

Overview authors should describe the interpretation of the results based on the balance of benefits, harms and costs, and the quality of evidence. If there were few reliable reviews, this should be reported. Because recommendations for an action usually need additional information, such as patient's preference and values, and resource utilization, recommendations based only on the results of overview is usually optimistic. If overview authors try to advocate actions, they should provide several particular factors and assumptions that might influence a decision.

We also recommend overview authors address current evidence and the important clinical issues that remain unresolved after conducting overview of reviews to make explicit recommendations for future research. These important clinical factors include PICOS and method of literature search.

#### 13.2.5.4 Funding

Describe sources of funding for the overview of reviews.

**Explanation.** Overview authors should report any funding that supported the overview. If no funding is received, they should state this. Overview authors should also describe the role of funders for the overview.

## 13.3 Conclusion

Although overviews of systematic reviews have some of similar reporting and methodological standards to systematic reviews, there may be some differences as well as additional challenges [1]. These differences include whether eligibility criteria include reviews method, whether searches are limited to high-yield information sources, whether additional searches are conducted to identify other relevant primary studies, how discordant findings between reviews are handled, how the quality of reviews is assessed, and how the results of reviews are synthesized (Table 13.2). In addition, there are some differences between reporting of umbrella reviews and meta-epidemiologic studies, such as data synthesis and implication. Although we offered a reporting and methodological checklist for overviews of systematic

reviews, a consistent nomenclature and further assessment of its validity, reliability, sensitivity, and responsiveness are needed.

## References


1. Becker KA, Oxman AD. Overviews of reviews. In: Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions, version 5.1.0. The Cochrane Collaboration; 2011. Available from: http://handbook.cochrane.org/. Last accessed on 20 May 2015.
2. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. J Clin Epidemiol. 2012;65:1267–73.
3. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzshe PC, Ionnidis JP, Clarke M, Devereaux PJ, Kleijne J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.
4. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA. 2000;283:2008–12.
5. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7:e49667.
6. Li L, Tian J, Tian H, Sun R, Liu Y, Yang K. Quality and transparency of overviews of systematic reviews. J Evid Based Med. 2012;5:166–73.
7. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009;62:1013–20.
8. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med. 2008;148:776–82.
9. Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. Eval Health Prof. 2002;25:12–37.
10. Onishi A, Furukawa TA. Publication bias is underreported in systematic reviews published in high-impact-factor journals: metaepidemiologic study. J Clin Epidemiol. 2014;67:1320–6.
11. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007;147:224–33.
12. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. CMAJ. 1997;156:1411–6.
13. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44:1271–8.
14. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–6.
15. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams Jr JW, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt GH, GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008;336:1106–10.
16. Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, Kessels AG, Guyatt GH, GRADE Working Group. A GRADE working group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.
17. Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration; 2011. Available from: http://handbook.cochrane.org/. Last accessed on 20 May 2015.

# Part III

# Case Studies

# Case Study of Allied Health and Complementary and Alternate Medicine

14

Saravana Kumar

**Abstract**

With increasing production of, and access to, research evidence, umbrella reviews are gaining prominence as a structured and systematic means for summarising the best, high-level evidence to answer a clinical and review question. Umbrella reviews may act as an important, timely and efficient resource as, due to its very nature, it summarises existing reviews through a systematic, rigorous, transparent and universally recognised set of processes. Allied health and complementary and alternate medicine comprise of a collective of health-care professions which are uniquely diverse, in terms of their roles and responsibilities, and yet are grouped under the umbrella term "allied health" and "complementary and alternate medicine". Umbrella reviews, as a methodology to systematically review the best research evidence, is in its infancy in these professions. This chapter summarises four purposively selected umbrella reviews conducted in allied health and complementary and alternate medicine and discusses some of the opportunities and challenges in undertaking umbrella reviews within this field.

## 14.1 Introduction

Health care in the twenty-first century has undergone large-scale changes in response to a number of challenges. Within the developed world, the challenges have come in the form of an ageing population with multiple comorbidities, the rise of the non-communicable diseases, pressures on meagre health-care resources due to financial

S. Kumar, BSc, MPT, PhD
School of Health Sciences, International Centre for Allied Health Evidence, University of South Australia, City East Campus, Adelaide 5000, Australia
e-mail: saravana.kumar@unisa.edu.au



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_14

constraints and burgeoning demand, explicit recognition for quality to underpin health-care service delivery (avoid underuse, overuse, and misuse of health care), uneven distribution of health-care professionals (oversupply in metro/urban region and undersupply in rural/remote regions) and an increasingly informed consumer who is an active participant in health-care decision making (unlike the historical passive receipt of health care) [1]. Many of these challenges may also exist in the developing world, where due to societal, financial and resource disadvantages there may also be challenges associated with communicable and non-communicable diseases within the same population groups. Evidence-based practice (EBP) [2] has been considered as a vehicle to address some of these challenges. The philosophy underpinning EBP calls for using research evidence to inform health-care decision making while taking into account patients morals, values and beliefs, the clinical expertise and experiences of the health professional and information from the practice context [3].

It is in this context, umbrella reviews can play an important role. As umbrella reviews are underpinned by systematic review methodology, umbrella reviews use a systematic, rigorous, transparent and universally recognised set of processes which enables systematic reviews to be considered as the highest level of evidence for intervention, aetiology, prognosis, diagnostic and screening intervention questions (Level 1) [4]. Given that the evidence base is growing rapidly, due to which the evidence base also continually evolves, umbrella reviews are ideal vehicles to summarise existing highest level of evidence and provide solutions to questions in a timely, efficient and structured manner. By using existing secondary research, umbrella reviews negate the need to engage with primary research and as such are ideal tools for any health-care stakeholder.

## 14.2 What Are Allied Health and Complementary and Alternate Therapies?

Unlike other areas such as medicine, undertaking systematic reviews in allied health and complementary and alternate medicine pose a number of unique challenges. A key challenge is the definitional confusion and ambiguity that underpins allied health and complementary and alternate medicine. There is no standard or agreed definition of allied health, the most common approach has been to group a number of professions which are not medical/nursing as "allied health" and this has resulted in as many as 100 or more distinct professions being labelled as "allied health" [5]. Recently though, there has been some attempt to have a universal definition of allied health, which is built on the tasks performed by health professionals such as therapy, education, assessment and management rather than a mere designation [6]. In addition to definitional ambiguity, there is great deal of diversity in terms of educational and registration requirements, ability to practice as primary contact practitioners, referral and prescription rights and professional supervision and regulatory requirements which vary across jurisdictions [6]. Similarly, for complementary and alternate medicine, there is no uniform definition [7]. Complementary and alternate

medicine could be considered as those health-care practices that are not integral part of conventional medicine [8] and are often denoted by a number of terms (such as alternate therapy, holistic therapy and traditional medicine). In recent times, the term integrative medicine has also been used which indicates the integration of conventional medicine with complementary and alternate medicine [9]. Due to this definitional confusion and ambiguity, there may be little uniformity and lack of standardisation in reporting of primary research data in allied health and complementary and alternate medicine.

Other challenges in undertaking systematic reviews in allied health and complementary and alternate medicine include consistent access to and relevance of high-level research evidence across all professions (e.g. the role of randomised controlled trials for social work) [10]; provision of packages of care as part of interventions (which may be clinically relevant and appropriate but not suited for experimental research) [11, 12]; lack of detailed description of interventions [13] (resulting in inability to recognise the parameters of the intervention and for replication); heterogeneity in terms of patient population, comparators and measures of outcome; despite a large body of evidence from qualitative research paradigm, lack of recognition of qualitative research as useful, appropriate and relevant evidence [14]; and potential for publication [15] and language bias [16] (a large body of research on complementary and alternate medicine is available in languages other than English (such as Chinese and Hindi) and may not be accessible through mainstream databases).

Not surprisingly, these challenges witnessed in the conduct and reporting of systematic reviews also extend into the conduct and reporting of umbrella reviews (which use these systematic reviews). Nevertheless, they provide an insight into the processes and outcomes of these umbrella reviews and by doing so they highlight opportunities and challenges in evidence synthesis of overview of reviews in these fields. The purposively selected four case studies of umbrella reviews of allied health and complementary and alternate medicine in this chapter focus on (a) massage therapy [17], (b) exercise [18], (c) spinal manipulation [19] and (d) acupuncture [20].

## 14.3 Case Studies of Allied Health and Complementary and Alternate Medicine

### 14.3.1 Case Study 1: Massage Therapy

**Full citation** Kumar S, Beaton K, Hughes T. The effectiveness of massage therapy for the treatment of non-specific low back pain: a systematic review of systematic reviews. Int J Gen Med. 2013. doi:http://dx.doi.org/10.2147/IJGM.S50243.

**Background to the review** It is commonly thought that massage is a safe therapeutic modality without any significant risks or side effects. However, despite its popularity, there continues to be an ongoing debate on the effectiveness of massage in treating non-specific low back pain.

**Review objective** To provide a synthesis of the best available research evidence for the effectiveness of massage therapy for adults suffering from non-specific low back pain. Systematic reviews are considered to be the highest level of evidence for intervention questions.

**Inclusion criteria** Types of participants: Adults (≥18) suffering from non-specific acute, subacute or chronic low back pain (low back pain is defined as pain that is localised from the 12th rib to the inferior gluteal fold). "Non-specific" means that there is no specific cause of the low back pain such as neoplasms, infection, osteoporosis, arthritic conditions, fracture, radicular syndrome or inflammatory processes. Types of exposure: For the purpose of this review, massage therapy was defined as the manual manipulation of the soft tissues of the body for therapeutic purposes. Types of comparators: Sham or placebo treatment, medical interventions, physical therapy, electrical therapy (TENS, US, etc.), pharmaceutical interventions and other forms of alternative therapy. Types of outcomes: Patient self-report/subjective change of symptoms, assessment of pain, functional status as measured by validated tools and assessment of range of motion (ROM). Type of studies: In order for inclusion in this systematic review, reviews had to meet the PRISMA definition of a systematic review or meta-analysis.

**Search strategy** A systematic search was conducted in the following databases: EMBASE, MEDLINE, AMED, ICONDA, Academic search Premier, Australia/New Zealand Reference Centre, CINAHL, HealthSource, SPORTDiscus, PubMed, The Cochrane Library, Scopus, Web of Knowledge/Web of Science, PsycINFO, ProQuest Nursing and Allied Health source, investigating systematic reviews and meta-analyses from January 2000 to December 2012, and restricted to English language documents. Secondary searching (pearling) of reference lists of any reviews returned in this search was also pearled for additional systematic reviews or meta-analyses that may not have been found in the original search.

**Methodological quality assessment and data extraction** Upon the selection of relevant publications, two reviewers independently evaluated the methodological quality using the Centre for Evidence-Based Practice (CEBM) critical appraisal tool. The criteria were as follows: (1) what question (PICO) did the systematic review address, (2) is it unlikely that important/relevant studies were missed, (3) were the criteria used to select articles for inclusion appropriate, (4) were the included studies sufficiently valid for the type of question asked, (5) were the results similar from study to study and (6) how were the results presented. Following this process, the included systematic reviews were categorised as poor, moderate, good and excellent. The data from each systematic review was then extracted into a custom-built table which was derived from CEBM questions.

**Synthesis and analysis** Narrative synthesis.

**Results** Nine systematic reviews were included with varying methodological quality. Partly, this was due to the poor methodological quality of the primary research which informed the subsequent systematic reviews. Collectively though, the findings indicate that massage therapy may be an effective treatment in comparison to placebo and when compared to some active treatments such as relaxation. However, these effects were only noticed in the short term. When massage therapy was

compared to other forms of active treatment, such as mobilisation, standard medical care and acupuncture, the evidence is conflicting and contradictory.
**Conclusion** The findings from this umbrella review suggest that massage therapy may be effective in treating non-specific low back pain, especially in the short term. Due to contradictory and conflicting evidence, it is unclear what, if any, effects massage therapy has in the long term especially when compared against other common treatments.

### 14.3.2 Case Study 2: Exercise

**Full citation** Bidonde J, Busch AJ, Bath B, Milosavljevic S. Exercise for adults with fibromyalgia: an umbrella systematic review with synthesis of best evidence. Curr Rheumatol Rev. 2014;10:45–79.
**Background to the Review** Fibromyalgia is a common condition that affects a two to three per cent of the world's population. While its aetiology remains unclear, its impact can be far ranging including fatigue, lack of or disturbed sleep, stiffness, depression and cognitive problems. There are many pharmacological and non-pharmacological treatments for fibromyalgia. While there is evidence to suggest that these interventions may be effective, to date no umbrella reviews have been conducted on this topic.
**Review objective** This umbrella review aimed to evaluate the quality of systematic reviews of physical activity interventions for adults with fibromyalgia.
**Inclusion criteria** Types of participants: Adults of either gender diagnosed with FM according to a published criteria. Types of exposure: Any intervention in which participants perform a programme of regular physical activity over a period of time. Any interventions that did not involve bodily movements (such as interventions which involved soaking in water) were excluded. Types of comparators: Any control or comparison group. Types of outcomes: Multidimensional function (well-being or quality of life), pain, physical function (self-report or observational test) and adverse effects. Time frame: January 2007 to March 2013. Type of studies: Systematic review (meeting at least three of five criteria characterising a systematic review) of randomised controlled trials. Narrative reviews and reviews with no identified search strategy were excluded.
**Search strategy** A systematic search was conducted in the following six databases: Medline, EMBASE, CINAHL, AMED, the Cochrane Library, and DARE. No restrictions on languages were placed. Secondary searching (pearling) of reference lists of any reviews returned in this search were also pearled for additional systematic reviews or meta-analyses that may not have been found in the original search.
**Methodological quality assessment and data extraction** Two reviewers independently evaluated the methodological quality of included systematic reviews using the AMSTAR methodological quality measurement tool. The 11 AMSTAR items were scored (yes/no) to evaluate the adequacy of the important components of the method: search, selection criteria, validity assessment and synthesis. Inter-rater reliability was assessed using Kappa statistic. In this review, data was extracted at

two levels. First, data extraction focussed on the systematic reviews, and the second, individual randomised controlled trials which formed the basis of systematic reviews were also extracted.

**Synthesis and analysis** Narrative synthesis.

**Results** Nine systematic reviews were included in this umbrella review. There were positive findings for the effectiveness of diverse exercise interventions for pain, multidimensional function and self-reported physical function. This umbrella review, however, did not find any supporting evidence for emerging exercise therapies such as Qigong and Tai Chi. Furthermore, across the systematic reviews there were no serious harm performing exercise for individuals with fibromyalgia. While these are positive findings, due to variability in the interventions, recommendations regarding optimal dosages and modes could not be provided which impacts on making concrete recommendations for clinical practice.

**Conclusion** The findings from this umbrella review suggest that diverse exercises do have a positive impact in terms of pain, multidimensional function and self-reported physical function in individuals with fibromyalgia. However, there continues to persist methodological issues with the current evidence base. This is particularly evident when making recommendations regarding the parameters of exercise interventions. Therefore, stakeholders of fibromyalgia should be a discerning consumer of the current evidence base when using it to make health-care decisions.

### 14.3.3 Case Study 3: Spinal Manipulation

**Full citation** Ernst E, Canter PH. A systematic review of systematic reviews of spinal manipulation. J R Soc Med. 2006;99:192–6.

**Background to the review** Spinal manipulation is a commonly used and popular treatment for a range of health conditions but mostly musculoskeletal problems. It can be performed by a number of practitioners, namely, physiotherapists, chiropractors and osteopaths, and the benefits of spinal manipulation are thought to include increased range of motion, reduce pain, improve joint mobility and kinematics and increase proprioception. Despite its popularity, the evidence for the effectiveness of spinal manipulation continues to be variable.

**Review objective** This umbrella review summarised current evidence from systematic reviews and clear the confusion underpinning the effectiveness of spinal manipulation.

**Inclusion criteria** Types of participants: Participants who received spinal manipulation, irrespective of the medical condition. Types of exposure: For the purpose of this review, systematic reviews had to be concerned specifically with the effectiveness of spinal manipulation and to include evidence from at least two controlled clinical trials. Spinal manipulation when provided as part of a complex intervention was excluded. Type of studies: For inclusion in this systematic review, reviews had to include an explicit and repeatable method of searching the scientific literature and if there were explicit and repeatable inclusion and exclusion criteria for studies.

**Search strategy** Four databases were searched and they were Medline, Embase, AMED and Cochrane Database. Where possible searching was restricted to as reviews or meta-analyses and, in all cases, the search was restricted to articles published between 2000 and May 2005. No language restrictions were applied.
**Methodological quality assessment and data extraction** While a stand-alone methodological quality assessment process was not undertaken, in order to be included in this overview, systematic reviews had to include an explicit and repeatable method for searching the scientific literature and if there were explicit and repeatable inclusion and exclusion criteria for studies. These two items were derived from a methodological quality scoring system and therefore some degree of methodological quality assessment was built into the methods.
**Synthesis and analysis** Narrative synthesis.
**Results** This umbrella review identified 16 systematic reviews which reported on the effectiveness of spinal manipulation. Not surprisingly, many systematic reviews had a focus on musculoskeletal problems (such as back pain, neck pain, lower back pain and headache) but some also focussed on other health conditions such as dysmenorrhoea, infantile colic, asthma, allergy and cervicogenic dizziness. Collectively, the findings from this umbrella identified that the evidence base for spinal manipulation was generally not supportive except in the case of back pain. For back pain, there was evidence to indicate that spinal manipulation was superior to sham manipulation. However, spinal manipulation was not better than other conventional treatments. However, the authors do acknowledge that none of the systematic reviews conclusively demonstrated that spinal manipulation was ineffective.
**Conclusion** The findings from this umbrella review suggest that spinal manipulation may not be an effective treatment for any condition. The authors caution that given the opportunities for adverse events, recommending spinal manipulation as a treatment choice should be made with caution.

### 14.3.4 Case Study 4: Acupuncture

**Full citation** Gilbey A, Ernst E, Tani K. A systematic review of reviews of systematic reviews of acupuncture. Focus Altern Complement Ther. 2013. doi:10.1111/fct.12004.
**Background to the review** Amongst the many complementary and alternate medicine therapies, acupuncture is one of the most common and popular forms. With the growth in complementary and alternate medicines, there has been renewed focus on the effectiveness of its treatments, including acupuncture. However, much of the evidence base for acupuncture, historically, is variable.
**Review objective** To systematically review existing reviews of reviews of the literature on the effectiveness of acupuncture to better understand persistent contradictions in the body of the literature.
**Inclusion criteria** Types of participants: Participants who received acupuncture with any condition. Types of exposure: For the purpose of this review, systematic reviews had to be concerned specifically with the effectiveness of acupuncture

(e.g. electro-acupuncture, acupressure or auricular acupuncture, but not exclusively moxibustion). Acupuncture was the primary intervention for one or more conditions, or one of several interventions for one or more conditions. Type of studies: In order for inclusion in review, the review must be a critical overview, critical review, overview, review, summary or systematic review of systematic reviews.

**Search strategy** A systematic literature search was undertaken through the following databases: MEDLINE (via Ovid), Scopus and EbscoHost. The search terms used were (systematic review) AND (acupuncture OR acupressure) and the search was limited to the period from January 1991 to December 2011. No language restrictions were applied.

**Methodological quality assessment and data extraction** As this review was the first to review reviews of reviews, the authors devised their own critical appraisal tool to assess the methodological quality of included reviews. There were nine items which formed the basis of this tool. The items were: (1) Was the objective and reason for the SR of reviews explained? (2) Were inclusion criteria reported? (3) Were exclusion criteria reported? (4) Were the search methods reported? (5) Was the way in which reviews were selected reported? (6) Was the quality of the SR explicitly assessed? (7) Was the quality of the primary studies explicitly assessed? (8) Was a summary of findings of included reviews provided? (9) Were implications for practice and/or research given? Following the critical appraisal process (which was undertaken by two independent reviewers), reviews were categories into three categories: high methodological quality (i.e. had few or no minor flaws) if the total score was greater than or equal to 14, low methodological quality (i.e. had numerous flaws) if total scores were greater than or equal to seven and moderate quality if total scores ranged from 8 to 13.

**Synthesis and analysis** Narrative synthesis.

**Results** Eighteen reviews of reviews were included as part of this review. The methodological quality of the systematic reviews of reviews varied from poor to excellent. Generally, the findings from the 18 reviews were categorised into two categories: positive to inconsistent (or inconclusive) or inconsistent (or inconclusive) to negative. The results indicated that there is evidence of effectiveness of postive impact for acupunture for some types of pain relief, nausea and vomitting. On the other hand, there were negative findings for rheumatic arthritis and inconclusive to conclusive negative findings for smoking cessation.

**Conclusion** Findings from this review suggests that despite a large body of secondary research, firm conclusions about the effectiveness of acupuncture cannot be drawn. This is due to methodological quality issues and poor quality primary research which underpin many secondary research initiatives. The authors call for improved primary research studies (such as randomised controlled trials) in treatment areas where there is existing and emerging evidence.

## 14.4 Discussion

With increasing focus on evidence-based practice and the explicit recognition for embedding research evidence into health-care decision making, health-care stakeholders are increasingly relying on current best evidence to augment clinical expertise,

patient's morals, values and beliefs and the practice contexts. Umbrella reviews can play an important role in this context as they provide access to current best evidence using a systematic, rigorous, transparent and universally recognised set of processes. Furthermore, by providing access to current best evidence, umbrella reviews can be time and resource efficient and act as a "one stop shop" for a range of clinical review questions. Umbrella reviews cannot only shed light on the current evidence base for a particular topic, it can also focus on gaps in knowledge, which can be the impetus for further research and investment. These opportunities mean umbrella reviews will continue to gain increasing prominence in contemporary health-care decision making.

However, while umbrella reviews have a number of merits, they also suffer from a number of limitations which act as its "Achilles Heel". While no one design can answer all questions, it is important to recognise these limitations and challenges when considering the impact and utility of umbrella reviews in allied health and complementary and alternate medicine.

- Predominant intervention focus: Currently it is thought that umbrella reviews ideally lend themselves to clinical review questions which have a focus on an intervention. This is partly because, historically, systematic reviews have had a predominant focus on interventions and naturally this has now extended to umbrella reviews. Given the complexity and diversity of health care, there is now increasing focus on questions that address diagnostic, prognostic, aetiological and screening interventions and there are a number of systematic reviews in these fields. However, to date, umbrella reviews that focus on diagnostic, prognostic, aetiological and screening interventions are limited. In order to address these gaps, there needs to be recognition for other forms of research as credible evidence, especially in instances when intervention is not the most appropriate design (such as aetiological research).
- Efficacy vs. effectiveness: As highlighted previously, umbrella reviews, to date, have had an intervention focus reporting on the best available evidence on the effectiveness of a stand-alone intervention or a package of interventions. Often the systematic reviews, which form the basis of the umbrella reviews, rely on randomised controlled trials as the primary research design of choice due to its ability to minimise bias in the research process (hence occupying the highest level of evidence for primary research designs). While this is true, randomised controlled trials are ideally placed to evaluate the efficacy of an intervention (tested in an ideal or controlled setting) rather than the effectiveness of an intervention (tested in a real-world condition). Therefore, using randomised controlled trials to test the effectiveness of allied health and complementary and alternate medicine interventions, which are often packaged as complex interventions, may be inappropriate and erroneous [21, 22]. As a result, the findings from the randomised controlled trials (and the umbrella reviews) often are listed as no effect, non-significant and/or inconclusive findings. Not surprisingly, all the four included case studies (including one which is a systematic review of reviews of systematic reviews) echo, at least in part, no effect, non-significant and/or inconclusive findings.
- Lack of focus on qualitative research paradigm: Given that evidence-based practice has its origins in quantitative research, to date, qualitative research (and

hence its research) has remained in its shadow. While this is slowly changing (due to the recognition of qualitative research as an important contributor to evidence-based practice in the form of integrating patients morals, values and beliefs), there remains a long way to go in terms of secondary qualitative research. For example, unlike quantitative systematic reviews, qualitative systematic reviews using meta-synthesis is still in its infancy. Once this has matured, it can then be progressed to umbrella reviews, which can contribute to a generation of new knowledge.

- Definitional confusion and ambiguity: As highlighted at the beginning of this chapter, there remains confusion and ambiguity in terms of "whom" and "what" constitutes allied health and complementary and alternate medicine. This confusion and ambiguity also extends to the interventions provided by health professionals from allied health and complementary and alternate medicine. For example, one of the included case studies focused on spinal manipulation. This technique can be provided by physiotherapists, chiropractors and osteopaths and, depending on the professional grouping, geographical location of practice, and teaching and practice philosophies, may mean entirely different things to different people. This often results in comparing "apples with oranges", and it is no surprise that concrete conclusions cannot be drawn at the completion of the reviews. This can be solved by ensuring detailed reporting within the primary and secondary research.
- The devil is in the detail: Unlike some medical interventions (such as medications), allied health and complementary and alternate medicine have complex, packages of care. This is because allied health and complementary and alternate medicine often provide a number of interventions during a single occasion of service. Replicating this in a scientific research, such as a randomised controlled trial, is difficult and may be antithesis to the principles of rigorous research process (which is built on standardisation and control). Furthermore, even if complex interventions are tested, reporting of these interventions is limited as authors may skimp on details in the methods section (as they may considered it as being not overly relevant or as means to save words and use them for the results and discussion sections). As a result these interventions are poorly described which impacts on the ability to make recommendations for practice change (replicability). This is a reporting issue which can be resolved through thorough reporting of interventions.
- Quality (or lack thereof) of primary research: One of the common findings across all the four case studies was the quality (or lack thereof) of the included primary research. This is not an uncommon finding in allied health and complementary and alternate medicine literature. Any reviews of the literature (be it umbrella review, systematic review, literature review) can only report on the available primary research. If these primary researches are flawed, inevitably, the secondary researches which rely on these primary researches are impacted by these flaws. Therefore, careful consideration is required when making recommendations arising from poor quality primary research (which formed the basis of systematic reviews).

- Publication bias: One of the limitations of any secondary research is publication bias. Some critics of secondary research (in particular systematic reviews) argue that as often only studies with positive findings are published, studies with negative or no effects may not be published and hence not be accessible for nor included in systematic reviews. Some systematic reviews address this by incorporating searching strategies which go beyond the traditional sources of research (such as grey literature); however, this is not common practice due to time and resource implications. This is a limitation that has to be acknowledged as part of any umbrella reviews that incorporates systematic reviews with limited search strategies.
- Language bias: Much of the accessible health literature is available through databases which house mostly English language publications. While there is access to other language publications, this is mostly confined and is dependent on the databases, language skills of the reviewers and resource availability (for translation purpose). In areas such as complementary and alternate medicine, there may be large bodies of research in other languages which are not accessible through mainstream searching (such as Chinese medicine, Ayurveda) and hence not reported in systematic reviews and umbrella reviews. This may result in language bias. This, too, is a limitation that has to be acknowledged as part of any umbrella reviews that incorporates systematic reviews with a single language focus.

### Conclusion

This chapter has showcased four umbrella reviews from the fields of allied health and complementary and alternate medicine. It has highlighted a number of opportunities for umbrella reviews to play an important role in informing contemporary health-care decision making while also recognising some of the challenges moving forward. As allied health and complementary and alternate medicine fields are complex, considered thoughts and actions are required to fine-tune and reshape some methodological considerations and practical applications during the planning, conduct and reporting of umbrella reviews. By doing so, umbrella reviews can continue to remain a useful tool to synthesise the current best evidence in allied health and complementary and alternate medicine while providing meaningful and valuable evidence-based solutions to challenges in this field.

## References

1. Armstrong BK, Gillespie JA, Leeder SR, Rubin GL, Russell LM. Challenges in health and health care for Australia. Med J Aust. 2007;187(9):485–9.
2. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Br Med J. 1996;312:71–2.
3. Hoffmann T, Bennett S, Del Mar C. Evidence-based practice across the health professions. Sydney, Australia: Elsevier; 2001.
4. Australian Government: National Health and Medical Research Council (NHMRC). NHMRC additional levels of evidence and grades for recommendations for developers of guidelines-stage 2 consultation. Australia: Canberra; 2009.

5. Thomas SW, McLean L, Debnam A. The role of allied health in health care reform. N C Med J. 2011;72(5):417–9.
6. Turnbull C, Grimmer-Somers K, Kumar S, May E, Law D, Ashworth E. Allied, scientific and complementary health professionals: a new model for australian 'allied health'. Aust Health Rev. 2009;33(1):27–37.
7. Coulter ID, Willis EM. The rise and rise of complementary and alternative medicine: a sociological perspective. Med J Aust. 2004;180(11):587–9.
8. Zollman C, Vickers A. What is complementary medicine? Br Med J. 1999;319:693.
9. Zhang W, Pritzker SE, Hui KK. Factors affecting definitions of and approaches to integrative medicine: a mixed methods study examining China's integrative medicine development. Evid Based Complement Alternat Med. 2015. http://dx.doi.org/10.1155/2015/458765.
10. Dixon J, Biehal N, Green J, Sinclair I, Kay C, Parry E. Trials and tribulations: challenges and prospects for randomized controlled trials of social work with children. Br J Soc Work. 2014;44:1563–81.
11. Blackwood B. Methodological issues in evaluating complex healthcare interventions. J Adv Nurs. 2006;54(5):612–22.
12. Dutta J, Petticrew M. Challenges to evaluating complex interventions: a content analysis of published papers. BMC Public Health. 2013;13:568.
13. Wells M, Wiliams B, Treweek S, Coyle J, Taylor J. Intervention description is not enough: evidence from an in-depth multiple case study on the untold role and impact of context in randomised controlled trials of seven complex interventions. Trials. 2012;13:95. doi:10.1186/1745-6215-13-95.
14. Broeder J, Donze A. The role of qualitative research in evidence-based practice. Neonatal Netw. 2010;3:197–202(6).
15. Torgerson CJ. Publication bias: the achilles heel of systematic reviews? Br J Educ Stud. 2006. doi:10.1111/j.1467-8527.2006.00332.x.
16. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, Mierzwinski-Urban M, Clifford T, Hutton B, Rabb D. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care. 2012;28(2):138–44. doi:10.1017/S0266462312000086.
17. Kumar S, Beaton K, Hughes T. The effectiveness of massage therapy for the treatment of non-specific low back pain: a systematic review of systematic reviews. Int J Gen Med. 2013. doi:http://dx.doi.org/10.2147/IJGM.S50243.
18. Bidonde J, Busch AJ, Bath B, Milosavljevic S. Exercise for adults with fibromyalgia: an umbrella systematic review with synthesis of best evidence. Curr Rheumatol Rev. 2014;10:45–79.
19. Ernst E, Canter PH. A systematic review of systematic reviews of spinal manipulation. J R Soc Med. 2006;99:192–6.
20. Gilbey A, Ernst E, Tani K. A systematic review of reviews of systematic reviews of acupuncture. Focus Altern Complement Ther. 2013. doi:10.1111/fct.12004.
21. Kendall S. What do we mean by evidence? Implication for primary health care nursing. J Interprof Care. 1997;11(1):23–34.
22. Milanese S. The use of RCT's in manual therapy – are we trying to fit a round peg into a square hole? Man Ther. 2011. doi:10.1016/j.math.2011.02.007.

# Case Study in Dentistry

# 15

Clovis Mariano Faggion Jr and Utsana Tonmukayakul

**Abstract**

Systematic reviews in dentistry have been consistently published in the last two decades, although the publication of overviews of these reviews is a more recent phenomenon. Most of these few overviews published use a variety of tools to evaluate the methodological quality of reviews included such as the Glenny, OQAQ, CASP, and AMSTAR tools. Probably, the most used is the AMSTAR checklist that is becoming widely recognized as a reference for evaluating reviews. These overviews identified methodological limitations in different levels: at primary study level (those included in the systematic review) and at systematic review level. Thus, an overview of reviews has a unique advantage of providing at the same time the "big picture" of evidence and an in-depth view of primary studies forming this evidence.

Most overviews reported in this chapter were developed in specific fields of dentistry, periodontology, and implant dentistry.

A selection of reviews on economic evaluation also shows that this subject should receive more attention in dentistry. Based on our best knowledge, up to date, there is no overview of systematic reviews of economic evaluation in dentistry. The published reviews suggested that the standard of conducting and reporting of economic evaluation studies should be improved. As with overviews of systematic reviews on clinical issues, these reviews identified several methodological limitations in primary studies evaluating economic aspects of dentistry.

C.M. Faggion Jr, DDS, Dr. Med. Dent, PD (✉)
Department of Periodontology and Operative Dentistry, Faculty of Dentistry, University of Münster, Waldeyerstraße 30, 48149 Münster, Germany
e-mail: clovisfaggion@yahoo.com

U. Tonmukayakul
Deakin Health Economics, Deakin University, Melbourne, Australia



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_15

## 15.1 Introduction

The overview of systematic reviews (SRs) plays an important role in the whole hierarchy of research [1]. The idea of having different levels of research evidence interconnected seems to be relevant in an era where a huge number of scientific documents are published every year. For example, animal experiments provide research output for planning and performing clinical trials; clinical trials are selected to produce SRs which are the basis for conducting evidence-based clinical guidelines [1]. Reliable results from each phase of the research process are pivotal for generating accurate output in the subsequent phases of the research process.

SRs with meta-analyses are considered the last step in this process, where these reviews are able to collect all potential information from primary studies already published. Anyway, it is almost impossible to have only one SR for each important research question. Different research groups around the world produce similar SRs about the same topic, but, in many cases, with different views on evaluating the evidence. So, to have a comprehensive view about the evidence of some specific topic, it is necessary to perform an overview of these reviews. Thus, the great advantage of an overview of SRs is that we may have all potential evidence condensed in only one document [2].

The concept of overviews of SRs is relatively new in dentistry. The current chapter will cover some examples of overviews of SRs published in some dental fields in order to provide the reader with information on the current knowledge in this field. Most cases are focused on the periodontology and implant dentistry fields. An overview of SRs on economic evaluation in oral health will be also reported.

## 15.2 Definitions

Some terms may be used to define documents which evaluate a sample of SRs. For example, a more general term is "overview" of reviews; "umbrella review" may be related to a specific type of overviews of reviews focused on a specific clinical topic (e.g., a drug or a condition). Meta-epidemiological studies may be overviews of reviews with a non-clinical first topic (e.g., they may focus on funding issues or small study effects). In this chapter, for the sake of simplicity, we will treat any document evaluating at least two SRs as "overview."

## 15.3 Overviews of Systematic Reviews in Dentistry

The field of dentistry is featured by presenting several clinical subfields. For example, periodontology deals with diseases affecting hard and soft tissues supporting the teeth; endodontology deals with diseases affecting tissues in the tooth root canals; orthodontics deals with the mechanical movement of teeth to correct the positioning and function of teeth, etc. Although these clinical fields are

heterogeneous on types of treatment delivered, the quality of evidence supporting the therapeutic approaches in these fields seems to be of similar level [3]. Most examples of overviews of systematic reviews in dentistry will be focused on a few dentistry subfields, but we understand, nevertheless, that the present information may be representative to other clinical fields in dentistry.

### 15.3.1 The Case of Periimplantitis Treatment

Periimplantitis is a disease which affects the soft and hard tissues supporting a dental implant, and it is primarily caused by bacterial plaque on the dental implant surface [4]. If not treated, periimplantitis might cause the loss of the supporting tissues and, consequently, the loss of the dental implant. An overview of SRs on this issue was published in 2010 [5]. In this overview, only two SRs on periimplantitis treatment met the eligibility criteria and were therefore included. Therapeutic approaches supported by these two SRs were subgingival debridement alone, subgingival debridement plus antibiotics, and regenerative procedures for more severe periimplantitis cases.

Although the methodological quality of these two SRs was considered rigorous, several limitations of primary studies included in these reviews were identified. For example, these limitations included trials with underpowered samples, very short-term follow-ups, and biased treatment arms, i.e., treatment arms with different combinations of therapies which avoided more robust conclusions about the effectiveness of therapies. In this overview, three checklists (CASP, QUOROM, and AMSTAR) were used to evaluate the methodological quality of SRs.

The Critical Appraisal Skills Programme (CASP) [6] was developed in the beginning of the 1990s, and it comprises several checklists related to different types of study design. The CASP checklist for SRs is formed by ten questions "to help the reader make sense of a review" as the checklist states. Questions should be answered with YES, I CAN'T TELL, and NO. Although the answer of these ten questions may improve reader's knowledge about an SR, there is a lack of in-depth information to clearly evaluate the methodological quality of an SR.

QUOROM (Quality Of RepOrting of Meta-analyses) [7] is a checklist composed by 21 headings and subheading regarding the report of pivotal parts of a meta-analysis. This checklist was now updated and renamed as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [8]. PRISMA checklist is endorsed by several scientific journals with the intention of providing clearer information about how the study was conducted. Table 15.1 reports items described in QUOROM and PRISMA checklists.

AMSTAR (Assessment of Multiple SysTemAtic Reviews) is a validated checklist formed by 11 items created to evaluate the methodological quality of SRs [9]. Authors of the checklist claim it has good face and content validity. Items in the AMSTAR are reported as questions that should be answered with YES, I CAN'T TELL, NO, and NOT APPLICABLE. This checklist is becoming a reference for evaluating SRs in several medical fields, including dentistry.

**Table 15.1** Items reported in two checklists used to evaluate the quality of reporting of systematic reviews

| Checklist | QUOROM | PRISMA |
|---|---|---|
| Title | Identify the report as a meta-analysis of RCTs | (1) Identify the report as a systematic review, meta-analysis, or both |
| Abstract | Use a structured format<br>Objectives: The clinical question explicitly<br>Data sources: The databases (i.e., list) and other information sources<br>Review methods: The selection criteria (i.e., population, intervention, outcome, and study design); methods for validity assessment, data abstraction, and study characteristics and quantitative data synthesis in sufficient detail to permit replication<br>Results: Characteristics of the RCTs included and excluded; qualitative and quantitative findings (i.e., point estimates and confidence intervals); and subgroup analyses<br>Conclusion: The main results | Structured summary: (2) Provide a structured summary including, as applicable background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number |
| Introduction | The explicit clinical problem, biological rationale for the intervention, and rationale for review | Rationale: (3) Describe the rationale for the review in the context of what is already known<br>Objectives: (4) Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS) |

| | | |
|---|---|---|
| Methods | Searching: The information sources, in detail (e.g., databases, registers, personal files, expert informants, agencies, hand searching), and any restrictions (years considered, publication status, language of publication)<br>Selection: The inclusion and exclusion criteria (defining population, intervention, principal outcomes, and study design)<br>Validity assessment: The criteria and process used (e.g., masked conditions, quality assessment, and their findings)<br>Data abstraction: The process or processes used (e.g., completed independently, in duplicate)<br>Study characteristics: The type of study design, participants' characteristics, details of intervention, outcome definitions, and how clinical heterogeneity was assessed<br>Quantitative data synthesis: The principal measures of effect (e.g., relative risk), method of combining results (statistical testing and confidence intervals), handling of missing data, how statistical heterogeneity was assessed, a rationale for any a priori sensitivity and subgroup analyses, and any assessment of publication bias | Protocol and registration: (5) Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number<br>Eligibility criteria: (6) Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale<br>Information sources: (7) Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched<br>Search: (8) Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated<br>Study selection: (9) State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)<br>Data collection process: (10) Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators<br>Data items: (11) List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made<br>Risk of bias in individual studies: (12) Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis<br>Summary measures: (13) State the principal summary measures (e.g., risk ratio, difference in means)<br>Synthesis of results: (14) Describe the methods of handling data and combining results of studies, if one, including measures of consistency (e.g., I2) for each meta-analysis<br>Risk of bias across studies: (15) Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies)<br>Additional analyses: (16) Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were prespecified |

(continued)

**Table 15.1** (continued)

| Checklist | QUOROM | PRISMA |
|---|---|---|
| Results | Trial flow: Provide a meta-analysis profile summarizing trial flow (figure)<br>Study characteristics: Present descriptive data for each trial (e.g., age, sample size, intervention, dose, duration, follow-up period) (table)<br>Quantitative data synthesis: Report agreement on the selection and validity assessment; present simple summary results (for each treatment group in each trial, for each primary outcome); present data needed to calculate effect sizes and confidence intervals in intention-to-treat analyses (e.g., 2×2 tables of counts, means and SDs, proportions) | Study selection: (17) Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram<br>Study characteristics: (18) For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations<br>Risk of bias within studies: (19) Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12)<br>Results of individual studies: (20) For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot<br>Synthesis of results: (21) Present results of each meta-analysis done, including confidence intervals and measures of consistency<br>Risk of bias across studies: (22) Present results of any assessment of risk of bias across studies (see item 15)<br>Additional analysis: (23) Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see item 16]) |
| Discussion | Summarize key findings; discuss clinical inferences based on internal and external validity; interpret the results in light of the totality of available evidence; describe potential biases in the review process (e.g., publication bias); and suggest a future research agenda | Summary of evidence: (24) Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers)<br>Limitations: (25) Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias)<br>Conclusions: (26) Provide a general interpretation of the results in the context of other evidence and implications for future research |
| Funding | Not available | Funding: (27) Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review |

### 15.3.2 The Case of Animal Research in Dentistry

An overview of SRs of animal research in dentistry was recently published [10]. In this overview, 54 SRs including animal experiments were evaluated with the AMSTAR checklist. Only two SRs (from 54) were considered at high methodological quality. One important feature of the reviews included in this overview was the eligibility criteria for including primary studies: most reviews ($n=41$) included animal experiments together with clinical trials to provide an overall view of evidence in different levels (animal and clinical). Only 13 reviews included only animal studies ($n=13$).

The average AMSTAR score in these 44 reviews was 3 (median = 2), and this score demonstrates how poor is the methodological quality of SRs in oral health including animals.

The information on animal research is pivotal for guiding the development and conducting of clinical trials in humans. The more biased is the information on animal research, the higher will be the probability of generating misleading information for clinical trials. This situation may raise ethical concern by the misuse of animals and potentially dangerous therapies for humans [1]. It is important to mention that SRs is not "able" to improve the quality of primary studies included in the review, but improve power and precision of treatment effect estimates [11]. In oral health, it seems that both levels of evidence have limitations: the quality of animal experiments as well as the quality of SRs of these experiments [10].

### 15.3.3 Can Periodontitis Affect the Survival of Dental Implants?

In this overview of SRs about the relationship between periodontitis and the outcome of dental implants [12], nine SRs and three meta-analyses were included. The methodological evaluation of SRs was performed with two checklists: AMSTAR and R-AMSTAR. The latter is a modified form of AMSTAR which comprises the 11 items contained in AMSTAR plus several subitems [13]. R-AMSTAR is intended to generate more in-depth information about AMSTAR items, and it was already validated [13]. The authors of R-AMSTAR suggest the generation of quantitative values for the methodological scores, although this approach might be misleading. It seems very difficult to provide adequate values for each item assessed. In other words, it is challenging to determine the contribution of any item to the overall quality of the review. And this assumption is valid for both AMSTAR and R-AMSTAR. For example, let's assume that some thresholds for quality are set in AMSTAR: up to 4 items met (low methodological quality); 4–7 items met (moderate methodological quality); and 8–11 items met (high methodological quality). If two SRs are compared, suppose that the first SR has the first four items met only, and therefore it might be considered of low methodological quality; the second SR has the last five items (7–11) met, and therefore it might be considered of moderate methodological quality. Nevertheless, for a specific research question, items may have stronger or weaker influence on the overall quality. For example, questions 1–4 may have more

importance regarding the overall quality than the other questions. How to determine the weight of each item? One may consider that the first SR is in fact with higher methodological quality than the second one. Thus, the approach should be not focused on quantitative scores when SRs are methodologically evaluated.

In this overview, SRs with highest methodological quality reported an association between history of periodontitis and risk of implant failure. Nevertheless, the SRs included presented several methodological limitations which may reduce our confidence on these findings. For example, methodological tools evaluating primary studies were too heterogeneous across SRs, which hinder any attempt for comparison. Another example of methodological limitation is the lack of comprehensiveness of literature search strategies used in these SRs. Only 2 from 9 SRs included met the requirements established by AMSTAR regarding the comprehensiveness of literature search. These results may imply in higher risk of publication bias [14].

### 15.3.4 Use of Systemic Antimicrobials for Treating Periodontitis

Periodontitis affects the hard and soft tissue supporting the teeth, and its severe form may cause tooth loss [15]. The consequences of tooth loss may considerably impact quality of life [16]. The conventional treatment of periodontitis is performed with the mechanical removal of pathogenic bacteria (periodontitis main cause) from the tooth root surfaces. In order to improve treatment outcomes, some forms and severity of periodontitis may be treated with the adjunctive use of systemic antimicrobials [17]. Nevertheless, some potential short- and long-term side effects may be expected with the use of antimicrobials, such as allergy and diarrhea and potential bacterial resistance [17].

An overview of SRs of the use of systemic antimicrobials in periodontal patients was performed to thoroughly understand potential benefits and harms after their use [17]. Interestingly, the evaluation of nine SRs included did not reveal any information about true endpoints in the form of tooth survival. In other words, it is now known whether the use of systematic antimicrobials to treat periodontitis may save more teeth than the standard mechanical therapy. Nevertheless, patients treated with mechanical treatment plus systemic antimicrobials presented better outcomes in terms of surrogate endpoints, i.e., clinical attachment gain and pocket depth reduction. These improvements were noticed only in cases with more severe forms of disease. Surrogate endpoints have been proved to be sometimes inaccurate to explain the final outcomes, for example, in the case of some drugs that suppress ventricular arrhythmias, which are associated with an almost fourfold increase in the risk for death related to cardiac complications [18]. Three drugs related to the suppression of arrhythmias were approved by the American Food and Drug Administration (FDA), but later they were found to cause higher rates of death than the placebo group [19]. This case illustrates the danger of only relying on surrogated endpoints to support clinical decision-making.

In this overview, the report of side effects was very scarce, with almost no mention about bacterial resistance, a growing general problem by using antimicrobials [17]. The methodological quality of SRs included was evaluated with AMSTAR, which showed heterogeneity in the quality of reviews included.

### 15.3.5 Root Coverage Procedures

One of the effects of periodontitis is the recession of soft tissues covering the root surfaces. Recession may be also caused by daily traumatic tooth cleaning. In many cases, tooth recession does not require treatment and it might become stable over time. Nevertheless, in some cases, the recession needs to be treated to reduce tooth sensitivity or improve aesthetic outcomes [20]. An overview of SRs to evaluate the effectiveness of root coverage (RC) procedures in the treatment of recession-type defects was conducted [21]. In this overview, three checklists (Glenny, OQAQ, and AMSTAR) were used to evaluate the methodological quality of SRs included.

Glenny et al. [22] and OQAQ (Overview Quality Assessment Questionnaire) [23] checklists have similar questions, although Glenny checklist has a greater number of questions compared to OQAQ (15 and 10, respectively). OQAQ was one of the checklists used as basis for the development of AMSTAR [9]. In this overview, nine SRs were paper based and one was a Cochrane review. There was heterogeneity regarding the methodological quality of included SRs. Some methodological limitations include incomprehensive literature search strategies, unclear inclusion and exclusion criteria for selecting primary studies, and inadequate quality assessment approach used to evaluate primary studies included in the SRs. Only two (from ten SRs) followed in full the guidelines proposed by the Cochrane Collaboration. Table 15.2 reports items described in Glenny, OQAQ, CASP, and AMSTAR checklists.

### 15.3.6 Search Strategies Used in SRs

A recent meta-epidemiologic overview evaluated the search strategies used by authors of SRs published in the fields of periodontology and implant dentistry [24]. In this overview, some criteria defined by the Cochrane Collaboration and AMSTAR checklist were used to evaluate a large sample of SRs with meta-analyses ($N=146$). Although the great majority of SRs were paper based, Cochrane reviews ($N=19$) presented more complete search strategies than paper-based ones. The evaluation of some items in both types of reviews (paper based and Cochrane), such as the search for gray literature and no language restriction, demonstrated that there is still room for improving search strategies in these reviews. Search of gray literature and search without language restriction were performed in 34 % and 50 % of evaluated reviews, respectively. Electronic searches in more than one database were more consistently performed in Cochrane than paper-based reviews (100 % and 55 % of evaluated Cochrane and paper-based reviews, respectively).

**Table 15.2** Items reported in three checklists used to evaluate the methodological quality of systematic reviews

| Checklist | | | |
|---|---|---|---|
| Glenny | OQAQ | CASP | AMSTAR |
| 1. Did review address a focused question?<br>2. Did authors look for appropriate papers?<br>3. Do you think authors attempted to identify all relevant studies?<br>4. Did authors search for published and unpublished literature?<br>5. Were all languages considered?<br>6. Was any hand searching carried out?<br>7. Was it stated that the inclusion criteria were carried out by at least two reviewers?<br>8. Did reviewers attempt to assess the quality of the included studies?<br>9. If so, did they include this in the analysis?<br>10. Was it stated that the quality assessment was carried out by at least two reviewers?<br>11. Are the results given in a narrative or pooled statistical analysis?<br>12. If the results have been combined, was it reasonable to do so?<br>13. Are the results clearly displayed?<br>14. Was an assessment of heterogeneity made and reasons for variation discussed?<br>15. Were results of review interpreted appropriately? | 1. Were the search methods reported?<br>2. Was the search comprehensive?<br>3. Were the inclusion criteria reported?<br>4. Was selection bias avoided?<br>5. Were the validity criteria reported?<br>6. Was validity assessed appropriately?<br>7. Were the methods used to combine studies reported?<br>8. Were the findings combined appropriately?<br>9. Were the conclusions supported by the reported data?<br>10. What was the overall scientific quality of the overview? | 1. Did the review address a clearly focused question?<br>2. Did the authors look for the right type of papers?<br>3. Do you think all the important, relevant studies were included?<br>4. Did the review's authors do enough to assess the quality of the included studies?<br>5. If the results of the review have been combined, was it reasonable to do so?<br>6. What are the overall results of the review?<br>7. How precise are the results?<br>8. Can the results be applied to the local population?<br>9. Were all important outcomes considered?<br>10. Are the benefits worth the harms and costs? | 1. Was an "a priori" design provided?<br>2. Was there duplicate study selection and data?<br>3. Was a comprehensive literature search performed?<br>4. Was the status of publication (i.e., gray literature) used as an inclusion criterion?<br>5. Was a list of studies (included and excluded) provided?<br>6. Were the characteristics of the included studies provided?<br>7. Was the scientific quality of the included studies assessed and documented?<br>8. Was the scientific quality of the included studies used appropriately in formulating conclusions?<br>9. Were the methods used to combine the findings of studies appropriate?<br>10. Was the likelihood of publication bias assessed?<br>11. Was the conflict of interest stated? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies |

The planning and conducting of a comprehensive search strategy in any SR is pivotal for providing unbiased treatment effect estimates of therapies. Ideally, all relevant literature related to the proposed research question should be identified, with the risk of biasing estimates in case relevant information is left out. It is well established that studies with more positive or with more/less inflated estimates have higher chances of being published or published sooner than those studies not providing significant differences between therapies [25]. Publication bias may occur in different levels of research, for example, in animal research [26] as well as clinical trials [25].

### 15.3.7 Methodological Tools in SRs

An important aspect of an SR is the evaluation of quality/risk of bias (ROB) of primary studies included in the SR. Some evidence suggests that studies with high ROB may generate more inflated treatment effect estimates than those from studies with low ROB [27]. So it seems relevant that all SRs present a reliable way to evaluate primary studies.

An overview of a sample of 159 SRs with meta-analyses published in periodontology and implant dentistry evaluated the approaches used to assess quality/ROB of primary studies included in the SRs [28]. In this overview, the most prevalent approach to evaluate primary studies was the domain based, followed by checklists and scales. The Cochrane Collaboration recommends the evaluation of seven areas which may be sensitive to bias in a clinical study (i.e., randomized controlled trial): sequence generation, allocation concealment, blinding of participants, personnel and outcome assessors, incomplete outcome data, selective outcome reporting, and other bias [29]. The main advantage of domain-based approaches over checklists and scales is that they allow the transparent report of the rationale used to support the judgment about different ROB levels. So, the readers receive comprehensive information to judge by themselves whether the results of the assessment are valid. From this group of tools (domain based, checklist, and scale), the scale has probably the most important limitations, because it normally involves scores and thresholds. Thus, as it was reported earlier in the text, the final results may be misleading.

The evaluation of these 159 SRs found several limitations on the approaches used by authors of SRs to evaluate primary studies. For example, authors normally do not differentiate quality from ROB or sometimes mix different concepts in their assessments such as methodological quality and quality of reporting [29]. Furthermore, there were some improvements over the years regarding some aspects of the use of methodological tools. For example, from 2010 to 2013, SRs published in periodontology and implant dentistry used more validated tools for evaluating methodological quality of primary studies than in previous years. Another improvement in SRs published in more recent years was the level of reporting of overall quality/ROB of primary studies included in the SRs.

### 15.3.8 Reporting Abstracts of SRs

The reporting of abstracts of SRs is important to inform readers about the relevance or not of the study. Monthly, a huge number of studies are published, including SRs and meta-analyses, which makes difficult to the reader to evaluate in depth all this material. Thus, many readers have the opportunity to check the abstract only. Therefore, the abstract should have enough details to inform the reader whether the reading of the full-text article is necessary.

A recent overview of abstracts of SRs with meta-analyses in periodontology and implant dentistry revealed a lack of uniform standard of reporting [30], and this lack of standard may hinder the understanding of the reader. For example, the strength of evidence and measure of precision (i.e., confidence interval) were reported in only less than half the selected abstracts. Measures of consistency in meta-analyses such as $I^2$ statistics [31] were reported in only 5 % of the selected sample of abstracts. In this sample of SRs, Cochrane abstracts reported the limitations of evidence and precision better than paper-based ones.

### 15.3.9 Reporting Funding Source in SRs

One important issue when publishing research is the disclosure of potential conflict of interests of authors and the economic support of this research. Some evidence suggests that research supported by industry may generate more positive results than research supported by nonprofit organizations [32]. This issue has been receiving much attention in the last few years, and some consider them so important that suggest their inclusion as item for evaluating ROB in studies [33].

One overview of SRs evaluated the type of sponsorship of SRs published in periodontology and implant dentistry [34]. From 146 SRs evaluated, only 14 (10 %) SRs discussed the possible influence of sponsorship on the meta-analytic findings. This apparent inadequate evaluation may negatively impact a comprehensive understanding of the research findings, because authors of SRs may also have some conflict of interest on providing more "positive" information about one of the therapies being evaluated in the SR. The 11th item of AMSTAR checklist recommends the evaluation of both sponsorship information of the SR and all studies included in the SR.

### 15.3.10 The Case of Economic Evaluation in Dentistry

Economic evaluation is a comprehensive analysis that considers and compares all costs and outcomes associated with healthcare interventions [35]. Economic evaluation provides evidence-based information for clinicians and decision-makers on how to efficiently use limited resources and whether such intervention could offer good value for money. This technique was firstly applied in dentistry in the 1970s [36] and has been employed vastly on dental caries prevention programs e.g. community water fluoridation, application of different forms of fluoride, etc. The

economic aspect has been considered in clinical guideline development for the management of wisdom teeth [37] and dental check-up [38] by the National Institute for Clinical Excellence, UK [39, 40]. Due to the increased awareness of the importance of evidence-based healthcare policy, the number of economic evaluations has increased in recent years. As a consequence, there is a demand for methods to synthesize and interpret the results of the previous studies that address the same issue which can be done by applying the systematic review framework.

Critical appraisal is a standard practice in conducting systematic review. In economic evaluation literature review, the Drummond checklist is a common tool and it is recommended in the Cochrane review handbook [41]. The Drummond checklist considers description of interventions, study design, identification detail of intervention, measurement and valuation of costs and consequences, discounting, clear results with sensitivity and uncertainty analysis, and discussion of results in context of policy relevance and existing literature. The Drummond checklist has also been widely used as a guideline of reporting of economic evaluation studies. The Drummond checklist is a prototype of economic evaluation checklists which were developed, e.g., Adams et al. [42], Gerard [43], and the *British Medical Journal (BMJ)* Checklist [44]. Up to date, there is no overview of reviews of economic evaluation in dentistry. Therefore, this section will report some reviews on economic evaluation in dentistry. Special attention will be given to methodological limitations in economic evaluation in dentistry, retrieved from existing SRs. Some information on strengths and weaknesses of checklists to evaluate economic studies included in SRs will be briefly reported. This information may guide author in the development of future SRs on this important topic.

To begin with, a recent SR assessed the methodological quality of published economic evaluation in dentistry [45]. The authors adopted suggested search terms by the Cochrane Oral Health Group and McMaster costing with the Medical Subject Heading (MeSH). Electronic searches of MEDLINE, NHS Economic Evaluation Database, and the Cochrane library from 1975 to 2013 were conducted to identify publications considering costs and outcomes in dentistry. Hand searching from relevant reference lists was also performed. Three independent authors reviewed the retrieved publications for inclusion and one author applied the 10-point Drummond checklist for quality appraisal of the publications [44]. Median scores, a conservative approach, were used as cutoff point to identify well-performed studies.

One hundred and fourteen publications were assessed with 73 (79 %) of both partial and full economic evaluations that were published after the year 2000. The majority of the existing publications (53 %) were about caries preventive procedures. An increase in proportion of partial economic evaluation was observed. Partial economic evaluation focuses on costs of an intervention (with or without comparing with its alternative), whereas full economic evaluation compares all costs and outcomes of at least two alternatives [35]. Of the 114 publications, 35 were partial economic evaluations and the other 79 publications are full economic evaluation. A significant improvement in quality of partial economic evaluations published after the year 2000 was observed. However, such trend was not distinctive among full economic evaluations.

This SR identified a few methodological pitfalls in published economic evaluation in dentistry, e.g., confusion over the use of labels for types of economics evaluation, neglected to reported sensitivity analysis and discount information, and insufficient details of measurement and valuation of outcomes and costs. These methodological limitations were also reported in previous systematic reviews.

A literature review done by Yule et al. [46] published in 1986 pointed out three problems in economics in dentistry: (a) enumerating and measuring costs, (b) enumerating and measuring benefits, and (c) the timing of costs and benefits, and uncertainty. The authors found that opportunity costs, e.g., travel costs of the patients, time spent on traveling to the dentist, etc., were often neglected in some of the economic evaluations in dentistry. Moreover, the costs and benefits occurring over time in dental care programs were not always converted to present values. In other words, the technique of discounting was lightly employed. The authors also highlighted serious problem of the measures of the success for DMFS measure. The DMFS index was used in a number of cost-effectiveness analysis of caries preventions. Because this DMFS measure is a summation of all decayed, missing, and filled surfaces, it cannot identify dental health implication of a losing tooth and resource savings that could yield from the reduction in decayed, missing, and filled surfaces.

Another SR that was published in 2003 by Källestål et al. tried to identify the most cost-effectiveness dental caries preventions [47]. The authors comprehensively assessed the quality of the clinical evidences together with the health-economic quality appraisal tool called Drummond checklist. The results of reviewing 74 references that identified from MEDLINE from 1966 to May 2003 found the lacking of well-conducted studies as well as contradictory evidence in the existing economic evaluation studies. Therefore, it is difficult to identify which dental caries prevention methods are offering best value for money.

An SR of Mariño et al. aimed to objectively retrieve, synthesize, and describe available information of economic evaluation of dental caries prevention programs [48]. Out of 67 eligible literatures which were electronically searched from MEDLINE, EconLit, and ISI from January 1975 to April 2012, the authors found a steady increase in number of economic evaluation of dental caries prevention in the early 2010s, but there was room for improving the quality of reporting technical aspects. Confusion over the appropriate labels was evident in some existing studies, e.g., full economic analysis was assigned to a study that failed to have a comparator. Although most published economic evaluations of dental caries prevention clearly stated and explained the importance of performing and economic analysis, most failed to justify the reason of lack of information on the perspective of analysis and mostly neglected to report sensitivity analysis and discounting information. These findings were also consistent with the systematic review conducted by Källestål et al. above in which economic evaluation in dentistry was low in evidence value with contradictory results.

Another SR of economic evaluation of diagnostic methods applied to dentistry also emphasized that there was a room for improvement in economic evaluation in dentistry [49]. The authors searched literature in Medline, Web of Science, The Cochrane library, and the NHS Economic Evaluation databases. Hand searching

was performed up to February 2013 to complement the literature search. Reviewers applied a protocol based on the QUADAS tool regarding diagnostic methods and the Drummond checklist for economic evaluation. The result indicated that the choice of economic evaluation method was not well justified and the perspective of the study was not stated in the previous publications.

Economic evaluation considers all costs and outcomes of the interested interventions. So both costs and effectiveness data must be based on a good quality of evidence. Findings of existing SRs up to date highlighted the lack of sound effectiveness data. Thus, these methodological limitations inhibit cost-effectiveness analysis in dentistry, particularly in the dental caries field. O'Connell and Griffin [50] provided an overview of strategies for conducting economic evaluation with an emphasis on costing methods of oral health interventions which can considerably improve the weaknesses indicated in the existing SRs.

Besides the methodological pitfalls in economic evaluation in dentistry as outlined above, the critical appraisal tool that is generally applied in SR of economic evaluation in healthcare also contains some limitations. Quality assessment tools allow systematic and transparent assessment of methodological quality of the reviewed studies. The Drummond checklist was initially developed as a guideline for economic evaluation reporting, but it has been widely used for methodological quality assessment purpose. The Drummond checklist contains items that are related to quality and transparency of the results reporting and interpretation of results rather than validity (Table 15.3). In fact, there is no gold standard critical appraisal tool in economic evaluation [48], and validity and reliability for assessing the methodological quality of economic evaluations has been little examined. Because of the primary purpose of the Drummond checklist, there is an absence of scoring scheme in which these explicit quantitative measures allow economic evaluations to be ranked according to their quality. There is no suggestion on how to score, weigh

**Table 15.3** Drummond checklist [35]

| Items | Criteria |
|---|---|
| 1 | Was a well-defined question posed in an answerable form? |
| 2 | Was a comprehensive description of the competing alternatives given? (that is, can you tell who did what to whom, where and how often?) |
| 3 | Was the effectiveness of the programmes or services established? |
| 4 | Were all the important and relevant costs and consequences for each alternative identified? |
| 5 | Were costs and consequences measured accurately in appropriate physical units (for example, hours of nursing time, number of physician visits, lost work-days, gained life-years)? |
| 6 | Were costs and consequences valued credibly? |
| 7 | Were costs and consequences adjusted for differential timing? |
| 8 | Was an incremental analysis of costs and consequences of alternatives performed? |
| 9 | Was allowance made for uncertainty in the estimates of costs and consequences? |
| 10 | Did the presentation and discussion of study results include all issues of concern to users? |

each item of the Drummond checklist, as well as the cutoff point to decide whether to include or exclude such publications in the review. The impact of critical appraisal on the decision to include economic evaluation studies that meet some but not all criteria of methodological study was little known [51]. However, other well-developed checklists that had excellent reliability and validity, e.g., the BMJ checklist could plausibly be used as quality assessment tools.

SRs of economic evaluations are challenging because economic evaluation studies are based on a unique set of results and contexts. The comparison and combination of the results are difficult because of the nature of economic evaluation research methodology. Given inherent limitation of the existing critical appraisal tool, some methodological pitfalls of economic evaluations in dentistry were identified.

**Conclusion**

Dentistry has been the focus of many interesting efforts at evidence synthesis based on secondary or tertiary level analyses (i.e., umbrella reviews, overviews of reviews, and meta-epidemiologic studies). Whereas the limitations inherent to several studies conducted in the past in this field cannot be overcome by such study designs, the examples presented in the prior sections of this chapter highlight the great potential of these novel approaches to evidence integration and appraisal.

## References

1. Faggion Jr CM. Animal research as a basis for clinical trials. Eur J Oral Sci. 2015;123:61–4.
2. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011;11:15.
3. Faggion Jr CM. Is the evidence supporting dental procedures strong? A survey of Cochrane systematic reviews in oral health. J Evid Based Dent Pract. 2012;12:131–4.
4. Lindhe J, Meyle J, Group D of European Workshop on Periodontology. Peri-implant diseases: Consensus Report of the Sixth European Workshop on Periodontology. J Clin Periodontol. 2008;35(8 Suppl):282–5.
5. Faggion Jr CM, Schmitter M. Using the best available evidence to support clinical decisions in implant dentistry. Int J Oral Maxillofac Implants. 2010;25:960–9.
6. Critical Appraisal Skills Programme (CASP). 2014. CASP Checklists. http://media.wix.com/ugd/dded87_a02ff2e3445f4952992d5a96ca562576.pdf. Oxford: CASP.
7. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet. 1999;354:1896–900.
8. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009;62:1006–12.
9. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
10. Faggion Jr CM, Listl S, Giannakopoulos NN. The methodological quality of systematic reviews of animal studies in dentistry. Vet J. 2012;192:140–7.
11. Akobeng AK. Understanding systematic reviews and meta-analysis. Arch Dis Child. 2005;90:845–8.
12. Faggion Jr CM, Giannakopoulos NN. Critical appraisal of systematic reviews on the effect of a history of periodontitis on dental implant loss. J Clin Periodontol. 2013;40:542–52.

13. Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, Maida CA. From systematic reviews to clinical recommendations for evidence-based health care: validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J. 2010;4:84–91.
14. Sterne JAC, Egger M, Moher D, editors. Addressing reporting biases, chap. 10. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of intervention, version 5.1.0 (updated March 2011). The Cochrane Collaboration. 2011. Available from www.cochrane-handbook.org.
15. Page RC, Engel LD, Narayanan AS, Clagett JA. Chronic inflammatory gingival and periodontal disease. JAMA. 1978;240:545–50.
16. Gerritsen AE, Allen PF, Witter DJ, Bronkhorst EM, Creugers NH. Tooth loss and oral health-related quality of life: a systematic review and meta-analysis. Health Qual Life Outcomes. 2010;8:126.
17. Faggion Jr CM, Cullinan MP, Atieh M, Wasiak J. An overview of systematic reviews of the use of systemic antimicrobials for the treatment of periodontitis. Br Dent J. 2014; 217:443–51.
18. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. N Engl J Med. 1991;324:781–8.
19. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;125:605–13.
20. Douglas de Oliveira DW, Oliveira-Ferreira F, Flecha OD, Gonçalves PF. Is surgical root coverage effective for the treatment of cervical dentin hypersensitivity? A systematic review. J Periodontol. 2013;84:295–306.
21. Chambrone L, Faggion Jr CM, Pannuti CM, Chambrone LA. Evidence-based periodontal plastic surgery: an assessment of quality of systematic reviews in the treatment of recession-type defects. J Clin Periodontol. 2010;37:1110–8.
22. Glenny AM, Esposito M, Coulthard P, Worthington HV. The assessment of systematic reviews in dentistry. Eur J Oral Sci. 2003;111:85–92.
23. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44:1271–8.
24. Faggion Jr CM, Atieh MA, Park S. Search strategies in systematic reviews in periodontology and implant dentistry. J Clin Periodontol. 2013;40:883–8.
25. Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. Cochrane Database Syst Rev. 2007;(2):MR000011.
26. ter Riet G, Korevaar DA, Leenaars M, Sterk PJ, Van Noorden CJ, Bouter LM, Lutter R, Elferink RP, Hooft L. Publication bias in laboratory animal research: a survey on magnitude, drivers, consequences and potential solutions. PLoS One. 2012;7, e43404.
27. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352:609–13.
28. Faggion Jr CM, Huda F, Wasiak J. Use of methodological tools for assessing the quality of studies in periodontology and implant dentistry: a systematic review. J Clin Periodontol. 2014;41:625–31.
29. Higgins JPT, Altman DG, Sterne JAC, editors. Assessing risk of bias in included studies, chap. 8. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March 2011). The Cochrane Collaboration. 2011. Available from www.cochrane-handbook.org.
30. Faggion Jr CM, Liu J, Huda F, Atieh M. Assessment of the quality of reporting in abstracts of systematic reviews with meta-analyses in periodontology and implant dentistry. J Periodontal Res. 2014;49:137–42.
31. Fletcher J. What is heterogeneity and is it important? BMJ. 2007;334:94–6.
32. Lundh A, Sismondo S, Lexchin J, Busuioc OA, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev. 2012;12:MR000033.

33. Bero LA. Why the Cochrane risk of bias tool should include funding source as a standard item. Cochrane Database Syst Rev. 2013;12:ED000075.
34. Faggion Jr CM, Atieh M, Zanicotti DG. Reporting of sources of funding in systematic reviews in periodontology and implant dentistry. Br Dent J. 2014;216:109–12.
35. Drummond MF, Sculpher MJ, Torrance GW, O'Brien B, Stoddart GL. Method for the economic evaluation of health care programmes. New York: Oxford University Press; 2005.
36. Niessen LC, Douglass CW. Theoretical considerations in applying benefit-cost and cost-effectiveness analyses to preventive dental programs. J Public Health Dent. 1984;44(4): 156–68.
37. Song F, O'Meara S, Wilson P, Golder S, Kleijnen J. The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth. Health Technol Assess. 2000;4:1–55.
38. Davenport C, Elley K, Salas C, Taylor-Weetman CL, Fry-Smith A, Bryan S, et al. The clinical effectiveness and cost-effectiveness of routine dental checks: a systematic review and economic evaluation. Southampton: Health Technol Assess. 2003;7(7):1–138.
39. National Institute for Clinical Excellence. Guidance on the Extraction of Wisdom Teeth National Institute for Clinical Excellence. 2000. Available from: http://www.nice.org.uk/nicemedia/pdf/wisdomteethguidance.pdf.
40. National Institute for Clinical Excellence. Dental recall: recall interval between routine dental examination: National Institute for Clinical Excellence. 2004. Available from: http://www.nice.org.uk/nicemedia/live/10952/29486/29486.pdf.
41. Shemilt I, Mugford M, Byford S, Drummond M, Eisenstein E, Knapp M, Mallender J, Mcdaid D, Vale L, Walker D. 15.5.2 Critical appraisal of methodological quality. In: Higgins JPT, Sally G, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley-Blackwell; 2011.
42. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. Med Care. 1992;30(3):231–43.
43. Gerard K. Cost-utility in practice: a policy maker's guide to the state of the art. Health Policy. 1992;21(3):249–79.
44. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ. 1996;313(7052): 275–83.
45. Tonmukayakul U, Calache H, Clark R, Wasiak J, Faggion CM Jr. Systematic review and quality appraisal of economic evaluation publications in dentistry. J Dent Res. 2015;94:1348–54.
46. Yule BF, van Amerongen BM, van Schaik MCM. The economics and evaluation of dental care and treatment. Soc Sci Med. 1986;22(11):1131–9.
47. Källestål C, Norlund A, Söder B, Nordenram G, Dahlgren H, Petersson LG, et al. Economic evaluation of dental caries prevention: a systematic review. Acta Odontol Scand. 2003;61: 341–6.
48. Mariño R, Khan A, Morgan M. Systematic review of publications on economic evaluations of caries prevention programs. Caries Res. 2013;47:265–72.
49. Christell H, Birch s, Horner K, Lindh C, Rohlin M, The SEDENTEXCT consortium. Economic evaluation of diagnostic methods used in dentistry. A systematic review. J Dent. 2014;42: 1361–71.
50. O'Connell JM, Griffin S. Overview of methods in economic analyses of behavioral interventions to promote oral health. J Public Health Dent. 2011;71:S101–18.
51. Walker DG, Wilson RF, Sharma R, Bridges J, Niessen L, Bass EB, Frick K. Best practices for conducting economic evaluations in health care: a systematic review of quality assessment tools. Methods research report. (Prepared by Johns Hopkins University Evidence-based Practice Center under contract No. 290-2007-10061-I.) AHRQ Publication No. 12(13)-EHC132-EF. Rockville: Agency for Healthcare Research and Quality; 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

# Case Study in Neonatology

16

David Andrew Osborn

**Abstract**

Neonatology represents per se a challenging clinical setting, where decision-making faces multifaceted issues and where the conduct of high-quality and large randomised trials is not easy. Umbrella reviews, overviews of reviews and meta-epidemiologic studies offer an important and efficient set of tools to integrate the evidence base in this peculiar discipline, while pinpointing areas of uncertainty either because of poor precision in effect estimates or at risk of inaccuracy due to sources of bias.

## 16.1 Introduction

The case study in neonatology is an overview of systematic reviews with the primary aim of preventing or minimising the use of mechanical ventilation in newborn care. Newborn infants are at high risk of respiratory distress requiring admission to a neonatal intensive care unit (NICU) and need for respiratory support. This is associated with substantial neonatal mortality, morbidity [1] and subsequent neurodevelopmental impairments. Many treatments have been found in systematic review to reduce the burden or preterm delivery, mature the foetus, reduce admission to NICU, prevent respiratory distress and avoid mechanical ventilation and/or reduce its duration. Some of these treatments have also been found to reduce mortality and

D.A. Osborn, MB BS, MM, FRACP, PhD
Central Clinical School, Discipline of Gynecology, Obstetrics and Neonatology, University of Sydney, Sydney, NSW, Australia

RPA Newborn Care, Royal Prince Alfred Hospital, Missenden Road, Camperdown, Sydney, NSW 2050, Australia
e-mail: david.osborn@sydney.edu.au



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_16

incidence of neurodevelopmental impairments in the survivors of neonatal intensive care. The overview is prevention focused so assumes the neonatologist/paediatrician will have both an antenatal as well as postnatal focus.

## 16.2 Recognising Risk for Mechanical Ventilation

Need for mechanical ventilation is almost universal for infants born extremely premature (<28 weeks gestation) [1] with rapidly reducing rates for very preterm (28 to <32 weeks) [1], late preterm (33 to <37 weeks) [2–4], near-term (37–38 weeks) and term gestation infants [4–7] (Tables 16.1 and 16.2). Recognising risk allows appropriate targeting of treatments for prevention and minimisation of respiratory morbidity. Gestational age and absence of labour are the major determinants of risk. In determining optimal place of delivery (in a hospital capable of providing ventilator support), to avoid a risk >20 % for ventilator support, the cut off is 33 weeks for mothers who labour, but is increased to 35 weeks when there is no labour (elective caesarean section) [2, 10].

Common diagnoses in premature infants are respiratory distress syndrome (hyaline membrane disease or surfactant deficiency) [11, 12] and congenital pneumonia [13]. Common diagnoses at near-term gestation are respiratory distress syndrome [4, 6, 7, 9, 14] and transient tachypnoea of the newborn [14]. The term infant is at lowest risk of respiratory morbidity [5, 6, 8, 14–17], with common diagnoses including persistent pulmonary hypertension, transient tachypnoea of the newborn, congenital pneumonia, pneumothorax and meconium aspiration syndrome [4]. Postmature infants (≥42 weeks gestation) are at increased risk of perinatal asphyxia and meconium aspiration syndrome [4, 8, 18]. The clinician should recognise the antecedents to these diagnoses including clinical presentations associated with prematurity increasing the likelihood of lung immaturity (respiratory distress syndrome) and infection, preterm rupture of membranes or clinical chorioamnionitis increasing the likelihood of infection, and postmature delivery increasing the likelihood of meconium aspiration syndrome. Infants born by elective caesarean section, especially before 39 weeks gestation, are at much higher risk of respiratory distress syndrome [4, 6, 7, 9, 14].

**Table 16.1** Survival, risk of mechanical ventilation, chronic lung disease and use of home oxygen in infants ($n=4454$) born between $23^{+0}$ and $31^{+}$ [6] weeks gestation 2007–2011 [1]

| | Gestation (weeks) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Outcome | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
| Overall survival % | 25 | 60 | 75 | 85 | 90 | 95 | 97 | 98 | 99 |
| Mechanical ventilation % | >95 | >95 | 95 | 90 | 80 | 70 | 55 | 40 | 30 |
| Chronic lung disease (36 weeks) postmenstrual age % | 25 | 45 | 50 | 45 | 30 | 20 | 10 | <5 | <5 |
| Home oxygen % | 10 | 20 | 15 | 15 | 5 | 5 | <5 | <1 | <1 |

**Table 16.2** Risk of serious respiratory morbidity or mechanical ventilation in population cohorts of late preterm and term infants

| Outcome | Years | *n* | Gestation (weeks) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | ≥42 |
| Mechanical ventilation % | 1992–1999 [2] | 3102 | 62 | 47 | 28 | 19 | 13 | 7 | 4 | | | | | | |
| Mechanical ventilation 1st day % | 1999–2002 [7] | 13,258 | | | | | | | | 1.9 | 0.9 | 0.4 | 0.4 | 0.4 | 0 |
| Mechanical ventilation % | 1997–2006 [3] | 7434[a] | | | | | 6.9 | 1.8 | 1.0 | 0.2 | | | | | |
| Mechanical ventilation % | 2010 [5] | 2273 | | | | | | | | * | 0.3 | 0.2 | | | |
| Mechanical ventilation % | 2003 [8] | 2,527,766[a] | | | | | | | | 0.57 | 0.32 | 0.28 | 0.29 | 0.38 | |
| Serious respiratory morbidity[b] % | 1990–1998 [4] | 14,531 | | | | | 22 | 8.5 | 3.9 | 1.4 | 1.1 | 0.7 | 0.8 | 1.1 | 1.6 |
| Serious respiratory morbidity[c] in elective caesarean deliveries % | 1998–2006 [9] | 34,458 | | | | | | | | 1.9 | 0.9 | 0.2 | 0.0 | 0.0 | |
| Serious respiratory morbidity[c] in intended vaginal deliveries % | | | | | | | | | | 0.4 | 0.2 | 0.1 | 0.1 | 0.1 | |

[a]Low-risk deliveries only
[b]Any ventilator support or the use of oxygen supplement for more than 24 h
[c]Oxygen therapy for more than 2 days, nasal continuous positive airway pressure or need for mechanical ventilation
*Numbers too small for accurate estimate

The following forms a clinician's checklist for decreasing risk of ventilation of newborn infants. The data is derived from an overview of predominately Cochrane Database of Systematic Reviews reporting the evidence from randomised controlled trials for effectiveness of interventions in pregnant women and infants. Data for major respiratory outcomes, mortality and critical morbidities are given where appropriate.

## 16.3 Antenatal Preparation for Delivery: Prevention of Preterm Birth (Table 16.3)

Interventions with proven efficacy for preventing preterm birth in women at risk of premature delivery include use of tocolytics for inhibition of labour, progesterone and cervical cerclage.

### 16.3.1 Tocolytics for Threatened Preterm Labour

Calcium channel blockers (e.g. nifedipine) are more effective than other tocolytic agents in reducing births within 7 days of treatment and before 34 weeks gestation [19]. They reduce the incidence of respiratory distress syndrome (9 trials, 763 infants; RR 0.63, 95 % CI 0.46, 0.88), intraventricular haemorrhage, necrotising enterocolitis and jaundice. There was a marked reduction in adverse maternal side effects compared to use of other tocolytics (e.g. betamimetics). Care should be taken in applying the evidence as trials excluded women who had contraindications to calcium channel blockers or betamimetics including foetal distress, chorioamnionitis, severe preeclampsia and abruptio placentae.

### 16.3.2 Tocolytics for Women with Preterm Premature Rupture of Membranes (PPROM)

PPROM accounts for one-third of preterm births. However, the evidence for improving respiratory outcomes for women with PPROM is less clear. In women with PPROM, use of tocolytics (8 trials/408 women) reduced birth within 48 h of treatment, but had no effect on perinatal death and increased maternal morbidity including chorioamnionitis [20]. There was no effect on respiratory distress syndrome but an increased need for mechanical ventilation of the newborn (1 trial, 81 infants; RR 2.46, 95 % CI 1.14, 5.34). Tocolytics cannot be recommended for women presenting with PPROM as they do not affect mortality and increase the risk of mechanical ventilation.

### 16.3.3 Antibiotics for Women with PPROM

Antibiotics for preterm premature rupture of membranes [21] (22 trials, 6872 women and their babies) increase duration of pregnancy, reduce maternal and infant infection but do not change perinatal mortality. Improved respiratory

**Table 16.3** Systematic reviews of antenatal interventions for preventing respiratory morbidity and mechanical ventilation in preterm infants

| Intervention | Outcome | Trials/participants | Effect [95 %CI] |
|---|---|---|---|
| Calcium channel blockers for inhibiting preterm labour [19] | Birth before 34 weeks gestation | 3/328 | RR 0.79 [0.65, 0.96] |
| | Perinatal mortality | 10/810 | RR 1.65 [0.74, 3.64] |
| | Respiratory distress syndrome | 9/763 | RR 0.63 [0.46, 0.88] |
| Tocolytics for preterm premature rupture of membranes [20] | Birth within 48 h | 6/354 | RR 0.55 [0.32, 0.95] |
| | Perinatal mortality | 7/402 | RR 1.67 [0.85, 3.29] |
| | Respiratory distress syndrome: all studies | 5/279 | RR 0.90 [0.70, 1.17] |
| | Studies using antenatal corticosteroids | 2/95 | RR 0.88 [0.55, 1.41] |
| | Intubation/mechanical ventilation | 1/81 | RR 2.46 [1.14, 5.34] |
| Antibiotics for preterm premature rupture of membranes [21] | Perinatal mortality | 12/6301 | RR 0.93 [0.76, 1.14] |
| | Birth within 7 days | 7/5965 | RR 0.79 [0.71, 0.89] |
| | Respiratory distress syndrome | 12/6287 | RR 0.95 [0.83, 1.09] |
| | Intubation/mechanical ventilation | 2/4924 | RR 0.90 [0.80, 1.02] |
| | Surfactant | 1/4809 | RR 0.83 [0.72, 0.96] |
| | Oxygen therapy | 1/4809 | RR 0.88 [0.81, 0.96] |
| | Chronic lung disease | 1/4809 | RR 0.91 [0.70, 1.17] |
| | Necrotising enterocolitis: | | |
| | Studies using amoxicillin-clavulanate | 2/1880 | RR 4.72 [1.57, 14.23] |
| | Studies using a macrolide | 3/2076 | RR 0.88 [0.45, 1.69] |
| Progesterone by any route in women with previous history of spontaneous preterm birth [22] | Birth before 34 weeks gestation | 5/602 | RR 0.31 [0.14, 0.69] |
| | Perinatal mortality | 6/1453 | RR 0.50 [0.33, 0.75] |
| | Respiratory distress syndrome | 3/1217 | RR 0.45 [0.17, 1.16] |
| | Intubation/mechanical ventilation | 3/633 | RR 0.40 [0.18, 0.90] |

(continued)

**Table 16.3** (continued)

| Intervention | Outcome | Trials/participants | Effect [95 %CI] |
|---|---|---|---|
| Progesterone in women with an ultrasound-identified short cervix [22] | Preterm birth before 34 weeks | 2/438 | RR 0.64 [0.45, 0.90] |
| | Perinatal mortality | 3/1389 | RR 0.74 [0.42, 1.29] |
| | Respiratory distress syndrome | 2/732 | RR 0.49 [0.29, 0.85] |
| | Intubation/mechanical ventilation | 1/274 | RR 0.65 [0.36, 1.16] |
| Progesterone in women with a multiple pregnancy [22] | Birth before 34 weeks gestation | 6/1758 | RR 0.97 [0.74, 1.27] |
| | Perinatal mortality | 7/4136 | RR 0.93 [0.45, 1.94] |
| | Respiratory distress syndrome | 6/5065 | RR 1.13 [0.94, 1.35] |
| | Intubation/mechanical ventilation | 4/3392 | RR 0.95 [0.78, 1.16] |
| Cervical stitch (cerclage) for preventing preterm birth in singleton pregnancy [23] | Perinatal mortality | 8/2391 | RR 0.78 [0.61, 1.00] |
| | Birth before 37 weeks gestation | 9/2898 | RR 0.80 [0.69, 0.95] |
| | Serious respiratory morbidity | 5/839 | RR 1.11 [0.66, 1.88] |
| | Caesarean section | 8/2817 | RR 1.19 [1.01, 1.40] |
| | Maternal pyrexia | 3/1245 | RR 2.39 [1.35, 4.23] |
| Cerclage versus progesterone for preventing preterm birth in singleton pregnancy [23] | Perinatal mortality | 1/79 | RR 1.12 [0.58, 2.16] |
| | Birth before 37 weeks gestation | 1/79 | RR 0.88 [0.60, 1.30] |
| Cerclage for preventing preterm birth in multiple pregnancy [24] | Perinatal mortality | 5/262 | RR 1.74 [0.92, 3.28] |
| | Preterm birth less than 34 weeks | 4/83 | 1.16 [0.44, 3.06] |
| | Respiratory distress syndrome | 3/116 | 1.70 [0.15, 18.77] |

| | | | |
|---|---|---|---|
| Antenatal corticosteroid [25] | Perinatal mortality | 13/3627 | RR 0.77 [0.67, 0.89] |
| | Respiratory distress syndrome | 21/4038 | RR 0.66 [0.59, 0.73] |
| | Respiratory support | 4/569 | RR 0.69 [0.53, 0.90] |
| | Days of mechanical ventilation | 2/198 | MD −3.5 [−5.1, −1.9] |
| | Chronic lung disease | 6/818 | RR 0.86 [0.61, 1.22] |
| Repeat antenatal corticosteroid [26] | Perinatal mortality | 9/5554 | RR 0.94 [0.71, 1.23] |
| | Respiratory distress syndrome | 8/3206 | RR 0.83 [0.75, 0.91] |
| | Intubation/mechanical ventilation | 6/4918 | RR 0.84 [0.71, 0.99] |
| | Days of respiratory support | 1/37 | MD 0.30 [−0.90, 1.50] |
| | Chronic lung disease | 8/5393 | RR 1.06 [0.87, 1.30] |
| | Birth weight (g) | 9/5626 | MD −76 [−118, −34] |
| Induction of labour for improving birth outcomes for women at or beyond term [27] | Perinatal mortality | 17/7407 | RR 0.31 [0.12, 0.81] |
| | Mortality | 17/7407 | RR 0.37 [0.10, 1.38] |
| | Meconium aspiration syndrome | 8/2371 | RR 0.50 [0.34, 0.73] |

*CI* confidence interval, *MD* mean difference, *RR* relative risk

outcomes included reduced need for oxygen therapy and surfactant, although there was no significant effect on mechanical ventilation (2 trials, 4924 infants; RR 0.90, 95 % CI 0.80, 1.02). Use of co-amoxicillin-clavulanate increased the incidence of necrotising enterocolitis and so should be avoided. Use of erythromycin increases latency to delivery and improves some respiratory outcomes in infants, although evidence is insufficient for an effect on mechanical ventilation. Antibiotics can be recommended for women with PPROM to reduce maternal and infant morbidity.

### 16.3.4 Asymptomatic Women at Risk of Preterm Birth

Women with a previous history of spontaneous preterm birth or an ultrasound-identified short cervix or a multiple pregnancy are at increased risk of recurrent preterm birth. Potential approaches to reduce this risk include expectant management, use of progesterone or use of cervical cerclage.

### 16.3.5 Progesterone

In women with a previous preterm birth, progesterone prevented birth before 34 weeks gestation and reduced perinatal mortality and incidence of mechanical ventilation (3 trials, 633 infants; RR 0.40, 95 % CI 0.18, 0.90) [22]. For women with an ultrasound-identified short cervix, progesterone prevented birth before 34 weeks gestation, had no significant effect on perinatal mortality or mechanical ventilation (1 trial, 274 infants; 0.65 RR 0.36, 1.16) but reduced the incidence of respiratory distress syndrome (2 trials, 732 infants; RR 0.49, 95 % CI 0.29, 0.85) [22]. In women with a multiple pregnancy, there is no evidence of an effect of progesterone on preterm birth, perinatal mortality or mechanical ventilation (4 trials, 3392 infants; RR 0.95, 95 % CI 0.78, 1.16).

### 16.3.6 Cervical Cerclage

Systematic review of 12 trials involving 3328 pregnant women at high risk of pregnancy loss (due to previous preterm birth or short cervix on ultrasound) found cerclage compared with no treatment (9 trials) reduced preterm delivery but resulted in no clear difference in perinatal mortality or neonatal illness including respiratory distress syndrome [23]. More women needed caesarean section and experienced side effects including pyrexia. One small trial of cervical cerclage versus progesterone reported no significant difference in obstetric outcomes for the mother and newborn. A second systematic review examined the effect of cervical cerclage in women with a multiple pregnancy and found no evidence of benefit in perinatal death, preterm birth rates or neonatal ill health [24].

### 16.3.7 Summary

In women with a singleton pregnancy with a history of previous preterm birth or short cervix on ultrasound, use of progesterone reduces preterm birth but had no effect on incidence of respiratory distress syndrome [23]. The benefits were not seen in women with a multiple pregnancy [24]. Although cervical cerclage reduces the incidence of preterm birth, there is no evidence for an effect on perinatal mortality or neonatal morbidity.

## 16.4 Antenatal Preparation for Delivery: Maturation of the Foetus (Table 16.3)

### 16.4.1 Antenatal Corticosteroids

In women at risk of preterm birth (21 trials including 3885 women and 4269 infants), antenatal corticosteroids reduce perinatal mortality, respiratory distress syndrome, need for respiratory support (4 trials, 569 infants; RR 0.69, 95 % CI 0.53, 0.90) and duration of mechanical ventilation [25]. The Australian and New Zealand Clinical Practice Guidelines 2015 [28] recommend use of a single course of antenatal corticosteroid for women at imminent risk of preterm birth who are less than 32 weeks and 6 days gestation, regardless of the reason why the women are considered to be at risk of preterm birth. The optimal timing is considered to be when the woman's preterm birth is expected within the next 48 h.

### 16.4.2 Repeat Doses of Antenatal Corticosteroids

Use of a repeat dose of antenatal corticosteroid reduced the incidence of respiratory distress syndrome and mechanical ventilation (6 trials, 4918 infants; RR 0.84, 95 % CI 0.71, 0.99) [26]. Treatment with multiple repeat doses of corticosteroid was associated with a reduction in mean birth weight. The Australian and New Zealand Clinical Practice Guidelines 2015 [28] recommend use of a repeat dose of antenatal corticosteroid in women at risk of early preterm birth when gestational age is 32 weeks and 6 days or less, and preterm birth is planned or expected within the next 7 days, even if birth is likely within 24 h. The repeat dose should not be less than 7 days following a single course of antenatal corticosteroids and should be given regardless of the reason the woman is considered at risk of preterm birth. It is recommended that up to a maximum of three, single, repeat doses only be used in view of growth concerns.

## 16.5 Preventing Respiratory Morbidity at Term Gestation (Table 16.3)

A high proportion of pregnancies are delivered electively (induction of labour or caesarean section) near or at term for maternal or foetal indications. Elective caesarean section at near-term gestation (37–38 weeks) carries a moderate risk

of respiratory morbidity in the infant [4, 9, 14], substantially higher risk than intended vaginal delivery [9]. However, there is insufficient evidence from trials to recommend routine use of antenatal corticosteroids at or near term [28–30]. Delaying delivery until 39 weeks gestation is associated with the lowest incidence of respiratory morbidity [4, 9, 14] and combined perinatal and infant mortality [31].

### 16.5.1 Induction of Labour for Improving Birth Outcomes for Women at or beyond Term

Induction of labour at or beyond term reduces perinatal mortality (17 trials, 7407 infants; RR 0.31 95 % CI 0.12, 0.81) and meconium aspiration syndrome (8 trials, 2371 infants; RR 0.50, 95 % CI 0.34, 0.73) [27]. There were also fewer caesarean sections in the induction group compared with the expectant management group. The majority of trials adopted a policy of induction at 41 completed weeks (287 days) or more.

## 16.6 Postnatal Respiratory Support Strategies for Avoiding Mechanical Ventilation in Preterm Infants (Table 16.4)

From a historical perspective [trials 1967–1970], before routine use of antenatal steroids and postnatal surfactant, mechanical ventilation significantly reduced neonatal mortality (5 trials, 359 infants; RR 0.86 95 % CI 0.74, 1.00) [42].

### 16.6.1 Continuous Positive (or Negative) Airway Pressure Support

Subsequently [trials 1973–2007], application of continuous distending pressure (either positive or negative) for respiratory distress in preterm infants was found to reduce need for mechanical ventilation (5 trials, 314 infants; RR 0.72, 95 % CI 0.56, 0.91) and reduce mortality, but increased pneumothorax compared to use of intermittent positive pressure ventilation [33]. In addition, earlier initiation of continuous distending pressure for respiratory distress syndrome in preterm infants further reduced rates of mechanical ventilation (6 trials, 165 infants; RR 0.55, 95 % CI 0.32, 0.96) with no effect on mortality or chronic lung disease [34].

In the surfactant era prior to routine use of continuous positive airway pressure (CPAP) for infants with respiratory distress, routine intubation for prophylactic surfactant reduced mortality compared to later intubation and rescue surfactant [43]. Of note is the relatively low rate of use of antenatal corticosteroids in the trials.

In the most recent review update [43], routine intubation for prophylactic surfactant versus later intubation and selective use of surfactant in infants on CPAP there was no difference in mortality or chronic lung disease, but a significant increase in

**Table 16.4** Postnatal respiratory support and surfactant strategies for avoiding mechanical ventilation

| Intervention | Outcome | Trials/participants | Effect [95 %CI] |
|---|---|---|---|
| Mechanical ventilation for newborn infants with respiratory failure due to pulmonary disease [32] [trials 1967–1970] | Mortality | 5/359 | RR 0.86 [0.74, 1.00] |
| | Pneumothorax/air leak | 2/275 | RR 2.75 [0.72, 10.45] |
| Continuous distending pressure versus intermittent positive pressure ventilation for respiratory distress in preterm infants [33] [trials 1973–2007] | Mortality | 6/355 | RR 0.52 [0.32, 0.87] |
| | Intubation/mechanical ventilation | 5/314 | RR 0.72 [0.56, 0.91] |
| | Chronic lung disease (28 days) | 3/260 | RR 1.22 [0.44, 3.39] |
| | Pneumothorax/air leak | 6/355 | RR 2.64 [1.39, 5.04] |
| Early versus delayed initiation of continuous distending pressure for respiratory distress syndrome in preterm infants [34] [trials 1977–1981] | Mortality | 6/165 | RR 0.68 [0.34, 1.38] |
| | Intubation/mechanical ventilation | 6/165 | RR 0.55 [0.32, 0.96] |
| | Chronic lung disease (36 weeks) | 1/36 | RR 1.12 [0.08, 16.52] |
| | Pneumothorax/air leak | 5/144 | RR 0.84 [0.37, 1.91] |
| Nasal intermittent positive pressure ventilation versus nasal continuous positive airway pressure in preterm and term infants to prevent intubation [35] | Intubation/mechanical ventilation | 6/620 | RR 0.44 [0.31, 0.63] |
| Laryngeal mask airway surfactant in preterm infants with respiratory distress syndrome [36] | Intubation/mechanical ventilation | 1/26 | RR 1.00 [0.25, 4.07] |
| Rescue thin catheter surfactant versus delayed intubation with selective surfactant in preterm infants on continuous positive airway pressure [37, 38] | Mortality | 2/431 | 1.12 [0.58, 2.18] |
| | Chronic lung disease (36 weeks) | 2/399 | 0.72 [0.49, 1.07] |
| | Mortality or chronic lung disease | 2/431 | 0.83 [0.60, 1.13] |
| | Intubation/mechanical ventilation | 2/431 | 0.71 [0.62, 0.81] |
| Rescue thin catheter surfactant versus rescue InSure technique in preterm infants on continuous positive airway pressure [39–41] | Mortality | 3/318 | RR 1.23 [0.63, 2.42] |
| | Chronic lung disease (36 weeks) | 3/281 | RR 0.53 [0.25, 1.13] |
| | Intubation/mechanical ventilation | 3/318 | RR 0.82 [0.60, 1.12] |
| | Pneumothorax/air leak | 3/318 | RR 0.70 [0.28, 1.77] |

*CI* confidence interval, *MD* mean difference, *RR* relative risk

combined mortality and chronic lung disease for infants treated with intubation and prophylactic surfactant. However, this review did not include all data from trials [44, 45] comparing the practice of prophylactic intubation and surfactant versus CPAP with intubation for late rescue surfactant, potentially due to difficulties around definitions relating to care practices. Unpublished meta-analysis of all data from current trials [44–46] found no significant difference in mortality (3 trials; 2172 infants; RR 0.80, 95 % CI 0.64, 1.01), chronic lung disease (3 trials, 1905 infants; RR 0.97, 95 % CI 0.86, 1.10) or combined mortality and chronic lung disease (3 trials; 2172 infants; RR 0.94, 95 % CI 0.85, 1.04) suggesting further research is required to determine optimal surfactant practices, especially in extremely premature infants.

### 16.6.2 Nasal Intermittent Positive Pressure Ventilation

A non-Cochrane systematic review [35] of nasal intermittent positive pressure ventilation (NIPPV) compared to CPAP found that NIPPV was effective at preventing need for mechanical ventilation (6 trials, 620; RR 0.44, 95 % CI 0.31, 0.63). For other neonatal morbidities, outcomes were reported for trials using NIPPV as a primary mode of ventilation combined with trials assessing infants being extubated. Morbidities were not significantly different.

## 16.7 Non-invasive Surfactant Strategies in Preterm Infants for Preventing Mechanical Ventilation (Table 16.4)

There are several potential methods for non-invasive surfactant delivery including intra-amniotic surfactant prior to delivery (no randomised trials found) [53], intrapartum pharyngeal surfactant administration (no randomised trials found) [54], nebulised surfactant administration [55], laryngeal mask airway surfactant administration [36] and thin catheter-delivered surfactant.

### 16.7.1 Nebulised Surfactant

Systematic review found insufficient evidence to assess the efficacy of nebulised surfactant administration for prevention of mechanical ventilation (1 trial, 32 infants; RR 1.2, 95 % CI 0.46, 3.15) [55].

### 16.7.2 Laryngeal Mask Airway Surfactant Administration

Systematic review found insufficient evidence to assess the efficacy of laryngeal mask airway surfactant administration for prevention of mechanical ventilation (1 trial, 26 infants; RR 1.00, 95 % CI 0.25, 4.07) [36].

### 16.7.3 Thin Catheter-Delivered Surfactant

There is no current published up-to-date systematic review of thin catheter-delivered surfactant for prevention of mechanical ventilation in preterm infants. A literature search (June 2015) identified 5 trials [37–41] including 749 infants assessing the effect of thin catheter-delivered surfactant in infants with respiratory distress. No trial has assessed the effect of prophylactic thin catheter surfactant.

Two trials compared use of rescue thin catheter-delivered surfactant to later intubation and rescue surfactant in extremely [38] and very preterm infants [37] on CPAP with respiratory distress. Meta-analysis found rescue thin catheter surfactant reduced need for mechanical ventilation (2 trials, 431 infants; RR 0.71, 95 % CI 0.62, 0.81) with no significant effect on mortality, chronic lung disease or mortality and combined chronic lung disease.

Three trials [39–41] compared use of rescue thin catheter-delivered surfactant to intubation, rescue surfactant and immediate extubation ('InSurE' technique) in preterm infants with respiratory distress on CPAP. Meta-analysis found that rescue thin catheter surfactant reduced need for IPPV (bag and mask ventilation) and severe saturations during the procedure (3 trials, 318 infants; RR 0.11, 95 % CI 0.06, 0.18), with no significant effect on mechanical ventilation, mortality, chronic lung disease or mortality and combined chronic lung disease compared to the InSurE technique.

## 16.8 Intubation and Surfactant Strategies for Minimising Mechanical Ventilation (Table 16.5)

### 16.8.1 Premedication and Techniques of Intubation

There are no current systematic reviews addressing the use of premedication for endotracheal intubation (a Cochrane review is pending) or use of a videolarygoscope for intubation (awaiting publication of recent trials). Several other approaches to surfactant administration have emerged: endotracheal intubation, surfactant administration and extubation (InSurE technique); endotracheal intubation, surfactant administration and brief mechanical ventilation (aim to extubate within 1 h); and several non-invasive methods of surfactant delivery. Older trials and reviews of premedication may not be applicable to current practices particularly in extremely premature infants.

A systematic review found not enough evidence to demonstrate any differences in the effect of nasal versus oral intubation for mechanical ventilation of newborn babies in neonatal intensive care [56].

### 16.8.2 Early Versus Delayed Surfactant

Early surfactant treatment for neonatal respiratory distress syndrome reduced mortality, rates of chronic lung disease and pneumothorax compared to delayed selective surfactant treatment [47].

**Table 16.5** Surfactant strategies for preventing or minimising mechanical ventilation

| Intervention | Outcome | Trials/participants | Effect [95 %CI] |
|---|---|---|---|
| Early versus delayed selective surfactant treatment for neonatal respiratory distress syndrome [47] [trials 1992–2002] | Mortality | 6/3577 | RR 0.84 [0.74, 0.95] |
| | Chronic lung disease | 3/3050 | RR 0.69 [0.55, 0.87] |
| | Pneumothorax/air leak | 5/3545 | RR 0.69 [0.59, 0.82] |
| Prophylactic versus selective use of surfactant in preterm infants [43].<br>Studies without routine application of continuous positive airway pressure [trials 1991–1997] | Mortality | 8/2761 | RR 0.69 [0.56, 0.85] |
| | Chronic lung disease (36 weeks) | 9/2789 | RR 1.30 [0.77, 2.17] |
| | Pneumothorax/air leak | 8/2760 | RR 0.79 [0.63, 0.98] |
| Prophylactic versus selective use of surfactant in preterm infants [43].<br>Studies with routine application of continuous positive airway pressure [trials 2010–2011] | Mortality | 2/1746 | RR 1.24 [0.97, 1.58] |
| | Chronic lung disease (36 weeks) | 1/402 | RR 1.12 [0.99, 1.26] |
| | Mortality or chronic lung disease | 2/1744 | RR 1.12 [1.02, 1.24] |
| | Pneumothorax/air leak | 1/1316 | RR 1.08 [0.73, 1.60] |
| Early surfactant administration with brief ventilation versus selective surfactant and continued mechanical ventilation for preterm infants with or at risk for respiratory distress syndrome [48] [1994–2005] | Mortality | 6/396 | RR 0.52 [0.17, 1.56] |
| | Chronic lung disease (28 days) | 4/262 | RR 0.51 [0.26, 0.99] |
| | Intubation/mechanical ventilation | 6/664 | RR 0.67 [0.57, 0.79] |
| | Days of mechanical ventilation | 3/278 | MD –0.36 [–0.81, 0.10] |
| | Pneumothorax/air leak | 6/664 | RR 0.52 [0.28, 0.96] |
| Pre- versus post-ventilatory surfactant in preterm infants [49] | Mortality | 1/651 | RR 1.20 [0.90, 1.60] |
| | Chronic lung disease | 1/651 | RR 1.35 [0.93, 1.94] |
| | Pneumothorax/air leak | 1/651 | RR 1.02 [0.60, 1.71] |
| | Chronic lung disease (36 weeks) | 1/505 | RR 1.42 [0.99, 2.02] |
| Multiple versus single doses of exogenous surfactant for the prevention or treatment of neonatal respiratory distress syndrome [50] [trials 1990–1995] | Mortality | 3/1220 | RR 0.59 [0.44, 0.78] |
| | Chronic lung disease | 3/1221 | RR 1.13 [0.83, 1.54] |
| | Pneumothorax/air leak | 3/1220 | RR 0.70 [0.52, 0.94] |

| | | | |
|---|---|---|---|
| Surfactant for meconium aspiration syndrome in full-term/near-term infants [51] | Mortality | 4/326 | RR 0.98 [0.41, 2.39] |
| | Treatment with ECMO | 2/208 | RR 0.64 [0.46, 0.91] |
| | Chronic lung disease | 1/168 | RR 0.47 [0.12, 1.80] |
| | Pneumothorax/air leak | 3/269 | RR 0.82 [0.39, 1.73] |
| | Days of mechanical ventilation | 3/158 | RR 0.60 [−0.41, 1.62] |
| Lung lavage for meconium aspiration syndrome in newborn infants [52] | Mortality | 2/88 | RR 0.42 [0.12, 1.46] |
| | Treatment with ECMO | 2/47 | RR 0.27 [0.04, 1.86] |
| | Mortality or use of ECMO | 2/88 | RR 0.33 [0.11, 0.96] |
| | Pneumothorax | 2/88 | RR 0.38 [0.08, 1.90] |

*CI* confidence interval, *MD* mean difference, *RR* relative risk

In preterm infants with respiratory distress syndrome, early surfactant administration with brief ventilation (aim to extubate within 1 h) reduced need for mechanical ventilation (6 trials, 664 infants; RR 0.67 95 % CI 0.57, 0.79), pneumothorax and chronic lung disease at 28 days compared to selective surfactant and continued mechanical ventilation [48].

### 16.8.3 Prophylactic Versus Delayed Selective Surfactant

For trials without routine use of CPAP in the control arm, meta-analysis found a reduction in mortality (8 trials, 2761 infants; RR 0.69, 95 % CI 0.56, 0.85) from routine intubation for prophylactic surfactant compared to delayed intubation and selective surfactant administration [43]. In contrast, meta-analysis of trials with routine use of CPAP found no significant difference in mortality or chronic lung disease, and a significant increase in combined mortality and chronic lung disease for infants treated with intubation and prophylactic surfactant. The optimal respiratory support and surfactant strategy, especially in extremely premature infants, is currently unclear. There are now trials assessing other strategies of non-invasive surfactant administration for infants on CPAP.

An overview of systematic reviews of prophylactic and early surfactant provides evidence for the earliest possible administration of surfactant especially in extremely premature infants [43, 48]. However, the optimal strategy is unclear particularly in infants with adequate antenatal corticosteroid cover managed with CPAP. There is increasing evidence becoming available regarding the relative safety of less invasive methods of surfactant administration (see above).

### 16.8.4 Pre- Versus Post-ventilatory Surfactant in Preterm Infants

A single randomised controlled trial including 651 infants reported no significant difference in mortality, chronic lung disease, pneumothorax or air leak or chronic lung disease for infants given endotracheal surfactant prior to initiation of positive pressure ventilation [49]. There is no proven benefit from endotracheal surfactant prior to resuscitation of the newborn infant.

### 16.8.5 Multiple versus Single Doses of Exogenous Surfactant in Preterm Infants

Repeated dosing of surfactant in ventilated preterm infants improved oxygenation and ventilatory support parameters and reduced mortality (3 trials, 1220 infants; RR 0.59, 95 % CI 0.44, 0.78) and rates of pneumothorax.

## 16.9 Surfactant in Full-Term/Near-Term Infants (Table 16.5)

### 16.9.1 Surfactant for Meconium Aspiration Syndrome

Surfactant treatment of ventilated infants with meconium aspiration syndrome did not affect mortality, pneumothorax, air leak or duration of mechanical ventilation, but reduced requirement for ECMO (2 trials, 208 infants; RR 0.64, 95 % CI 0.46, 0.91) [51]. The trials that reported benefit in respiratory parameters and prevention of ECMO gave 100–150 mg/kg surfactant every 6 h for four doses, with infants treated earlier having the greatest response [57, 58].

Dilute surfactant lung lavage for meconium aspiration syndrome did not affect mortality, treatment with ECMO, or pneumothorax but reduced combined mortality and treatment with ECMO (2 trials, 88 infants; RR 0.33, 95 % CI 0.11, 0.96) [52]. There are no trials reporting the comparison of surfactant treatment compared to dilute surfactant lung lavage in infants with meconium aspiration.

## 16.10 Minimising Mechanical Ventilation and Its Side Effects for Ventilated Infants (Table 16.6)

### 16.10.1 Analgesia, Sedation and Neuromuscular Blockade of Ventilated Newborn Infants

#### 16.10.1.1 Midazolam

Systematic review identified 3 trials including 122 infants and found higher sedation scores in the midazolam group compared to the placebo group, the importance of which is unclear [59]. One study reported a higher incidence of adverse neurological events, and meta-analysis (2 trials, 89 infants) found no difference in duration of mechanical ventilation but longer duration of NICU stay in the midazolam group compared to the placebo group.

#### 16.10.1.2 Morphine

Systematic review found 13 trials including 1505 infants assessing the effect of opioids in ventilated newborn infants [60]. Infants given opioids had reduced premature infant pain profile scores. Meta-analysis found no difference in duration of mechanical ventilation or hospital stay and no significant difference in mortality or neonatal morbidity.

#### 16.10.1.3 Neuromuscular Blockade

In the subgroup of trials that enrolled preterm infants with asynchronous ventilation, neuromuscular blockade (with pancuronium) was associated with reduced severe intraventricular haemorrhage and pneumothorax with no effect on mortality [61]. In the subgroup of trials not selecting for asynchrony at entry, there was no

**Table 16.6** Minimising mechanical ventilation and its side effects for ventilated infants

| Intervention | Outcome | Trials/participants | Effect [95%CI] |
|---|---|---|---|
| Premedication for endotracheal intubation in mechanically ventilated neonates | No current systematic review | | |
| Videolaryngoscopy for intubation of newborn infants | No current systematic review | | |
| Nasal versus oral intubation for mechanical ventilation of newborn infants [56] | Extubation failure | 2/160 | RR 1.64 [0.75, 3.59] |
| Intravenous midazolam infusion for sedation of infants in the neonatal intensive care unit [59] | Days of mechanical ventilation | 2/89 | MD 3.60 [−0.25, 7.44] |
| | Days of NICU stay | 2/89 | MD 5.44 [0.40, 10.49] |
| Opioids for neonates receiving mechanical ventilation [60] | Days of mechanical ventilation | 6/1197 | MD 1.24 [−0.29, 2.77] |
| | Days of NICU stay | 3/129 | MD 1.80 [−7.03, 10.62] |
| Neuromuscular blockade for newborn infants receiving mechanical ventilation [61] (trials selecting for asynchrony at entry) | Mortality | 3/103 | RR 0.95 [0.49, 1.84] |
| | Severe intraventricular haemorrhage | 3/103 | RR 0.55 [0.34, 0.89] |
| | Pneumothorax | 2/46 | RR 0.29 [0.11, 0.77] |
| | Days of ventilation reported as medians | | |
| Neuromuscular paralysis for newborn infants receiving mechanical ventilation [61] (trials not selecting for asynchrony at entry) | Mortality | 2/161 | RR 1.41 [0.94, 2.10] |
| | Severe intraventricular haemorrhage | 1/193 | RR 0.91 [0.39, 2.14] |
| | Pneumothorax | 3/354 | RR 1.06 [0.71, 1.60] |
| | Chronic lung disease (36 weeks) | 1/154 | RR 1.23 [0.80, 1.88] |
| | Days of ventilation reported as medians | | |

| | | | |
|---|---|---|---|
| Long versus short inspiratory times in neonates receiving mechanical ventilation [62] | Mortality | 5/694 | RR 1.26 [1.00, 1.59] |
| | Chronic lung disease | 4/654 | RR 0.91 [0.66, 1.24] |
| | Pneumothorax/air leak | 5/685 | RR 1.56 [1.25, 1.94] |
| High-frequency positive pressure ventilation (rate >= 60 breaths per minute)versus convention ventilation [63] | Death | 3/585 | RR 0.80 [0.62, 1.03] |
| | Pneumothorax | 3/585 | RR 0.69 [0.51, 0.93] |
| | Chronic lung disease (28 days) | 3/585 | RR 1.09 [0.78, 1.51] |
| Assist-control ventilation or synchronous intermittent mandatory ventilation versus conventional ventilation [63] | Mortality | 5/1729 | RR 1.19 [0.95, 1.49] |
| | Chronic lung disease | 2/1310 | RR 0.90 [0.75, 1.08] |
| | Pneumothorax/air leak | 6/1769 | RR 1.03 [0.80, 1.34] |
| | Extubation failure | 4/1056 | RR 0.93 [0.68, 1.28] |
| | Hours of mechanical ventilation | 4/1402 | MD −34.78 [−62.11, −7.44] |
| Volume-targeted ventilation versus pressure-limited ventilation in the neonate [64] | Mortality in hospital | 7/554 | RR 0.80 [0.53, 1.20] |
| | Mortality or chronic lung disease (36 weeks) | 5/439 | RR 0.73 [0.57, 0.93] |
| | Chronic lung disease (36 weeks) | 5/413 | 0.73 [0.53, 1.00] |
| | Pneumothorax/air leak | 8/589 | RR 0.46 [0.25, 0.84] |
| | Days of mechanical ventilation | 6/431 | MD −2.36 [−3.90, −0.83] |
| Elective high-frequency oscillatory ventilation versus conventional ventilation in preterm infants [65] | Mortality | 15/2885 | RR 0.98 [0.83, 1.14] |
| | Chronic lung disease (36 weeks) | 15/2369 | RR 0.89 [0.81, 0.99] |
| | Mortality or chronic lung disease | 15/2885 | RR 0.90 [0.78, 1.03] |
| | Pneumothorax/air leak | 12/2766 | RR 1.19 [1.05, 1.34] |

(continued)

**Table 16.6** (continued)

| Intervention | Outcome | Trials/participants | Effect [95%CI] |
|---|---|---|---|
| High-frequency oscillatory ventilation versus conventional ventilation for infants with severe pulmonary dysfunction born at or near term [66] | Mortality: elective HFOV | 1/118 | RR 5.93 [0.71, 49.19] |
| | Rescue HFOV | 1/79 | RR 0.51 [0.05, 5.43] |
| | Failed therapy*: elective HFOV | 1/118 | RR 10.64 [0.59, 193.23] |
| | Rescue HFOV | 1/79 | RR 0.73 [0.47, 1.13] |
| | Received ECMO: elective HFOV | 1/79 | RR 2.05 [0.86, 4.92] |
| | Air leak: elective HFOV | 1/118 | RR 2.37 [0.22, 25.43] |
| | Rescue HFOV | 1/79 | RR 0.68 [0.21, 2.24] |
| | Chronic lung disease: elective HFOV | 1/112 | RR 1.14 [0.75, 1.73] |
| | Rescue HFOV | 1/79 | RR 2.26 [0.86, 5.90] |
| | Days of mechanical ventilation: elective HFOV | 1/112 | MD 0.70 [−0.97, 2.37] |
| Inhaled nitric oxide for respiratory failure in preterm infants [67] | Mortality: infants treated <3 days | 9/1006 | RR 1.04 [0.90, 1.20] |
| | Infants treated >3 days | 2/624 | RR 1.06 [0.64, 1.74] |
| | Routine use | 3/1800 | RR 0.91 [0.74, 1.11] |
| | Chronic lung disease: infants treated <3 days | 8/681 | RR 0.89 [0.76, 1.04] |
| | Infants treated >3 days | 2/573 | RR 0.89 [0.78, 1.02] |
| | Routine use | 3/1658 | RR 0.94 [0.84, 1.05] |
| | Mortality or CLD: infants treated <3 days | 8/958 | RR 0.94 [0.87, 1.01] |
| | Infants treated >3 days | 2/624 | RR 0.90 [0.80, 1.02] |
| | Routine use | 3/1800 | RR 0.93 [0.86, 1.01] |

| | | | |
|---|---|---|---|
| Nitric oxide for respiratory failure in infants born at or near term [68] | Mortality | 9/916 | RR 0.91 [0.60, 1.37] |
| | Requirement for ECMO | 8/810 | RR 0.63 [0.54, 0.75] |
| | Mortality or requirement for ECMO | 9/915 | RR 0.68 [0.59, 0.79] |
| Sildenafil for pulmonary hypertension in neonates [69] | Mortality | 3/77 | RD −0.38 [−0.60, −0.16] |

*CI* confidence interval, *MD* mean difference, *RR* relative risk
* Numbers too small for accurate estimate

difference in mortality or morbidity. Three trials reported no significant difference in median duration of mechanical ventilation for infants receiving neuromuscular blockade. The data were not included in meta-analysis.

## 16.11 Minimising Side Effects of Conventional Ventilation (Table 16.6)

### 16.11.1 Inspiratory Time and Rate

Use of longer inspiratory time (IT) was associated with increased mortality (5 trials, 694 infants; RR 1.26, 96 % CI 1.00, 1.59) and pneumothorax ventilation compared to shorter IT ventilation [62]. Subgroup analysis found benefit from use of short inspiratory times <0.5 s (reduced pneumothorax) and in infants with hyaline membrane disease (reduced mortality and pneumothorax).

A second systematic review found that high-frequency positive pressure ventilation (rates >60 breaths per minute) was associated with reduced incidence of air leaks including pneumothorax (3 trials, 585 infants; RR 0.69, 95 % CI 0.51, 0.93) compared to lower-rate ventilation (30–40 bpm) [63].

### 16.11.2 Synchronised Ventilation

Assist-control ventilation or synchronous intermittent mandatory ventilation (both forms of triggered ventilation) reduced duration of mechanical ventilation (4 trials, 1402 infants; MD −34.78 h, 95 % CI −62.11, −7.44) for infants managed on synchronised ventilation compared to untriggered conventional ventilation [63].

### 16.11.3 Volume-Targeted Ventilation versus Pressure-Limited Ventilation

Volume-targeted ventilation reduced mortality or chronic lung disease (5 trials, 439 infants; RR 0.73, 95 % CI 0.57, 0.93), pneumothorax/air leak (8 trials, 589 infants; RR 0.46, 95 % CI 0.25, 0.84) and duration of mechanical ventilation (6 trials, 431 infants; MD −2.36 days, 95 % CI −3.90, −0.83) compared to pressure-limited ventilation [64].

## 16.12 Minimising Side Effects of High-Frequency Oscillation Ventilation (HFOV) (Table 16.6)

### 16.12.1 Preterm Infants

Systematic review comparing HFOV with conventional ventilation revealed no evidence of effect on mortality. A reduction in chronic lung disease (15 trials, 2369 infants; RR 0.89, 95 % CI 0.81, 0.99) in survivors at term equivalent

gestational age was inconsistent across studies, and the reduction was of borderline significance [65]. The effect was similar in trials with a high lung volume strategy for HFOV targeting low inspired oxygen concentration ($FiO_2$) and trials with a high lung volume strategy with somewhat higher or unspecified target $FiO_2$. Subgroups of trials showed a significant reduction in chronic lung disease with HFOV when no surfactant was used, when piston oscillators were used for HFOV, when lung protective strategies for conventional ventilation were not used, when randomisation occurred at 2–6 h of age and when inspiratory/expiratory ratio of 1:2 was used for HFOV. In the meta-analysis of all trials, pulmonary air leaks (12 trials, 2766 infants; RR 1.19, 95 % CI 1.05, 1.34) occurred more frequently in the HFOV group. Duration of mechanical ventilation was not significantly different.

### 16.12.2 Term Infants

The systematic review of elective or rescue HFOV near or at term is underpowered to detect important differences in outcomes compared to conventional ventilation. A single trial [70] (118 infants) of elective HFOV versus conventional ventilation reported no significant difference in mortality, failed treatment, chronic lung disease, air leak or days of mechanical ventilation. A single trial [71] (79 infants) of rescue HFOV (candidates for ECMO) reported no significant difference in mortality, failed treatment, use of ECMO, chronic lung disease or air leak.

## 16.13 Pulmonary Vasodilators in Infants with Lung Disease (Table 16.6)

### 16.13.1 Nitric Oxide in Preterm Infants

Systematic review of nitric oxide for respiratory failure in preterm infants found no significant difference in mortality, chronic lung disease or mortality and chronic lung disease irrespective of indication including early rescue treatment (<3 days), routine use in infants with lung disease or later use based on risk of chronic lung disease [67]. Nitric oxide does not appear to improve outcomes of preterm infants with respiratory disease. There are no trials of echocardiographically directed treatment.

### 16.13.2 Nitric Oxide in Term Infants

Nitric oxide for respiratory failure in infants born at or near term did not affect mortality but reduced requirement for ECMO and combined mortality and requirement for ECMO (9 trials, 915 infants; RR 0.68, 95 % CI 0.59, 0.79) [68].

### 16.13.3 Sildenafil in Term Infants with Pulmonary Hypertension

Sildenafil for pulmonary hypertension in newborn infants reduced mortality (3 trials, 77 infants; risk difference −0.38, 95 % CI −0.60, −0.16) [69]. The trials were all conducted in low-resource settings without availability of HFOV or nitric oxide and used oral sildenafil 1–3 mg/kg/dose every 6 h.

## 16.14 Strategies for Facilitating Extubation and Avoiding Reintubation (Table 16.7)

The following strategies have been found to reduce duration of mechanical ventilation or rates of extubation failure/reintubation in preterm infants.

### 16.14.1 Methylxanthines

Use of methylxanthines for extubation in preterm infants: systematic review found reduced extubation failure (6 trials, 197; RR 0.48, 95 % CI 0.32, 0.71), postmenstrual age at last positive pressure ventilation (1 trial, 667; MD −1.10 weeks, 95 % CI −1.64, −0.56), patent ductus arteriosus ligation, chronic lung disease, cerebral palsy and mortality or major disability by 18–21 months age [72]. The largest trial used caffeine citrate 20 mg/kg loading dose and 5–10 mg/kg/day maintenance. Systematic review of caffeine versus theophylline for apnoea in preterm infants found a reduced incidence of adverse effects for caffeine [89].

### 16.14.2 Ventilation Settings Prior to Extubation

Extubation from compared to extubation after a trial of endotracheal CPAP in intubated preterm infants: systematic review found reduced extubation failure (3 trials, 145 infants; RR 0.21, 95 % 0.06, 0.72) in infants extubated to head box oxygen, but not those extubated to CPAP [73].

### 16.14.3 Extubation to Respiratory Support

Use of nasal CPAP immediately after extubation reduces extubation failure (9 trials, 726 infants; RR 0.62, 95 % CI 0.51, 0.76) [74]. Use of nasal intermittent positive pressure ventilation (NIPPV) results in a reduction in respiratory failure post-extubation and endotracheal reintubation (3 trials, 159 infants; RR 0.39, 95 % CI 0.16, 0.97) compared to use of nasal CPAP [75].

**Table 16.7** Strategies for facilitating extubation and avoiding reintubation

| Intervention | Outcome | Trials/participants | Effect [95%CI ] |
|---|---|---|---|
| Prophylactic methylxanthines for extubation in preterm infants [72] | Extubation failure | 6/197 | RR 0.48 [0.32, 0.71] |
| | Postmenstrual age at last positive pressure ventilation | 1/667 | MD −1.10 [−1.64, −0.56] |
| | Patent ductus arteriosus ligation | 1/717 | RR 0.32 [0.20, 0.52] |
| | Chronic lung disease at term age | 1/672 | RR 0.81 [0.70, 0.93] |
| | Cerebral palsy | 1/644 | RR 0.54 [0.32, 0.92] |
| | Mortality or major disability by 18–21 months | 1/676 | RR 0.85 [0.73, 0.99] |
| Extubation from low-rate intermittent positive airway pressure versus extubation after a trial of endotracheal continuous positive airway pressure in intubated preterm infants [73] | Extubation failure: all trials | 3/174 | RR 0.45 [0.19, 1.07] |
| | Extubated to head box | 3/145 | RR 0.21 [0.06, 0.72] |
| | Extubated to continuous positive airway pressure | 1/57 | RR 1.69 [0.55, 5.14] |
| Nasal continuous positive airways pressure immediately after extubation in preterm infants [74] | Extubation failure | 9/726 | RR 0.62 [0.51, 0.76] |
| | Endotracheal reintubation | 9/726 | RR 0.87 [0.69, 1.08] |
| | Oxygen use (28 days) | 5/433 | RR 1.00 [0.81, 1.24] |
| Nasal intermittent positive pressure ventilation versus nasal continuous positive airway pressure for preterm neonates after extubation [75] | Respiratory failure post-extubation | 3/159 | RR 0.21 [0.10, 0.45] |
| | Endotracheal reintubation | 3/159 | RR 0.39 [0.16, 0.97] |
| | Chronic lung disease (36 weeks) | 2/118 | RR 0.73 [0.49, 1.07] |
| Short binasal prong versus single nasal prong continuous positive airway pressure to prevent extubation failure [76] | Mortality | 1/87 | RR 1.68 [0.30, 9.58] |
| | Extubation failure | 2/180 | RR 0.59 [0.41, 0.85] |
| | Days of mechanical ventilation | 1/87 | MD 3.39 [−10.22, 3.44] |
| | Days of respiratory support | 1/87 | MD −6.35 [−16.99, 4.29] |

(continued)

**Table 16.7** (continued)

| Intervention | Outcome | Trials/participants | Effect [95%CI ] |
|---|---|---|---|
| High-flow nasal cannula versus continuous positive airway pressure to prevent extubation failure [77] | Mortality | 3/475 | RR 0.66 [0.24, 1.82] |
| | Treatment failure | 3/475 | RR 1.22 [0.92, 1.61] |
| | Intubation/mechanical ventilation | 1/132 | RR 0.85 [0.45, 1.60] |
| | Chronic lung disease | 2/435 | RR 0.87 [0.68, 1.13] |
| | Nasal trauma | 2/343 | RR 0.73 [0.57, 0.93] |
| | Pneumothorax/air leak | 2/435 | RR 0.27 [0.04, 1.62] |
| Early (<8 days) postnatal corticosteroids for preventing chronic lung disease in preterm infants [78] | Mortality | 27/3720 | RR 1.00 [0.89, 1.13] |
| | Chronic lung disease 36 weeks | 21/3286 | RR 0.79 [0.71, 0.88] |
| | Mortality or chronic lung disease (36 weeks) | 22/3320 | RR 0.89 [0.84, 0.95] |
| | Failure to extubate by the 7th day | 7/956 | RR 0.75 [0.65, 0.86] |
| | Hyperglycaemia | 13/2175 | RR 1.34 [1.21, 1.48] |
| | Hypertension | 11/1996 | RR 1.85 [1.55, 2.22] |
| | Hypertrophic cardiomyopathy | 1/50 | RR 4.33 [1.40, 13.37] |
| | Growth failure | 1/50 | RR 6.67 [2.27, 19.62] |
| | Gastrointestinal bleeding | 12/1820 | RR 1.86 [1.35, 2.55] |
| | Intestinal perforation | 15/2523 | RR 1.81 [1.33, 2.48] |
| | Severe retinopathy of prematurity in survivors | 12/1575 | RR 0.77 [0.64, 0.94] |
| | Cerebral palsy | 12/1452 | RR 1.45 [1.06, 1.98] |
| | Mortality or cerebral palsy | 12/1452 | RR 1.09 [0.95, 1.25] |
| | Mortality or abnormal neurological exam | 5/829 | RR 1.23 [1.06, 1.42] |

| | | | |
|---|---|---|---|
| Late (>7 days) postnatal corticosteroids for chronic lung disease in preterm infants [79] | Mortality at latest reported age | 16/936 | RR 0.87 [0.67, 1.13] |
| | Chronic lung disease (36 weeks) | 8/471 | RR 0.72 [0.61, 0.85] |
| | Mortality or chronic lung disease (36 weeks) | 8/471 | RR 0.72 [0.63, 0.82] |
| | Failure to extubate by day 7 | 10/497 | RR 0.64 [0.56, 0.74] |
| | Hyperglycaemia | 14/1192 | RR 1.52 [1.25, 1.84] |
| | Glycosuria | 2/48 | RR 8.03 [2.43, 26.52] |
| | Hypertension | 12/1096 | RR 2.66 [1.58, 4.49] |
| | Necrotising enterocolitis | 8/952 | RR 0.98 [0.56, 1.73] |
| | Gastrointestinal bleeding | 7/992 | RR 1.38 [0.99, 1.93] |
| | Intestinal perforation | 2/95 | RR 0.36 [0.02, 8.05] |
| | Severe retinopathy of prematurity | 12/558 | RR 1.38 [1.07, 1.79] |
| | Hypertrophic cardiomyopathy | 4/238 | RR 2.76 [1.33, 5.74] |
| | Cerebral palsy | 11/777 | RR 1.14 [0.79, 1.64] |
| Intravenous dexamethasone for extubation of newborn infants [80] | Intubation/mechanical ventilation | 3/160 | RR 0.18 [0.04, 0.97] |
| Corticosteroid therapy for meconium aspiration syndrome [81] | Mortality | 2/85 | RR 0.95 [0.20, 4.58] |
| | Days of mechanical ventilation | 1/50 | MD −1.10 [−2.79, 0.59] |
| | Hours of oxygen therapy | 2/85 | MD 30.00 [8.40, 51.61] |
| | Pneumothorax/air leak | 1/35 | RR 0.64 [0.18, 2.26] |
| Diuretics for respiratory distress syndrome in preterm infants [82] | Mortality | 6/248 | RR 1.35 [0.71, 2.56] |
| | Failure to extubate within 3 days | 1/56 | RR 0.65 [0.45, 0.93] |
| | Chronic lung disease | 3/139 | RR 0.81 [0.41, 1.59] |

(continued)

**Table 16.7** (continued)

| Intervention | Outcome | Trials/participants | Effect [95%CI ] |
|---|---|---|---|
| Diuretics acting on the distal renal tubule for preterm infants with (or developing) chronic lung disease [83] | Not extubated after 8 weeks treatment | 1/33 | RR 0.46 [0.19, 1.11] |
| | Change in % inspiratory $O_2$ after 4 weeks | 1/43 | MD −4.0 [−11.65, 3.65] |
| | Days of oxygen | 1/43 | MD −12.00 [−51.58, 23.58] |
| | Days of hospital stay | 1/43 | MD −3.0 [−32.01, 26.01] |
| Intravenous or enteral loop diuretics for preterm infants with (or developing) chronic lung disease [84] | Change in % inspiratory $O_2$ after 1 week | 2/39 | MD −2.70 [−4.35, −1.05] |
| | Failure to extubate after 1 week | 1/13 | RR 0.53 [0.17, 1.68] |
| Air versus oxygen for resuscitation of infants at birth [85] | Mortality | 4/1275 | RR 0.71 [0.54, 0.94] |
| Lower versus higher oxygen concentration for delivery room stabilisation of preterm neonates [86] | Mortality: All | 6/484 | RR 0.65 [0.43, 0.98] |
| | Trials with allocation concealment | 4/267 | RR 1.00 [0.45, 2.24] |
| | Trials without allocation concealment | 2/207 | RR 0.53 [0.33, 0.88] |
| Early lower versus higher oxygen saturation targeting in preterm infants [87] | Mortality at discharge | 2/2441 | RR 1.16 (0.98, 1.37) |
| | Mortality (36 weeks) | 2/2441 | RR 1.20 (1.00, 1.44) |
| | Chronic lung disease | 2/729 | RR 0.99 (0.85, 1.16) |
| | Oxygen (36 weeks) | 2/2029 | RR 0.90 (0.81, 0.99) |
| | Retinopathy of prematurity | 2/2079 | RR 0.79 (0.63, 1.00) |
| Supplemental oxygen for the treatment of prethreshold retinopathy of prematurity [88] | Mortality | 1/649 | RR 1.29 [0.49, 3.42] |
| | Adverse pulmonary event by 3 months | 1/649 | RR 1.24 [1.06, 1.44] |

*CI* confidence interval, *MD* mean difference, *RR* relative risk

Short binasal prong CPAP reduced extubation failure (2 trials, 180 infants; RR 0.59, 95 % CI 0.41, 0.85) compared to single nasal prong CPAP [76].

Systematic review (3 trials, 475 infants) of high-flow nasal cannula (HFNC) versus CPAP found no significant difference in mortality, treatment failure, reintubation, chronic lung disease, pneumothorax or air leak and a reduction in nasal trauma (2 trials, 343 infants; RR 0.73, 95 % CI 0.57, 0.93) [77]. However, it is likely the trials did not have routine mouth closure for infants treated with nasal CPAP.

### 16.14.4 Corticosteroids in Preterm Infants with Lung Disease

There are several reviews of use of postnatal corticosteroids in preterm infants to treat severe respiratory disease, facilitate extubation and prevent or treat chronic lung disease.

#### 16.14.4.1 Early (<8 Days) Postnatal Corticosteroids for Preventing Chronic Lung Disease in Preterm Infants

Early postnatal corticosteroid treatment (≤7 days), facilitates extubation, reduces the risk of chronic lung disease and patent ductus arteriosus but causes adverse effects including gastrointestinal bleeding, intestinal perforation, hyperglycaemia, hypertension, hypertrophic cardiomyopathy and growth failure. Long-term follow-up studies report an increased risk of abnormal neurological examination and cerebral palsy [78].

#### 16.14.4.2 Late (>7 Days) Postnatal Corticosteroids for Chronic Lung Disease in Preterm Infants

Giving corticosteroids to infants at least 7 days age reduces need for assisted ventilation (failure to extubate by day 7: 10 trials, 497 infants; RR 0.64 95 % CI 0.56, 0.74), chronic lung disease and possibly neonatal mortality. High doses are associated with gastrointestinal bleeding and hypertension. There is little evidence for long-term complications. The review concludes that the use of postnatal corticosteroids should be limited to late use in those infants who cannot be weaned from assisted ventilation and to minimise the dose and duration of any course of treatment [79].

### 16.14.5 Diuretics in Preterm Infants with Lung Disease

Systematic review of use of diuretic for ventilated infants with respiratory distress syndrome found a reduction in failure to extubate within 3 days (1 trial, 56 infants; RR 0.65, 95 % CI 0.45, 0.93) but no difference in other clinical outcomes [82]. Systematic review of intravenous or loop diuretics (frusemide) in preterm infants <3 weeks of age developing chronic lung disease found inconsistent effects or no detectable effect. In infants >3 weeks of age with chronic lung disease, a single intravenous dose of 1 mg/kg of frusemide improved lung compliance and airway resistance for 1 h. Chronic administration of frusemide improves both oxygenation and lung compliance [84]. Systematic review of thiazide diuretics (with or without

spironolactone) found that in preterm infants >3 weeks of age with lung disease, acute and chronic administration of distal diuretics improves pulmonary mechanics, but infants required increased electrolyte supplementation. The review is underpowered to report reliable clinical outcomes [83].

## 16.15 Oxygen Targeting in Newborn Infants (Table 16.6)

### 16.15.1 At Resuscitation for Term Infants

Initial use of air instead of oxygen at resuscitation reduced mortality (4 trials, 1275 infants; RR 0.71, 95 % CI 0.54, 0.94) in term infants receiving resuscitation, although the findings are affected by methodological concerns of included trials [85].

### 16.15.2 At Resuscitation for Preterm Infants

A non-Cochrane systematic review found no difference in mortality (4 trials, 267 infants; RR 1.00, 95 % CI 0.45, 2.24) from resuscitation of preterm infants with lower versus higher oxygen in trials with adequate with allocation concealment [86].

### 16.15.3 Early Saturation Targeting in Preterm Infants

Lower early saturation targeting ($SpO_2$ <90 %) may increase mortality (up to 36 weeks PMA) (2 trials, 2441 infants; RR 1.20, 95 % CI 1.00, 1.44) [87].

### 16.15.4 Late Saturation Targeting in Preterm Infants

Higher late saturation targeting ($SpO_2$ >95 %) increased pulmonary adverse events to 3 months (1 trial, 649 infants; RR 1.24, 95 % CI 1.06, 1.44) [90].

### 16.15.5 Summary

For resuscitation of term infants, the initial use of air is supported by current evidence. For preterm infants, optimal initial oxygen concentrations for resuscitation are unclear. However, an oxygen saturation target of around 90–95 % is supported by evidence from early and late oxygen saturation targeting trials.

**Conclusion**

Preventing mechanical ventilation and minimising its side effects requires an integrated perinatal and neonatal approach to care. Table 16.8 provides a checklist for practices. Antenatal strategies to improve respiratory outcomes include identifying

**Table 16.8** Summary of interventions for preventing mechanical ventilation and reducing its side effects

| Patient population | Intervention | Outcomes affected |
|---|---|---|
| *Antenatal prevention of respiratory disease* | | |
| Women in preterm labour | Calcium channel blockers (tocolytic) [19] | Reduced respiratory distress syndrome, intraventricular haemorrhage, necrotising enterocolitis and jaundice. Reduced maternal side effects compared to betamimetics |
| Women with prelabour premature rupture of membranes | Antibiotics (macrolide) [21] | Reduced need for oxygen therapy and surfactant. Reduced maternal infective morbidity |
| Pregnant women with previous preterm birth<br>Pregnant women with a short cervix on ultrasound | Progesterone [22] | Reduced perinatal mortality and mechanical ventilation |
| All women at risk of preterm birth before 34 weeks plus 6 days gestation [28] | Antenatal corticosteroids [25] | Reduced perinatal mortality, respiratory distress syndrome, need for respiratory support and duration of mechanical ventilation |
| All women at risk of preterm birth 32 weeks and 6 days gestation [28] | Repeat antenatal corticosteroids [26] | Reduced incidence of respiratory distress syndrome and mechanical ventilation |
| Pregnant women beyond term gestation | Induction of labour [27] | Reduced perinatal mortality and meconium aspiration syndrome |
| *Preventing intubation in newborns* | | |
| Preterm infants with respiratory distress | Nasal continuous positive airway pressure (CPAP) [33] | Reduce need for mechanical ventilation and reduce mortality, but increased pneumothorax |
| | Early versus delayed nasal CPAP [34] | Further reduced rates of mechanical ventilation compared to later initiation nasal CPAP |
| | Nasal intermittent positive pressure ventilation [35] | Further reduce rates of mechanical ventilation compared to nasal CPAP |
| Preterm infants on nasal CPAP | Thin catheter surfactant administration [37, 38] | Reduced need for mechanical ventilation |

(continued)

**Table 16.8** (continued)

| Patient population | Intervention | Outcomes affected |
|---|---|---|
| *Minimising and reducing side effects of mechanical ventilation* | | |
| Preterm infants with respiratory distress | Early intubation and surfactant administration [47] | Reduced mortality, rates of chronic lung disease and pneumothorax compared to delayed selective surfactant treatment |
| | Prophylactic surfactant [43] | Reduced mortality compared to delayed selective surfactant in trials without routine use of nasal CPAP |
| | | No difference in mortality and chronic lung disease but possible increase in combined mortality and chronic lung disease compared to delayed selective surfactant in trials with routine use of nasal CPAP (and high rates of antenatal corticosteroids) |
| | Early surfactant administration with brief ventilation [48] | Reduced need for mechanical ventilation, pneumothorax and chronic lung disease compared to selective surfactant and continued mechanical ventilation |
| Ventilated preterm infants | Repeat dosing of surfactant [26] | Improved oxygenation and ventilatory support parameters, reduced mortality and pneumothorax |
| Conventionally ventilated preterm infants | Short inspiratory time (<0.5 s) [62] | Reduced mortality and pneumothorax |
| | Synchronised ventilation [63] | Reduced duration of mechanical ventilation |
| | Volume-targeted ventilation [64] | Reduced mortality or chronic lung disease, pneumothorax and duration mechanical compared to pressure-limited ventilation |
| Ventilated infants with meconium aspiration syndrome | Surfactant [51] | Reduced requirement for ECMO |
| | Surfactant lavage [52] | Reduced combined mortality and treatment with ECMO |
| Ventilated term infants with respiratory failure | Nitric oxide [68] | Reduced requirement for ECMO and combined mortality and requirement for ECMO |
| | Sildenafil [69] | Infants reduced mortality (nitric oxide not available) |

| *Strategies for facilitating extubation and avoiding reintubation* | | |
|---|---|---|
| Ventilated preterm infants | Use of methylxanthines [72] | Reduced extubation failure, postmenstrual age at last positive pressure ventilation, patent ductus arteriosus ligation, chronic lung disease, cerebral palsy and mortality or major disability by 18–21 months age |
| | Low-rate intermittent positive airway pressure ventilation prior to extubation [73] | Reduced extubation failure compared to extubation after a trial of endotracheal CPAP |
| | Immediate nasal CPAP post-extubation [74] | Reduced extubation failure |
| | Nasal intermittent positive pressure ventilation post-extubation [75] | Reduced extubation failure compared to nasal CPAP |
| Preterm infants unable to be weaned from the ventilator | Late (>7 days) postnatal corticosteroids [79] | Reduced need for assisted ventilation, chronic lung disease and possibly neonatal mortality. High doses are associated with gastrointestinal bleeding and hypertension |
| Infants on nasal CPAP | Short binasal prongs [76] | Reduced extubation failure |
| Preterm infants in the first month | Oxygen saturation targeting 90–95 % [91] | May reduce mortality compared to oxygen saturation targeting <90 % |
| Preterm infants with or at risk of retinopathy | Oxygen saturation targeting 90–95 % [90] | Reduced pulmonary adverse events to 3 months |

women at risk of preterm or complicated near-term and postmature delivery; targeted use of tocolytics (preterm labour), progesterone and/or cervical cerclage (singleton pregnancies in high risk women) and antibiotics (women with PPROM); and maturation of the foetus with corticosteroids. Postnatal strategies for preventing mechanical ventilation include use of nasal CPAP, nasal IPPV and thin catheter surfactant. Surfactant strategies for improving outcomes include early and repeated use where indicated. Mechanical ventilation strategies include use of short inspiratory time and higher-rate ventilation, a goal of brief mechanical ventilation (<1 h), extubation from low-rate ventilation rather than endotracheal tube CPAP, use of caffeine to facilitate extubation and extubation to nasal CPAP or nasal IPPV where required.

## References

1. Bolisetty S, Legge N, Bajuk B, Lui K, New South W, the Australian Capital Territory Neonatal Intensive Care Units' Data C. Preterm infant outcomes in New South Wales and the Australian capital territory. J Paediatr Child Health. 2015;51(7):713–21.
2. Patel H, Beeby PJ, Henderson-Smart DJ. Predicting the need for ventilatory support in neonates 30–36 weeks' gestational age. J Paediatr Child Health. 2003;39:206–9.
3. Melamed N, Klinger G, Tenenbaum-Gavish K, Herscovici T, Linder N, Hod M, Yogev Y. Short-term neonatal outcome in low-risk, spontaneous, singleton, late preterm deliveries. Obstet Gynecol. 2009;114:253–60.
4. Yoder BA, Gordon MC, Barth Jr WH. Late-preterm birth: does the changing obstetric paradigm alter the epidemiology of respiratory complications? Obstet Gynecol. 2008;111:814–22.
5. Ghartey K, Coletta J, Lizarraga L, Murphy E, Ananth CV, Gyamfi-Bannerman C. Neonatal respiratory morbidity in the early term delivery. Am J Obstet Gynecol. 2012;207:292.e1–4.
6. Hansen AK, Wisborg K, Uldbjerg N, Henriksen TB. Risk of respiratory morbidity in term infants delivered by elective caesarean section: cohort study. BMJ. 2008;336:85–7.
7. Tita AT, Landon MB, Spong CY, Lai Y, Leveno KJ, Varner MW, Moawad AH, Caritis SN, Meis PJ, Wapner RJ, Sorokin Y, Miodovnik M, Carpenter M, Peaceman AM, O'Sullivan MJ, Sibai BM, Langer O, Thorp JM, Ramin SM, Mercer BM, Eunice Kennedy Shriver NM-FMUN. Timing of elective repeat cesarean delivery at term and neonatal outcomes. N Engl J Med. 2009;360:111–20.
8. Cheng YW, Nicholson JM, Nakagawa S, Bruckner TA, Washington AE, Caughey AB. Perinatal outcomes in low-risk term pregnancies: do they differ by week of gestation? Am J Obstet Gynecol. 2008;199:370.e1–7.
9. Hansen AK, Wisborg K, Uldbjerg N, Henriksen TB. Elective caesarean section and respiratory morbidity in the term and near-term neonate. Acta Obstet Gynecol Scand. 2007;86:389–94.
10. Doyle LW, Bowman E, Davis P, Halliday J. Preterm infants 30–36 weeks' gestation in Victoria--where should they be delivered? Aust N Z J Obstet Gynaecol. 1998;38:298–300.
11. Dunn PM. The respiratory distress syndrome of the newborn: immaturity versus prematurity. Arch Dis Child. 1965;40:62–5.
12. Usher RH, Allen AC, McLean FH. Risk of respiratory distress syndrome related to gestational age, route of delivery, and maternal diabetes. Am J Obstet Gynecol. 1971;111:826–32.
13. Lawn JE, Wilczynska-Ketende K, Cousens SN. Estimating the causes of 4 million neonatal deaths in the year 2000. Int J Epidemiol. 2006;35:706–18.
14. Morrison JJ, Rennie JM, Milton PJ. Neonatal respiratory morbidity and mode of delivery at term: influence of timing of elective caesarean section. Br J Obstet Gynaecol. 1995;102:101–6.
15. Bastek JA, Sammel MD, Pare E, Srinivas SK, Posencheg MA, Elovitz MA. Adverse neonatal outcomes: examining the risks between preterm, late preterm, and term infants. Am J Obstet Gynecol. 2008;199:367.e1–8.

16. Parikh LI, Reddy UM, Mannisto T, Mendola P, Sjaarda L, Hinkle S, Chen Z, Lu Z, Laughon SK. Neonatal outcomes in early term birth. Am J Obstet Gynecol. 2014;211:265.e1–11.
17. Bastek JA, Srinivas SK, Sammel MD, Elovitz MA. Do neonatal outcomes differ depending on the cause of preterm birth? A comparison between spontaneous birth and iatrogenic delivery for preeclampsia. Am J Perinatol. 2010;27:163–9.
18. De Los Santos-Garate AM, Villa-Guillen M, Villanueva-Garcia D, Vallejos-Ruiz ML, Murguia-Peniche MT, Network NEs. Perinatal morbidity and mortality in late-term and post-term pregnancy. NEOSANO perinatal network's experience in Mexico. J Perinatol. 2011;31:789–93.
19. King JF, Flenady VJ, Papatsonis DN, Dekker GA, Carbonne B. Calcium channel blockers for inhibiting preterm labour. Cochrane Database Syst Rev. 2003;(1):CD002255.
20. Mackeen AD, Seibel-Seamon J, Grimes-Dennis J, Baxter JK, Berghella V. Tocolytics for preterm premature rupture of membranes. Cochrane Database Syst Rev. 2011;(10):CD007062.
21. Kenyon S, Boulvain M, Neilson JP. Antibiotics for preterm rupture of membranes. Cochrane Database Syst Rev. 2013;(12):CD001058.
22. Dodd JM, Jones L, Flenady V, Cincotta R, Crowther CA. Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. Cochrane Database Syst Rev. 2013;(7):CD004947.
23. Alfirevic Z, Stampalija T, Roberts D, Jorgensen AL. Cervical stitch (cerclage) for preventing preterm birth in singleton pregnancy. Cochrane Database Syst Rev. 2012;(4):CD008991.
24. Rafael TJ, Berghella V, Alfirevic Z. Cervical stitch (cerclage) for preventing preterm birth in multiple pregnancy. Cochrane Database Syst Rev. 2014;(9):CD009166.
25. Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database Syst Rev. 2006;(3):CD004454.
26. Crowther CA, McKinlay CJ, Middleton P, Harding JE. Repeat doses of prenatal corticosteroids for women at risk of preterm birth for improving neonatal health outcomes. Cochrane Database Syst Rev. 2011;(6):CD003935.
27. Gulmezoglu AM, Crowther CA, Middleton P, Heatley E. Induction of labour for improving birth outcomes for women at or beyond term. Cochrane Database Syst Rev. 2012;(6):CD004945.
28. Antenatal Corticosteroid Clinical Practice Guidelines Panel. Antenatal corticosteroids given to women prior to birth to improve fetal, infant, child and adult health: clinical practice guidelines. Auckland: Liggins Institute, The University of Auckland; 2015.
29. Ahmed MR, Sayed Ahmed WA, Mohammed TY. Antenatal steroids at 37 weeks, does it reduce neonatal respiratory morbidity? A randomized trial. J Matern Fetal Neonatal Med. 2014;12:1486–90.
30. Stutchfield P, Whitaker R, Russell I, Antenatal Steroids for Term Elective Caesarean Section Research T. Antenatal betamethasone and incidence of neonatal respiratory distress after elective caesarean section: pragmatic randomised trial. BMJ. 2005;331:662.
31. Rosenstein MG, Cheng YW, Snowden JM, Nicholson JM, Caughey AB. Risk of stillbirth and infant death stratified by gestational age. Obstet Gynecol. 2012;120:76–82.
32. Henderson-Smart DJ, Wilkinson A, Raynes-Greenow CH. Mechanical ventilation for newborn infants with respiratory failure due to pulmonary disease. Cochrane Database Syst Rev. 2002;(4):CD002770.
33. Ho JJ, Subramaniam P, Henderson-Smart DJ, Davis PG. Continuous distending pressure for respiratory distress syndrome in preterm infants. Cochrane Database Syst Rev. 2002;(2);CD002271.
34. Ho JJ, Henderson-Smart DJ, Davis PG. Early versus delayed initiation of continuous distending pressure for respiratory distress syndrome in preterm infants. Cochrane Database Syst Rev. 2002;(2):CD002975.
35. Tang S, Zhao J, Shen J, Hu Z, Shi Y. Nasal intermittent positive pressure ventilation versus nasal continuous positive airway pressure in neonates: a systematic review and meta-analysis. Indian Pediatr. 2013;50:371–6.
36. Abdel-Latif ME, Osborn DA. Laryngeal mask airway surfactant administration for prevention of morbidity and mortality in preterm infants with or at risk of respiratory distress syndrome. Cochrane Database Syst Rev. 2011;(7):CD008309.

37. Gopel W, Kribs A, Ziegler A, Laux R, Hoehn T, Wieg C, Siegel J, Avenarius S, von der Wense A, Vochem M, Groneck P, Weller U, Moller J, Hartel C, Haller S, Roth B, Herting E, Neonatal N. German. Avoidance of mechanical ventilation by surfactant treatment of spontaneously breathing preterm infants (AMV): an open-label, randomised, controlled trial. Lancet. 2011;378:1627–34.
38. Kribs A, Roll C, Gopel W, Wieg C, Groneck P, Laux R, Teig N, Hoehn T, Bohm W, Welzing L, Vochem M, Hoppenz M, Buhrer C, Mehler K, Stutzer H, Franklin J, Stohr A, Herting E, Roth B, Investigators NT. Nonintubated surfactant application vs conventional therapy in extremely preterm infants: a randomized clinical trial. JAMA Pediatr. 2015;169(8):723–30.
39. Kanmaz HG, Erdeve O, Canpolat FE, Mutlu B, Dilmen U. Surfactant administration via thin catheter during spontaneous breathing: randomized controlled trial. Pediatrics. 2013;131:e502–9.
40. Heidarzadeh M, Mirnia K, Hoseini MB, Sadeghnia A, Akrami F, Balila M, Ghojazadeh M, Shafai F. Surfactant administration via thin catheter during spontaneous breathing: randomized controlled trial in alzahra hospital. Iran J Neonatology. 2013;4:5–9.
41. Mohammadizadeh M, Ardestani AG, Sadeghnia AR. Early administration of surfactant via a thin intratracheal catheter in preterm infants with respiratory distress syndrome: feasibility and outcome. J Res Pharm Pract. 2015;4:31–6.
42. Henderson-Smart DJ, Wilkinson A, Raynes-Greenow CH. Mechanical ventilation for newborn infants with respiratory failure due to pulmonary disease. Cochrane Database Syst Rev. 2002;(40):CD002770.
43. Rojas-Reyes MX, Morley CJ, Soll R. Prophylactic versus selective use of surfactant in preventing morbidity and mortality in preterm infants. Cochrane Database Syst Rev. 2012;(3):CD000510.
44. Dunn MS, Kaempf J, de Klerk A, de Klerk R, Reilly M, Howard D, Ferrelli K, O'Conor J, Soll RF. Vermont oxford network DRMSG. Randomized trial comparing 3 approaches to the initial respiratory management of preterm neonates. Pediatrics. 2011;128:e1069–76.
45. Sandri F, Plavka R, Ancora G, Simeoni U, Stranak Z, Martinelli S, Mosca F, Nona J, Thomson M, Verder H, Fabbri L, Halliday H, Group CS. Prophylactic or early selective surfactant combined with nCPAP in very preterm infants. Pediatrics. 2010;125:e1402–9.
46. Network SSGotEKSNNR, Finer NN, Carlo WA, Walsh MC, Rich W, Gantz MG, Laptook AR, Yoder BA, Faix RG, Das A, Poole WK, Donovan EF, Newman NS, Ambalavanan N, Frantz 3rd ID, Buchter S, Sanchez PJ, Kennedy KA, Laroia N, Poindexter BB, Cotten CM, Van Meurs KP, Duara S, Narendran V, Sood BG, O'Shea TM, Bell EF, Bhandari V, Watterberg KL, Higgins RD. Early CPAP versus surfactant in extremely preterm infants. N Engl J Med. 2010;362:1970–9.
47. Bahadue FL, Soll R. Early versus delayed selective surfactant treatment for neonatal respiratory distress syndrome. Cochrane Database Syst Rev. 2012;(11):CD001456.
48. Stevens TP, Harrington EW, Blennow M, Soll RF. Early surfactant administration with brief ventilation vs. selective surfactant and continued mechanical ventilation for preterm infants with or at risk for respiratory distress syndrome. Cochrane Database Syst Rev. 2007;(4):CD003063.
49. Kendig JW, Ryan RM, Sinkin RA, Maniscalco WM, Notter RH, Guillet R, Cox C, Dweck HS, Horgan MJ, Reubens LJ, Risemberg H, Phelps DL. Comparison of two strategies for surfactant prophylaxis in very premature infants: a multicenter randomized trial. Pediatrics. 1998;101:1006–12.
50. Soll R, Ozek E. Multiple versus single doses of exogenous surfactant for the prevention or treatment of neonatal respiratory distress syndrome. Cochrane Database Syst Rev. 2009;(1):CD000141.
51. El Shahed AI, Dargaville P, Ohlsson A, Soll RF. Surfactant for meconium aspiration syndrome in full term/near term infants. Cochrane Database Syst Rev. 2007;(3):CD002054.
52. Hahn S, Choi HJ, Soll R, Dargaville PA. Lung lavage for meconium aspiration syndrome in newborn infants. Cochrane Database Syst Rev. 2013;(4):CD003486.
53. Abdel-Latif ME, Osborn DA, Challis D. Intra-amniotic surfactant for women at risk of preterm birth for preventing respiratory distress in newborns. Cochrane Database Syst Rev. 2010;(1):CD007916.

54. Abdel-Latif ME, Osborn DA. Pharyngeal instillation of surfactant before the first breath for prevention of morbidity and mortality in preterm infants at risk of respiratory distress syndrome. Cochrane Database Syst Rev. 2011;(3):CD008311.
55. Abdel-Latif ME, Osborn DA. Nebulised surfactant in preterm infants with or at risk of respiratory distress syndrome. Cochrane Database Syst Rev. 2012;(10):CD008310.
56. Spence K, Barr P. Nasal versus oral intubation for mechanical ventilation of newborn infants. Cochrane Database Syst Rev. 2000;(2):CD000948.
57. Findlay RD, Taeusch HW, Walther FJ. Surfactant replacement therapy for meconium aspiration syndrome. Pediatrics. 1996;97:48–52.
58. Lotze A, Mitchell BR, Bulas DI, Zola EM, Shalwitz RA, Gunkel JH. Multicenter study of surfactant (Beractant) use in the treatment of term infants with severe respiratory failure. Survanta in term infants study group. J Pediatr. 1998;132:40–7.
59. Ng E, Taddio A, Ohlsson A. Intravenous midazolam infusion for sedation of infants in the neonatal intensive care unit. Cochrane Database Syst Reviews. 2012;(6):CD002052.
60. Bellù R, de Waal KA, Zanini R. Opioids for neonates receiving mechanical ventilation. Cochrane Database Syst Rev. 2008;(1):CD004212.
61. Cools F, Offringa M. Neuromuscular paralysis for newborn infants receiving mechanical ventilation. Cochrane Database Syst Rev. 2005;(2):CD002773.
62. Kamlin C, Davis PG. Long versus short inspiratory times in neonates receiving mechanical ventilation. Cochrane Database Syst Rev. 2004;(4):CD004503.
63. Greenough A, Dimitriou G, Prendergast M, Milner AD. Synchronized mechanical ventilation for respiratory support in newborn infants. Cochrane Database Syst Rev. 2008;(1):CD000456.
64. Wheeler K, Klingenberg C, McCallion N, Morley CJ, Davis PG. Volume-targeted versus pressure-limited ventilation in the neonate. Cochrane Database Syst Rev. 2010;(11):CD003666.
65. Cools F, Henderson-Smart DJ, Offringa M, Askie LM. Elective high frequency oscillatory ventilation versus conventional ventilation for acute pulmonary dysfunction in preterm infants. Cochrane Database Syst Rev. 2009;(3):CD000104.
66. Henderson-Smart DJ, De Paoli AG, Clark RH, Bhuta T. High frequency oscillatory ventilation versus conventional ventilation for infants with severe pulmonary dysfunction born at or near term. Cochrane Database Syst Rev. 2009;(3):CD002974.
67. Barrington KJ, Finer N. Inhaled nitric oxide for respiratory failure in preterm infants. Cochrane Database Syst Rev. 2010;(12):CD000509.
68. Finer NN, Barrington KJ. Nitric oxide for respiratory failure in infants born at or near term. Cochrane Database Syst Rev. 2006;(4):CD000399.
69. Shah PS, Ohlsson A. Sildenafil for pulmonary hypertension in neonates. Cochrane Database Syst Rev. 2011;(8):CD005494.
70. Rojas MA, Lozano JM, Rojas MX, Bose CL, Rondon MA, Ruiz G, Pineros JG, Rojas C, Robayo G, Hoyos A, Celis LA, Torres S, Correa J, Columbian Neonatal Reseaerch N. Randomized, multicenter trial of conventional ventilation versus high-frequency oscillatory ventilation for the early management of respiratory failure in term or near-term infants in Colombia. J Perinatol. 2005;25:720–4.
71. Clark RH, Yoder BA, Sell MS. Prospective, randomized comparison of high-frequency oscillation and conventional ventilation in candidates for extracorporeal membrane oxygenation. J Pediatr. 1994;124:447–54.
72. Henderson-Smart DJ, Davis PG. Prophylactic methylxanthines for endotracheal extubation in preterm infants. Cochrane Database Syst Rev. 2010;(12):CD000139.
73. Davis PG, Henderson-Smart DJ. Extubation from low-rate intermittent positive airways pressure versus extubation after a trial of endotracheal continuous positive airways pressure in intubated preterm infants. Cochrane Database Syst Rev. 2001;(4):CD001078.
74. Davis PG, Henderson-Smart DJ. Nasal continuous positive airways pressure immediately after extubation for preventing morbidity in preterm infants. Cochrane Database Syst Rev. 2003;(2):CD000143.
75. Davis PG, Lemyre B, de Paoli AG. Nasal intermittent positive pressure ventilation (NIPPV) versus nasal continuous positive airway pressure (NCPAP) for preterm neonates after extubation. Cochrane Database Syst Rev. 2001;(2):CD003212.

76. De Paoli AG, Davis PG, Faber B, Morley CJ. Devices and pressure sources for administration of nasal continuous positive airway pressure (NCPAP) in preterm neonates. Cochrane Database Syst Rev. 2008;(1):CD002977.
77. Wilkinson D, Andersen C, O'Donnell CP, De Paoli AG. High flow nasal cannula for respiratory support in preterm infants. Cochrane Database Syst Rev. 2011;(5):CD006405.
78. Halliday HL, Ehrenkranz RA, Doyle LW. Early (<8 days) postnatal corticosteroids for preventing chronic lung disease in preterm infants. Cochrane Database Syst Rev. 2010;(1):CD001146.
79. Halliday HL, Ehrenkranz RA, Doyle LW. Late (>7 days) postnatal corticosteroids for chronic lung disease in preterm infants. Cochrane Database Syst Rev. 2009;(1):CD001145.
80. Davis PG, Henderson-Smart DJ. Intravenous dexamethasone for extubation of newborn infants. Cochrane Database Syst Rev. 2001;(4):CD000308.
81. Ward M, Sinn J. Steroid therapy for meconium aspiration syndrome in newborn infants. Cochrane Database Syst Rev. 2003;(4):CD003485.
82. Stewart A, Brion LP, Soll R. Diuretics for respiratory distress syndrome in preterm infants. Cochrane Database Syst Rev. 2011;(12):CD001454.
83. Stewart A, Brion LP, Ambrosio-Perez I. Diuretics acting on the distal renal tubule for preterm infants with (or developing) chronic lung disease. Cochrane Database Syst Rev. 2011;(9):CD001817.
84. Stewart A, Brion LP. Intravenous or enteral loop diuretics for preterm infants with (or developing) chronic lung disease. Cochrane Database Syst Rev. 2011;(9):CD001453.
85. Tan A, Schulze A, O'Donnell CP, Davis PG. Air versus oxygen for resuscitation of infants at birth. Cochrane Database Syst Rev. 2005;(2):CD002273.
86. Brown JV, Moe-Byrne T, Harden M, McGuire W. Lower versus higher oxygen concentration for delivery room stabilisation of preterm neonates: systematic review. PLoS One. 2012;7:e52033.
87. Group BIUKC, Group BIAC, Group BINZC, Stenson BJ, Tarnow-Mordi WO, Darlow BA, Simes J, Juszczak E, Askie L, Battin M, Bowler U, Broadbent R, Cairns P, Davis PG, Deshpande S, Donoghoe M, Doyle L, Fleck BW, Ghadge A, Hague W, Halliday HL, Hewson M, King A, Kirby A, Marlow N, Meyer M, Morley C, Simmer K, Tin W, Wardle SP, Brocklehurst P. Oxygen saturation and outcomes in preterm infants. N Engl J Med. 2013;368:2094–104.
88. Lloyd J, Askie L, Smith J, Tarnow-Mordi W. Supplemental oxygen for the treatment of prethreshold retinopathy of prematurity. Cochrane database of systematic reviews. 2003;(2):CD003482.
89. Henderson-Smart DJ, Steer PA. Caffeine versus theophylline for apnea in preterm infants. Cochrane Database Syst Rev. 2010;(1):CD000273.
90. Chen ML, Guo L, Smith LE, Dammann CE, Dammann O. High or low oxygen saturation and severe retinopathy of prematurity: a meta-analysis. Pediatrics. 2010;125:e1483–92.
91. Tarnow-Mordi WO, Darlow B, Doyle L. Target ranges of oxygen saturation in extremely preterm infants. N Engl J Med. 2010;363:1285.

# Case Study in Nursing

# 17

Walter De Caro, Anna Rita Marucci, Loreto Lancia, and Julita Sansoni

**Abstract**

Nursing complexity requires comprehensive tools for evidence appraisal and synthesis, able of taking into account several factors driving outcomes, resource use, and patient well-being. Umbrella reviews, overviews of reviews, and meta-epidemiologic studies offer a unique opportunity to capture and navigate such complexity, without disregarding the multiple evidence sources informing on nursing. In this chapter, a set of key umbrella reviews is presented on nursing which offer a poignant case study on the pros and cons of this kind of research design in this clinical and research discipline.

## 17.1 Introduction

Nurses are the largest group of clinical practitioners' workforce worldwide. They are positioned to make important contributions to improve health and quality of life. According to the International Council of Nurses (ICN), nursing includes

W. De Caro (✉) • A.R. Marucci
Nursing Research Unit, Department of Public Health and Infectious Disease, Sapienza University of Rome, Rome, Italy

Nursing Science, Department of Life Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy
e-mail: walter.decaro@uniroma1.it

L. Lancia
Nursing Science, Department of Life Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy

J. Sansoni
Nursing Research Unit, Department of Public Health and Infectious Disease, Sapienza University of Rome, Rome, Italy



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_17

"autonomous and collaborative care of individuals of all ages, families, groups and communities, sick or well and in all settings" [1]. Moreover nursing encompasses "the promotion of health, prevention of illness, and the care of ill, disabled and dying people. Advocacy, promotion of a safe environment, research, participation in shaping health policy and in patient and health systems management, and education are also key nursing roles."

For the American Nursing Association (ANA), "nursing is the protection, promotion, and optimization of health and abilities, prevention of illness and injury, alleviation of suffering through the diagnosis and treatment of human response, and advocacy in the care of individuals, families, communities, and populations" [2].

There are six essential features of professional nursing:

1. Provision of a caring relationship that facilitates health and healing
2. Attention to the range of human experiences and responses to health and illness within the physical and social environments
3. Integration of objective data with knowledge gained from an appreciation of the patient or group's subjective experience
4. Application of scientific knowledge to the processes of diagnosis and treatment through the use of judgment and critical thinking
5. Advancement of professional nursing knowledge through scholarly inquiry and research

The majority of people are unaware that nurses conduct research. Fitzpratick and Joyce in an editorial wrote that "the person in the street" has little understanding as to "what nursing research is or its benefits to the health and welfare of all citizens" and at same time "consider nursing to be a subset of medicine" [3]. However, nursing is not directed by physicians, even though nurses have less power in comparison to physicians.

In addition to extensive medical expertise, nurses have a unique, holistic patient advocacy focus, a unique scope of practice, and a unique body of knowledge, including special expertise in areas such as patient education, wound care, and pain management. Research priorities in nursing must take into consideration individual and collective needs in health (clinical and public health), both in macro and micro social environments. Studies should be focused on the social structure which generates health or disease, without neglecting the presence of the actor's subjective world. The theoretical frameworks should be both inter- and transdisciplinary constructions.

However, "nursing is a human discipline that facilitates individuals and families wellbeing and communities using a scientific knowledge base within caring relationships" [4].

The future needs to reflect nursing's unique contribution to care and related outcomes within an interdisciplinary environment; nurses need knowledge to inform and transform care delivery, improving quality and safety of care.

Nurse must be active participants in research and in the development of scientific knowledge.

This chapter presents two literature reviews: one is an umbrella review which aims to clarify some issues on handover during daily care routine, and the second is focused on e-learning its effects on learning environment (i.e., universities) and knowledge building within nursing education.

Handover is an important moment in the daily healthcare routine and a key aspect on healthcare delivery; this is because a wrong or confused handover can lead to a wide range of problems both from an organizational point of view and in terms of patient safety. As far as e-learning is concerned, the use of information and computing technology within the academia should be explored in depth as there is an overall need to evaluate its effectiveness per se and in comparison to traditional methods of learning and teaching; this is to have an evaluation and a possible implementation within nursing studies which are traditionally strongly based on relationships and therefore could potentially lose some important insights and features with the use of e-learning.

## 17.2 Case Study 1: Handover and Nursing

### 17.2.1 Synopsis

#### 17.2.1.1 Introduction

Nursing handover means the exchange of information among nurses about patients' conditions. This process is essential for nursing practice in terms of quality of care, patient safety, and continuity of care. To date, there is no agreement with respect to the best way to carry out nursing handover. The purpose of this work was to synthesize the secondary literature with reference to evidence on nursing handover, methods, and tools used for handover process, paying attention for new research activities.

#### 17.2.1.2 Methods

Comprehensive searches of scientific literature (systematic review and integrative review) were conducted in five electronic databases (PUBMED, SCOPUS, CINAHL, COCKRANE DATABASE, CRD DATABASE); no language restrictions were applied. The search strategy consisted of keywords and medical subject headings for handover (and related term, handover, hand-off, handoff, sign out, shift report) AND nursing as population. In addition, searches throughout reference lists were conducted to identify additional citations.

#### 17.2.1.3 Results

Twelve revisions met the inclusion criteria.

#### 17.2.1.4 Conclusion

Further studies should be carried out in this area given the lack of quality studies that may show which is the best way to carry out handover process in terms of styles, content, and tools. A key aspect to pay attention is the context since it affects

handover content. Structured handover through EHR systems together with face-to-face handover is crucial in providing a better communication process and easier data access as well as in improving quality of care; besides that, there is a need to develop educational tools on the topic currently absent in health professional curriculum.

### 17.2.2 Introduction

Clinical care is continuum that also involves all the information that the health professions and paramedics exchanged both at the referral of a patient in a hospital by a specialist or primary care patient discharge from hospital (hereinafter, "handover"); it is one of the most critical aspect of a patient care and involves some key aspects of the clinical care process: transfer of information, professional responsibility, and accountability for patient care from one clinical team to another either temporarily or permanently, as focused in definitions.

Literature suggests several definitions of handover. Australian Medical Association and UK National Patient Safety Agency give as definition of handover: "the transfer of professional responsibility and accountability for some or all aspects of care for a patient, or group of patients, to another person or professional group on a temporary or permanent basis" [5]. Cohen and Hilligross define handover as "the exchange between health professionals of information about a patient accompanying either a transfer of control over, or of responsibility for, the patient" [6].

Handover represents an umbrella term of synonym terms or terms that can be traced to handover (i.e. hand-off, shift report, health record, shift change) or for area of transfer. Handover can refer to one of diverse transfers that exist in healthcare service context: from specific provider to similar provider (i.e., nurse to nurse) or for primary care service to secondary care or diagnostic department, between wards in similar department or in-hospital or for ambulance service to emergency department.

Handover is one of the main aspects in clinical governance, and it has been identified as one of the main concerns during patient's hospitalization. Generally speaking, handover process is performed by different healthcare professionals as nursing or medical teams; handover is frequently pressured by time constraints, and it can lead to miss important information due to poor structure and process [7, 8].

Also another area of interest is the fact that the handover in most cases is paper based and at the same time is unstructured language, expressed in natural process languages. Some area of overlap and redundancy exit with different types of forms used for writing information: at the same time, the use of standardized languages appears as limited.

We carried out an umbrella review which aims to address related research questions to summarize the best evidence in the field, individuate standard methods, and clarify areas where it is necessary for new research activities to focus on.

The issues of the whole process are overemphasized by the overall use of similar words as well as the lack of use of standardized language.

### 17.2.3 Objective

The scientific literature on nursing handover has been documented in a number of systematic reviews in recent years. We aimed to conduct an umbrella review to provide a more comprehensive overview in the field as follows:

- Summarize the best evidence on handover for nursing practice.
- Explore standards and methods used for handover process.
- Clarify area which are important topic to explore in future research activities research activities.

### 17.2.4 Methods

#### 17.2.4.1 Inclusion Criteria

1. Systematic review and integrative review
2. Focus on nursing handover (or related research)
3. Presence of abstract
4. English language publications only
5. No limits of year of publication

#### 17.2.4.2 Exclusion Criteria

1. Reviews focused on handover for physician or other healthcare professional
2. Reviews without clear selection process flowchart
3. Reviews with other outcome than nursing handover

#### 17.2.4.3 Quality Assessment

All eligible reviews were assessed independently by two researchers. The AMSTAR quality assessment tool (a measurement tool to assess systematic reviews) was used to evaluate reviews. AMSTAR is an 11-item tool to assess methodological quality of systematic reviews that has been found to have good reliability. A minimum score for inclusion is 7.

#### 17.2.4.4 Data Analysis

We conducted dual, independent data extraction using a standardized form. Disagreements were resolved by consensus or consultation with a third researcher. Within a review, studies were included in the analysis if they addressed nursing handover, no context defined.

### 17.2.5 Result

Twelve reviews were included in the present study. The flow of studies through the selection process was presented in Fig. 17.1. Summaries and overall findings of the included reviews are reported for each review in Table 17.1.

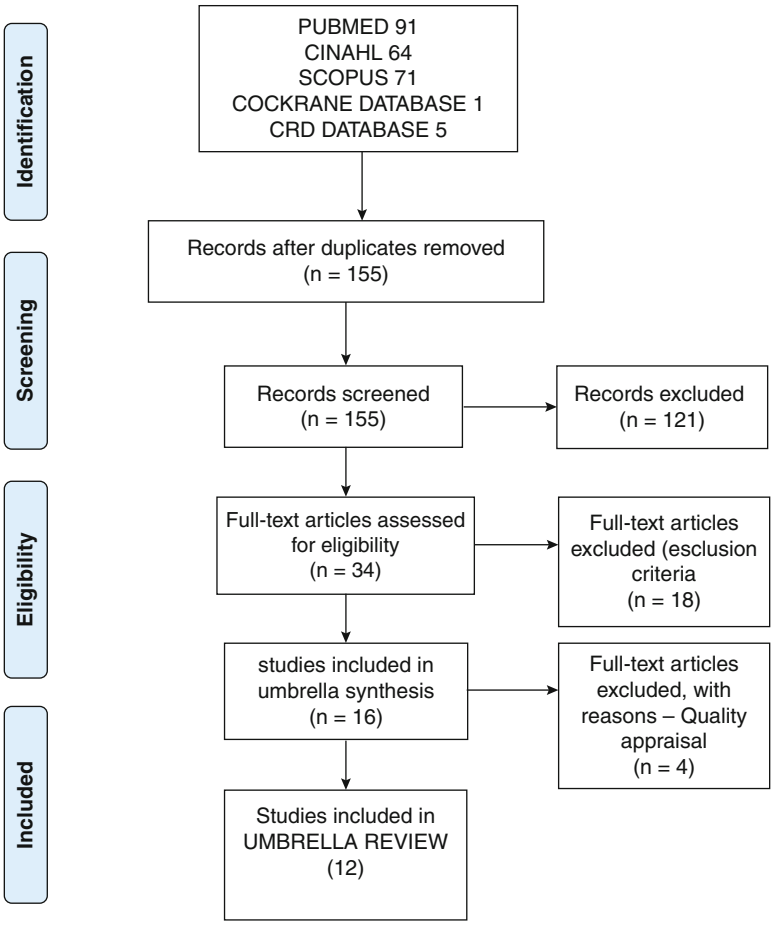



**Fig. 17.1** Flow chart for case study 1

### 17.2.6 Discussion

Articles included in the review show handover process in a number of different settings. Among studies under analysis, there are some focused on ambulance and hospital handover [9, 10]; some on intervention effectiveness evaluation [11–15]; some which explain the topic before, then set up a new handover standardization as the electronic one [16, 17]; and some focused on the content to identify process issues [18]. It is worth underlining that the categorization used here is developed with the main purpose and aim to summarize findings of this review; however, there are several articles which do not fall completely and exclusively into one category.

**Table 17.1** Key features of included reviews for case study 1

| First author | Title | Year | Number of included studies | Population target | Aim | Setting | Overall findings |
|---|---|---|---|---|---|---|---|
| Abraham J. | A systematic review of the literature on the evaluation of handoff tools: Implications for research and practice | 2014 | 36 | Medical and nursing staff | Evaluation studies of handoff tools | Critical care, non-critical care | The majority of the studies show to have electronic handovers. Nearly all EMR integrated handovers are set up for specific wards. Due to the fact that standardized handoff is devised for the specific ward, there is a need for more research with a larger sample/settings<br>Study design: most of the studies are observational studies with some limitations in terms of strength of evidences; furthermore almost all studies had very small sample sizes |
| Smeulers M. | Effectiveness of different nursing handover styles for ensuring continuity of information in hospitalised patients (Review) | 2014 | 0 | Nurses | To determine the effectiveness of interventions designed to improve hospital nursing handover | All teaching or university hospitals | The research question remains unanswered because none of the studies found was included in the review (no cluster RCT or RCT) which confirms the lack of quality nursing studies in this area |

(continued)

**Table 17.1** (continued)

| First author | Title | Year | Number of included studies | Population target | Aim | Setting | Overall findings |
|---|---|---|---|---|---|---|---|
| Flemming D. | How to improve change of shift handovers and collaborative grounding and what role does the electronic patient record system play? Results of a systematic literature review | 2013 | 60 | Medical and nursing staff | What are the typical errors and their consequences in handovers? How can they be overcome by conventional strategies and instruments? Are there any instruments to support collaborative grounding? | Outgoing shift and oncoming shift in healthcare institution | Several studies display that electronic handover tools integrated into EPR systems providing more and better information with respect to paper based handovers. Quality of handovers depend on the structure, quantity, and quality of information as well as on type of information, i.e., anticipatory guidance and other subjective information and holistic information about the patient as a clinical case, (not contained in conventional electronic record systems.) Some studies provide recommendations on how electronic handover tools integrated into EPR systems should be designed. Recommendations should be implemented and evaluated |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Jensen S.M. | Handover of patients: a topical review of ambulance crew to emergency department handover | 2013 | 18 | Physicians, nurses, and paramedics | To identify factors that influence handover process | Handover from ambulance crew to emergency department | Information gaps – lack of active listening during handover, loss of information for more major injuries, due to the failure to deliver in two steps (at the ambulance arrival and after starting treatments for stabilization); there are too many handovers: ambulance/ triage/nurse-in-charge Strategies to improve the process (tips) structuring the transfer of information with tools. Some studies suggest the use of computer to send information from the ambulance to the hospital Cultural differences and organizational aspects of some studies mention difficulties in verbal handover due to misunderstanding of health professionals |

(continued)

**Table 17.1** (continued)

| First author | Title | Year | Number of included studies | Population target | Aim | Setting | Overall findings |
|---|---|---|---|---|---|---|---|
| Kitson A.L. | What's my line? A narrative review and synthesis of the literature on registered nurses' communication behaviours between shifts | 2013 | 29 | Nursing staff (RN) | To identify gaps and inconsistencies in the handover process | Handovers in adult hospital settings | Overall lack of quality research on conceptual frameworks on complexities of handover. Article identifies seven consistent themes across seminal and empirical papers; however, the chronological evolution of these themes showed overall limited development<br>Currently research has not been able to investigate the role of the unit that leads in determining nurses' communication attitudes |
| Foster S. | The effects of patient handoff characteristics on subsequent care: A systematic review and areas for future research | 201 | 18 | Medical and nursing staff | Assessing the empirical evidence on the relationships between handoff characteristics and outcomes was to identify recurring methodological problems in previous research by examining the studies' quality with particular regard to their potential for causal inference | Patient handoffs within hospitals, including those between paramedics and the emergency department | This article displays that research brings very different results on the subject, and there is an overall difficulty to pair the studies Therefore, it is difficult to draw significant conclusion in order to develop projects on standardized form for handover (some studies found standardization form positive, some not); furthermore, standardized form varies among the different settings and is observational |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Staggers N. | Research on nursing handoffs for medical and surgical settings: An integrative review | 2012 | 30 | Nurses | To synthesize the outcomes of nursing handoff research (for future computerizing handoff) | Medical and surgical units | Primary role of the nurse in the handover, also at bedside (patients do not participate for a number of reasons)<br>Handover is crucial for nurses not only for the transfer of information but also for social, emotional (ritual), and educative aspects. Hence, the importance of preserving deliveries face- to-face (even with electronic systems)<br>Structured data improves handover process overall |
| Gordon M. | Educational interventions to improve handover in health care: a systematic review | 2011 | 10 | Medical and nursing staff (students included) | Assessment of educational intervention on handover | The setting was inpatient medical establishments, no acute inpatient setting | Several studies are methodologically weak; there is an overall lack in quality of the studies and are not replicable<br>Kirkpatrick model was used (only one study reaches level 3)<br>At international level, there are no recognized competences for handover delivery; therefore, there is a heterogeneity in the educational intervention |

(continued)

**Table 17.1** (continued)

| First author | Title | Year | Number of included studies | Population target | Aim | Setting | Overall findings |
|---|---|---|---|---|---|---|---|
| Collins S.A. | Content overlap in nurse and physician handoff artifacts and the potential role of electronic health records: A systematic review | 2011 | 36 | Nurse and physicians | Content overlap between nurse and physician hospital-based handoff | Acute care | Interdisciplinary differences in handover information<br>Interdisciplinary handover element list and its role in establishing common ground<br>Missing inpatient handover elements<br>EHR support of overlapping interdisciplinary handover information |
| Bost N. | Clinical handover of patients arriving by ambulance to the emergency department – A literature review | 2010 | 8 | Physicians, nurses, and paramedics | Critical review of research on clinical handover between the ambulance service and emergency department | Emergency department | BARRIERS TO DELIVERY TRANSFER<br>Lack of a common language and/or understanding between the different disciplines<br>Inattention during handover<br>Disproportionate amount of information exchanged<br>Lack of leadership/lack of teamwork<br>Staff perceived lack of a common language and framework for minimal information<br>Three out of the eight studies included in this review have recommended to standardized handover |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Riesenberg L.A. | Nursing handoffs:A systematic review of the literature | 2010 | 95 | Nurses | Identify strategies for effective handoffs | Nursing handoff | The research identified several strategies to improve handover process (7 categories) Communication skills, standardization strategies, technologic solutions, environmental strategies, educational and training leadership staff involvement |
| Arora V.M. | Hospitalist handoffs: A systematic review and task force recommendations | 2009 | 10 | Staff, patient, and system | To develop recommendations for hospitalist handoffs | During shift change and service change in same wards | Supports the use of a verbal handoff supplemented with written documentation or a technological solution in a structured format<br>Lack of studies able to build type A recommendations (all are of type C, except one of a B type)<br>Almost all studies have small samples, without control group and none are multicenter<br>Nonetheless, recommendations were drawn and divided into guidelines and recommendations for verbal handover and content delivery |

#### 17.2.6.1 Special Context (Ambulance to Emergency Department)

According to the articles under analysis, there is a high interest around handover issues between ambulance service and emergency department (articles come from the USA); this could be due to the fact that paramedics are not felt as highly qualified professionals. Several aspects make handover under this setting particularly challenging such as the large amount of information, the timing of communication, as well as the number of individuals involved (i.e., ambulance/triage/nurse).

Bost et al. state to use for their review Cochrane protocol to evaluate quality of articles; however, due to the lack of RCTs and intervention studies, this is impossible [9]. Therefore, the Polit and Beck model was used [19]. This is the same for Jensen, who decides to not rate in any way articles under analysis. Both revisions agree on the need, within this context, of developing an organizational culture that ensures the use of a common language, respect for hierarchy, and teamwork structure [10]. In the light of that, it would be important to develop educational modules in order to enhance communication among different professionals (physicians, nurses, paramedics).

Some studies included by Jensen recommend to use ICT facilities to send information directly from the ambulance to the emergency room in order to give a more rapid and straightforward answer to patients. In this way, there will be a significant reduction in time waste and transcription errors, and there will be a construction of a common language.

Authors emphasize relevance of the teamwork; although in this context, it is hard to build it due to the lack of recognition of the skills of other health professionals which leads to a challenging collaboration. Furthermore, these scholars recognize that it is timely to develop a standard handover process through tools as IMST-AMBO which are promising, seen that all health professionals in this peculiar handover process have been involved in validation.

#### 17.2.6.2 Tools and Intervention to Improve Handover Process

Several studies among those in analysis are focused on intervention evaluation targeting different aspects useful to improve handover process. The results of Cochrane review of 2014 are quite relevant; this study, carried out rigorously with respect to studies selection, aims to identify "the effectiveness of interventions designed to improve hospital nursing handover" (pag. 4). The study according to inclusion criteria (RCT or cluster RCTs) did not find any work which could be included [12].

Interventions under evaluation can be focused on wide range of aspects, e.g., content delivery (use/absence of form-template-checklist), verbal, written, recorded, mix model, and venue (nurses' room, bedside, and so on). According to that intervention, analysis could be focused on written handover in different venues (i.e., bedside vs office) or verbal vs nonverbal or recorded or on the type of content (use/absence of form-template-checklist). Evaluation results are carried out following well-known indicators for the assessment of adverse events, medication errors, complications, mortality, or sentinel event. Although authors conclude on the

absence of reliable evidence, there are several examples of researchers who have tried to assess the effectiveness of the styles of nursing handover in order to improve the safety and quality of care. This review (18 of 28 studies identified) pays attention to the introduction of new tools within local experience with a pre- and post-assessment in order to improve quality; these aspects leads to a difficult replicability. Other articles were excluded considering that they did not assess their effectiveness in terms of results but simply on the perception of satisfaction process by nurses and in some rare cases of patients (two).

Similar to that of Smeulers from a methodological rigor point of view are the works of Foster and Monser [20] and Abraham et al. [15], who conduct a systematic review to identify the relationships between the characteristics of handover and healthcare results through a search for evidences. The two studies, which were focused on handover between physicians and nurses, have similar conclusion; results are inconclusive as studies are really different, and it is difficult to pairing to drawn an overall conclusion.

Arora et al. [11] and Riensberg et al. [14] strive to develop recommendations and guidelines which could help nurses to improve handover process; the two studies have similar conclusion. Staff involvement (also found in Bost 2010 and Jensen et al.) [9, 10] and education training on handover process (both for physicians and nurses), across all studies, are the main points to work on. There is an overall call to introduce innovative ICT solutions, structured template checklist, and mnemonics standardized (e.g., SBAR, ISOBAR, etc.) together with verbal handover which is seen as key in order to assure data completeness (interactive – process) [16, 21]. Riensberg also highlights how environment represents a fundamental role in handover process to limit distractions, interruptions, and noises that can influence the final result and privacy during handover.

A particularly interesting study is that of Gordon et al. which analyzed the literature (handover) of medical and nursing with the specific aim to identify the effectiveness of the training in this process; the author used the Kirkpatrick model and one study was rated 3 [13].

In this study once more an element (common to all studies) is the lack of quality, characterized by studies with methodological weaknesses together with the lack of publication of educational material used during interventions (therefore not reproducible). More used methodologies are simulation, role playing, and the use of library materials to discuss and develop skills in handover process.

Emerged themes were information management, team working, communication, leadership, error awareness, and professional behavior. It should bear in mind that handover process among health professionals is not included in the curricula in universities which leads to a heterogeneity of educational interventions. In the light of the above, education in this area probably is taught during clinical practice, the observation of colleagues, first as a student and then as a newly hired, becoming a teaching tool for communicating data.

In the light of the above, there is an inherent complexity in having suggestions to develop new tools for handover; for instance, there are mixed and conflicting results on the use of form-template-checklist with some authors who recognize the

usefulness of standardization and some who do not; studies are overall weak as are observational studies with no control. Scholars point out that there is an overall lack of studies focused on nursing of theoretical construction both on handover and handover standardization.

#### 17.2.6.3 Handover Content

Flemming et al. identify more common errors associated with handover and their consequences [18], whether such errors could be overcome through the use of traditional or electronic tools. They evaluate different structured handovers, both oral and written with checklist and oral mnemonic highlighting that using structured handover there is less data loss.

The sample considered included both physicians and nurses, and a huge overlap was found in the information between the two groups (also in Collins et al.) [17]; therefore, a possible solution could be use a common EHR system. This can be differenciate in some aspects for the to groups where necessary. From a purely nursing a point of view, this suggests a difficulty in EHR systems to give room to the holistic nature of information (attention to the quality of information as well as the amount and structure). The loss of important information during patient care is found to be the most common issue; among analyzed studies, 43 out of 60 show that this issues can be sort out by introducing the use of traditional instruments (SBAR, ISOBAR, etc.) as well as ICT-based tools.

With a different focus, Stagger and Blaz in 2012 conducted an integrative review aimed at identifying outcomes of nursing research handover (for future computerizing handover) to identify critical content for an EHR system [16]. The study started with the identification of the main handover purpose which are transfer of patient information, building a team, and knowledge on actions of care which have to be communicated to a new shift. Several inconsistencies were found between handover and patients' real conditions. Recorded handover has more omission with respect to the unrecorded ones, but also more consistent with the real state of the patient.

There are several ways to carry out handover although many authors do not declare the type. Given the methodology used, which included qualitative studies, the importance of handover as ritual moment, which relieves anxiety and gives a sense of protection for nurses, emerged. According to Riensenberg et al., several problems were identified during the delivery as communication barriers, repeated interruptions, and high levels of noise [14].

Kitson et al. conducts a narrative review that aimed to understand how handover is developed in acute care by meta-narrative review and synthesis, in order to "tell a story" about handover, as part of a healthcare performance [21]. Using the metaphor of care settings, they described each key issue identified in both empirical studies and in seminar work included in the study.

Similarly the studies of Smeleurs et al. [12], Foster and Manser [20], Gordon and Findley [13], Arora et al. [11], Abraham et al. [15], and Flemming and Hübner [18] determines that the low quality of the studies is the main deficiency together with an inherent challenge in research in determining the correct communication/handover

among nurses. Different authors with different paths come to similar conclusions about the fact that handovers are subject to errors; omissions of important information determine errors in decision-making. Furthermore, there is an overall call to identify strategies to develop handover [16, 18, 21].

### 17.2.7 Conclusion

Handover in healthcare facilities is a high-risk moment in terms of patients' safety; indeed this moment entails several dangers such as interruption during care, adverse events, and legal issues. In 2006, the WHO identified handover as one of the five biggest issues for patient safety; the issues became even more serious when the legislation on employment timetable and schedule changed both in Europe and in the USA; with this change, shifts are shorter and there is an increase of handover in daily routine. Besides that, handover is strongly influenced by perceptions, personal characteristics, health professional's knowledge of who are those who exchange the information, as well as by organizational context and ward.

In the light of the above, nurses can potentially play an important role in the loss of data/information about the patient's needs. This loss exposes the patient to a risk for his/her health in terms, for instance, of treatment delay or wrong treatment.

Nursing handover is key in the healthcare process; nevertheless, there are not as yet unique recommendations or guidelines in the literature with respect to what and how nurses should communicate in nursing handover. This leads to a huge challenge in developing a standardized methodology.

All studies included in this review agree with the need of future research in this area, especially given the lack of strong evidences which can support the effectiveness of a way/methodology to transfer (through handover process) healthcare treatment information/data during daily routine.

The review carried out displays some suggestions which can be drawn from the studies analyzed (those with pre- and post-intervention) which however show weak results. For instance the need to keep face-to-face handover even when electronic handover are used to have a more inclusive number of information. Furthemore it should be take into account the setting and specific clients (i.e. older people) in order to evalutate if there is an inherent difficulty in communication, since patients may not be able to answer willingly to health professionals' questions. Indeed in hospital settings, there are several communication barriers due to interruptions, noise and lack of ICT system.

Even in setting where handover is ICT oriented, although the introduction of these tools (SBAR, MIST, and so on) was positive, there is an overall lack of identification of a handover methodology because setting were so different (nurses and patients) as well as different kind of tools leads to a wide range of different standardization handover tools which are not easily replicable in different context. There is furthermore a little presence of multicenter studies that would allow evaluation of broader intervention not just in one unit and an overall lack of educational specific content.

## 17.3 Case Study 2: E-Learning in Nursing Education in Academic Fields

### 17.3.1 Synopsis

#### 17.3.1.1 Introduction

The increasing use of an online learning explosion is kind of revolution that has deeply modified the traditional way of education. The aim of this overview is to conduct an overview of reviews about e-learning in nursing and other healthcare students' education in academic environment, by reviewing reviews.

#### 17.3.1.2 Method

A comprehensive database search was conducted using two electronic database: PubMed/MEDLINE, Ebsco/CINAHL, for the period 2004–2014. The search strategy consisted of keywords and medical subject headings for e-learning (and related terms like distance education, online learning, distance learning, mobile learning, Web-based learning) and nursing (or healthcare students) as population. In addition, searches throughout reference lists were conducted to identify additional citations. Two review authors independently screened results and extracted data from included studies, with any discrepancies settled by a third author.

#### 17.3.1.3 Results

Seventeen reviews were included for this overview of review. Three areas were identified: population (faculty and members), methodologies (blended learning, game/3D, PBL, and situated learning), and evaluation (comparison of e-learning with the traditional method, performance, students' satisfaction)

#### 17.3.1.4 Conclusions

This overview demonstrates that e-learning in nursing is a valid alternative to traditional learning. This study shows that there is a lack of robust evidence on this topic and that the field is constantly under development, especially in some areas as simulation or game/3D activities, althought not strong there are evidence of reduction in the cost of education (in terms of management for instance) as well as a more efficient management of the time for students and lecturers which reduce overall the economic effort afforded by universities and facilitate the management of education environment.

### 17.3.2 Introduction

E-learning education within the academia means new organizational issues. Universities, particularly in healthcare education sector, should have a clearer understanding of the impact of technology on learning. What can be seen as a problem at a first sight can become a formidable challenge for traditional academic institution, especially in healthcare education.

E-learning used in academia for nursing and healthcare professionals' education could be represented like a speeding train. Online learning explosion, just as a revolution, has deeply modified the traditional way of education [22], also in terms of necessity of sharing space and time, that during the years constituted an archetype of the formative moment [23].

Nowadays, there is no commonly accepted and clear definition for e-learning, but it generally refers to distance-based forms of learning rather than face-to-face interaction and every time traditional methods of learning are supported by online resources. The European Union (EU) defines e-learning as "the use of new multimedia technologies and the Internet to improve the quality of learning by facilitating access to resources and services as well as remote exchanges and collaboration" [24]. E-learning is an umbrella term in the literature; there are several terms which have similar meaning such as distance learning, digital learning, distance education, electronic learning, online learning, Web-based learning online education, and now mobile learning.

The key features as stated by Ganino of e-learning are the use of an internet connection and a technological device (computer, tablet, smartphone); enhancement of multimedia; independence from the constraints of physical presence and specific times (always and everywhere); continuous monitoring of the level of learning through self-assessment; interactivity with teaching materials, faculty members, tutors, and other students; and enhancement of social and collaborative learning [25].

In 2010, the American Nurses Association (ANA) recognized e-learning's benefit: "As the nurse of the future evolves, so must nursing educations. Curricula must be designed to adequately prepare competent entry-level nurses. The nurse shortage and program capacity limits demand efficient education process. Online, virtual, simulated, and competency-based learning are attempts to expand opportunities to students and increase efficiency" [26].

In a few years, distance learning has become central in the academic debate for health professional education. In the literature, there are doubts both with respect to the extensive application of e-learning in terms of job market and in terms of relationship and emotional closeness as is a key element for the success of education. Other concern are focused on infrastructure, security, and reliability. Research and extensive analysis can help to clarify direction and identify drop points.

### 17.3.3 Objective

To conduct an overview of reviews about e-learning in nursing and other healthcare students' education in academic environment.

### 17.3.4 Methods

The methodology used for this overview of reviews is aimed at identifying, appraising, and synthesizing evidence from systematic and integrative reviews in order to

synthesize and analyze the evidence generated focusing on e-learning for nursing students and students of other health professions.

#### 17.3.4.1 Search Strategy

The search was made up by text words and index terms into three domains: (1) E-LEARNING (and related terms: e-learning, distance learning, digital learning, distance education, electronic learning, online learning, Web-based learning online education, and now mobile learning.), (2) nursing, and (3) reviews and literature reviews. The Boolean operator "OR" was used to consolidate each domain; furthermore, "AND" operator was employed to cross-reference the three domains.

The search was conducted in August 2014 using the database EBSCO, CINAHL, and Pubmed.

#### 17.3.4.2 Inclusion Criteria

Only integrative and systematic reviews focused on e-learning within the academic environment for nursing and healthcare professionals were included, as indicated below:

- Period: 2003–2013
- Language: Italian – English
- Integrative or systematic review or systematic review with meta-analysis
- Only based within the academic environment
- Only focused on nursing and healthcare students
- Other outcome of review

#### 17.3.4.3 Exclusion Criteria

- Other review than integrative or systematic review
- Review focused on e-learning and CME (continuing medical education)
- Review without clear selection process
- Other language than English

#### 17.3.4.4 Quality Appraisal

The eligible reviews were evaluated for quality appraisal by two reviewers; independently, the quality of each review was assessed according to AMSTAR criteria. Minimum score for the inclusion was 7 (seven) in a maximum grade of 11 (eleven)

### 17.3.5 Result

The first stage of searching was conducted in the two databases, and relevant titles/abstracts were retrieved (780). After the duplicate studies were identified and deleted, two reviewers screened separately the title and abstract of candidate articles

for potential articles. After the full texts of potential studies had been obtained, two reviews (41), working independently, evaluated and selected the articles according to the inclusion criteria and select 27 eligible for inclusion. During the processes, any disagreements between the two reviewers were resolved through consensus. If consensus could not be reached, a third reviewer was consulted for a final decision.

After evaluation of quality criteria appraisal (AMSTAR), 17 were included in this overview, as indicated in Fig. 17.2 and in Table 17.2.

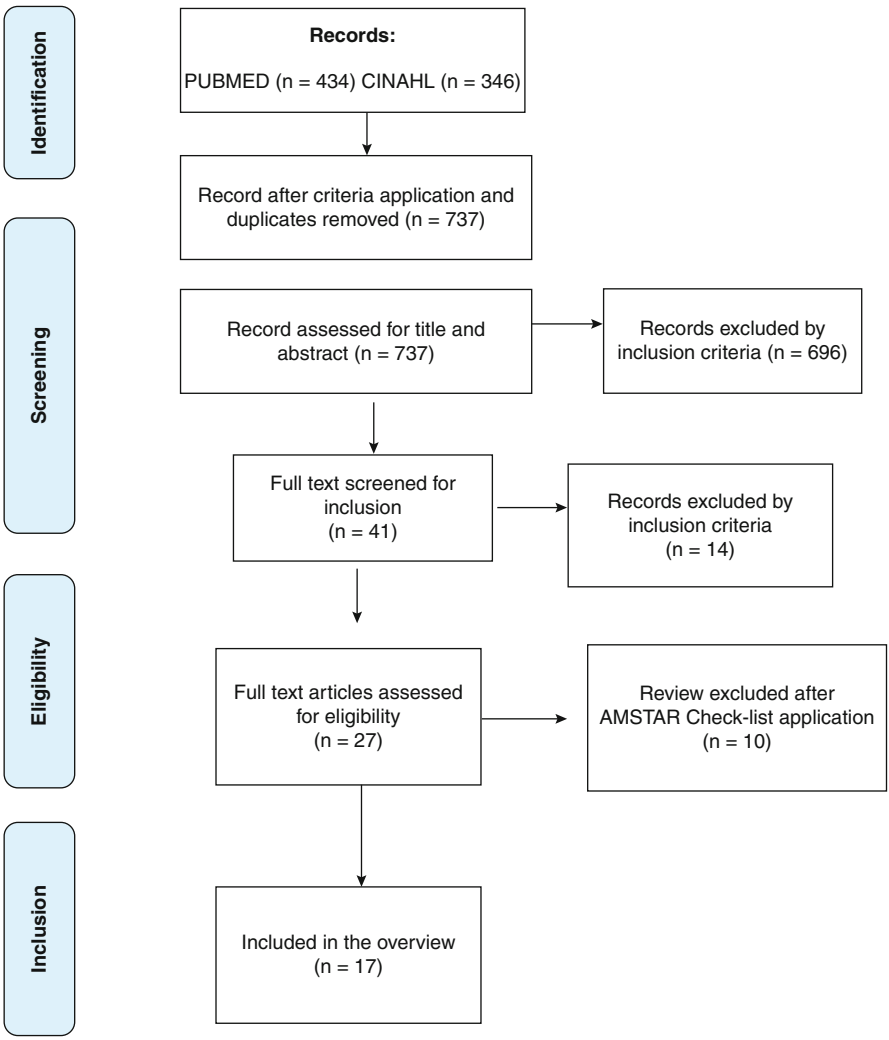



**Fig. 17.2** Flow chart for case study 2

**Table 17.2** Key features of the included reviews for case study 2

| Authors | Titles | Years | Studies | Population – focus | Key findings |
|---|---|---|---|---|---|
| Wilkinson et al. | Measurement of information and communication technology experience and attitudes to e-learning of students in the healthcare professions: integrative review | 2008 | 49 | Nursing student and faculty | Information literacy is a key aspect for reducing gaps in e-learning education<br>Necessity to develop and validate instruments to explore e-learning perspective |
| Cook DA et al. | Internet-based learning in the health professions: A meta-analysis | 2008 | 201 | Health professional student | Internet-based learning compared with no intervention has a consistent positive effect<br>Necessity for more trial to standardize application |
| Bloomfield JG et al. | Using computer assisted learning for clinical skills education in nursing: Integrative review | 2008 | 12 | Nurses and nursing student | Computer-assisted learning has potential as a method of teaching clinical skills in nursing<br>Necessity for more robust methods to investigate |
| Carroll et al. | UK health-care professionals' experience of on-line learning techniques: A systematic review of qualitative data | 2009 | 19 | Health care professional students | Flexibility as key element<br>Improve regular testing or assessment is necessary for evaluate acceptance and performance |
| Mancuso JM | Perceptions of distance education among nursing faculty members in North America | 2009 | 72 | Nursing Faculty | Necessity to establish rules for workload, compensation, support, development, and role of the faculty |
| Booth et al. | Applying findings from a systematic review of workplace-based e-learning: implications for health information professionals | 2009 | 29 | Physician, nurses, and health care students | Need to design and develop new application for support<br>Development of innovative methods of assessment as element to improve application |

**Table 17.2** (continued)

| Authors | Titles | Years | Studies | Population – focus | Key findings |
|---|---|---|---|---|---|
| Cook DA et al. | What do we mean by web-based learning? A systematic review of the variability of interventions | 2010 | 50 | Physician, nurses, and health care students | Exit too much variation in the technology to permit generalizable statements<br>Need more and better research to clarify our use of WBL |
| Cook DA et al. | Instructional design variations in internet-based learning for health professions education: A systematic review and meta-analysis | 2010 | 51 | Nurses and nursing students | Interactivity, practice exercises, repetition, and feedback improve learning outcomes and that interactivity, online discussion, and audio improve satisfaction for health professionals |
| Cook DA et al. | Time and learning efficiency in Internet-based learning: A systematic review and meta-analysis | 2010 | 20 | Health care professional students | Great variability of course delivery<br>Choice to use e-learning provides a logistic advantage for learner groups |
| Lathi M. | Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: A systematic review and meta-analysis | 2012 | 11 | Nurses and nursing students | No difference between e-learning or traditional learning<br>Develop and evaluate methods for education among nurses |
| Graafland m. et al. | Systematic review of serious games for medical education and surgical skills training | 2012 | 25 | Physician, nurses, and health care students | Blended and interactive learning – serious games may be applied to train both technical and nontechnical skills<br>Games need validation before integration into teaching curricula |
| Petty J. | Interactive, technology-enhanced self-regulated learning tools in healthcare education: A literature review | 2012 | 11 | Physician, nurses, and nursing students | Educators do not have to remain stagnant, and there is the need to develop new resources and curriculum delivery<br>E-learning engagement can be variable |

(continued)

**Table 17.2** (continued)

| Authors | Titles | Years | Studies | Population – focus | Key findings |
|---|---|---|---|---|---|
| Patterson BJ | Student outcomes of distance learning in nursing education: an integrative review | 2012 | 33 | Nursing students | Clear orientation to e-learning and consistent use of the Web-based learning platform are also essential to facilitate effective student use of the Web-based classroom |
| Rowe M et al. | The role of blended learning in the clinical education of healthcare students: A systematic review | 2012 | 14 | Nursing students | Further research is necessary before educators make assumptions about the long-term effects of blended learning in clinical education |
| Button D. et al. | E-learning & information communication technology (ICT) in nursing education: A review of the literature | 2013 | 28 | Nursing faculty and nursing students | Development of preregistration nursing curricula for e-learning and ICT technology is compulsory<br>Information (ICT) literacy is an essential learning skill for nurses<br>Nurse educators need more training; computer information technology is needed for nurse faculty |
| Feng JY et al. | Systematic review of effectiveness of situated e-learning on medical and nursing education | 2013 | 14 | Physician and health care professional students | Situated e-learning is an effective method to improve novice health professionals' performance |
| Du S et al. | Web-based distance learning for nurse education: a systematic review | 2013 | 69 | Nurses and nursing students | Web-based distance education has equivalent or better effects in improving knowledge and skills performance for nursing students |

Reviews are being widely heterogeneous in structure and content. However, revisions could be identified and some recurring themes could be grouped: population (students and faculty) evaluation (in terms of acceptance, performance, comparison with traditional system, and evaluation) and e-learning methodologies (i.e., blended, game/3D, situated learning).

### 17.3.6 Population

The World Health Organization and the American Nursing Association suggest that the use of e-learning can be efficient in reducing healthcare professional shortage [27]. For students, several reviews show positive elements such as flexibility of study, knowledge availability, design, and usability as well as communication with faculty, nevertheless no relevant advantages emerged with respect to the traditional way of delivering e-learning [27, 28].

At the same time, universities need to develop and validate instruments to explore students' experiences with e-learning and to develop models for engaging students. An interesting element, in line with the traditional academic learning, is that female students have better performance than male students [22]. E-learning methodologies for nursing students seem to fit better for graduate education, particularly in order to meet the needs of working students [29]. There is an overall agreement in the literature that in using e-learning more interaction and tutoring is required [30]. Web-/mobile-based learning for simulation-based education is the key element for the real placement of e-learning in nursing education offer [29, 31].

The faculty show some problems with respect to ICT literacy as well as in acceptance of e-learning for some courses (ethics, research); furthermore, e-learning is seen as more work to do [22, 30, 32–34]. Scholars and faculty should avoid to connote in terms of polarization and positive and negative e-learning activities. In the wide spectrum of e-learning in terms of content, delivery, interactivity it is not possible to consider it in a unique way, therefore currently e-learning analisys can be measleading. The most useful approach is to study how to use, in every single specific configurations, the most effective way of Web-based learning [32].

There are a number of issues which should be take into account, as the inherent dificulty for lectures and faculties to hava a clear understanding of students' educational need as well as to deliver a basic training in order to use adequately e-learning systems provided [34].

In the light of the above, there is a need to rebuild the traditional approach in terms of:

- Different time consumption
- Necessity of large pre-programming activities
- Different interaction activities
- Different type of evaluation (especially progress evaluation)

### 17.3.7 Methodology

E-learning delivery courses are heterogeneous; indeed it ranges from, for instance, simple remote support (i.e., teaching materials) to a type training with high levels of interaction, both synchronous and asynchronous (i.e., forum, chat audio-video, social. and so on), until simulated scenarios and the game/3D.

Most differences are present, for example, in terms of learning management system digital environment, platform format, either mobile (tablet, smart phones) or desktop [32, 35, 36]. The blended learning approach is the most widely used e-learning approach that combines "face-to-face" presence and online training activities. This approach has benefit for the university (less need of space), faculty (simultaneously face-to-face and online instruction), and for student (preserving social and educational interactions) [22, 29, 32, 37].

In 2011, Keyte et al. showed that in a number of studies, the necessity to provide paper copies of education material to the student to ensure the completion of learning activities had been highlighted. Some issues are connected to evaluate the level of attention payed by students during online sessions, technological difficulties at home and ability to read on screen literature in printed form (as Pdf) [37]. For student participation and consolidation of knowledge, Button et al. [22], recognizing the benefits offered by e-learning, have shown a better performance with the use of this mixed methodology like blended learning [22].

Gaming, simulation, and situated learning are used as educational approaches to motivate students to learn by using video interactivity game and game elements within learning environments. The main aim is to maximize engagement by capturing the interest of learners as well as inspiring them to continue learning. High cost of development, presence ICT specialists, and innovative skills can be a serious limitation to the use and dissemination of these instruments in the field of health education, where in any case should be used along with traditional learning methodology [31, 38, 39].

Several reviews compare traditional methods of education with e-learning strategies. Lahti et al. has shown that there is no significant difference between e-learning and traditional methods with regard to knowledge, skills, and satisfaction; however, e-learning represents an important way of the delivery of nursing education to be implemented time by time with respect to specific courses [28]. Even Wolbrink et al. and Cook et al. (2008) found no significant differences between the media and the traditional methods in terms of efficacy; at same time other scholars display equivalent or slightly better results in terms of gained knowledge, compared to traditional education [40, 41] and higher values of satisfaction with online courses. To sum up, e-learning with an high level of interactiviy show to have a better performance in comparison to a simple way of delivering (i.e. only textual material in electronic format) [35, 42]. E-learning is, conversely, optimal for the post-university training base, because they often have limited time and possibilities to follow a traditional teaching [29]. In terms of graduation or course conclusion, e-learning does not show better time performances with respect to traditional approaches. The

teaching strategies of e-learning, with the highest presence of feedback and interactivity, typically extend the time needed for the completion of learning activities, but in many cases improve the results [43].

Students who have difficulty with the traditional method can have greater confidence in the technical e-learning in our main virtue of the flexibility of this training, special support mechanisms and rapid assessment of learning [27].

### 17.3.8 Evaluation

Many reviews are related to assessments and evaluations on e-learning course delivery. Although the importance of this education is widely acknowledged as being important, there is no strong evidence for evaluating its process, as well as about the effects of new models or approaches [30]. Internet is full of Web resources on e-learning courses, divided by areas of specialization (e.g., informatics, intensive care). However, no organic framework exist allowing their maximum use by the students. The use of Web-based learning reports encouraging results to improve the knowledge, performance, and competence of the participants in specific nursing activities, with a high rate of satisfaction [42].

Information literacy is a key aspect in nursing practice and nurses' careers; a clear orientation to and consistent use of the Web-based learning platform are also essential to facilitate students in the development of their knowledge [22, 44]. At the same time, e-learning courses require the faculty for a training on information literacy especially for the "older generation" [44]. Button et al. have shown that 19 studies analyzed on 28 recommend to incorporate in the e-learning courses preliminary notions of ICT literacy; one of the major source of frustration, which hinders e-learning course appreciation, is the lack of clarity of the instructions for the use of e-learning courses [22]. This pre-training has an important impact on the outcome and performance of courses. Another element that must be taken into account is relative to synchronous interactions (chat or other real-time interaction): it is necessary to give priority to asynchronous activities (i.e., forum, homework) if the course is delivered in different countries distant in terms of time zone [35]. Visual, audio, and interactive contents can increase learning as well as to facilitate knowledge and satisfaction [22, 35, 45].

Some concerns persist for online degree programs about specific risks, due to the different methodology of control and progress of e-learning. According to the literature there is a overall need to establish a common understanding of e-learning in order to decide if use permanently in the academia elearning for nursing studies [28, 31, 34]. No statistical difference in terms of knowledge, skill, and satisfaction both for nursing (and health professional) students at undergraduate and postgraduate level emerged; in the light of that, there is an urgent need to develop robust quantitative instruments to measure the impact, effectiveness, and perceptions of students and educators [22, 28, 31, 41, 46].

### 17.3.9 Conclusion

This review clearly shows that currently there are multiple forms of e-learning in universities. However, no stronger evidence of best technologies/modality of e-learning exist currently in terms of impact on the acquisition of skills and knowledge for students and faculty. Common aspects that are key aspect for acceptance of e-learning are:

- Interactivity – necessity to synchronous and asynchronous interaction with other students and tutor/faculty
- Accessibility and flexibility – open access 24/7 is a key element for nurses and other healthcare professional
- Personalized feedback – encourage student involvement
- Tutoring – improving performance

The unresolved key issues are related to proper planning of activities, the specific training of faculty members, as well as the complexity of production of interactive-digital e-learning contents and practical laboratory activities. At the same time, specific control criteria should be defined for distance verification systems, comparing with the traditional systems.

The overview carried out shows that there is a lack of robust evidence on this topic and that the field is constantly under development. Nevertheless, the research analyzed displays that there is a reduction in the cost of education (in terms of management, for instance) as well as a more efficient management of the time for students and lecturers which reduce overall the economic effort afforded by universities and facilitate the management of education environment. Another key point is the overall need of guidelines and rules for knowledge assessment of e-learning students.

Beyond what was mentioned above, the main conclusion which can be drawn after this study is that further higher level research (i.e., RCT) is necessary in order to better understand and frame e-learning within nursing and healthcare profession, keeping in mind that e-learning is a very broad topic and that there are several e-learning tools; therefore, which tool can be used in which environment should be carefully analyzed.

### 17.3.10 Limitations and Strengths

The strength of this overview is to show that e-learning delivery is very heterogeneous, and this difference is certainly reflected in the literature, both for primary studies and reviews. In addition, the systematic and integrative reviews included in this study showed significant methodological differences in terms of analysis of population, e-learning methodology, and outcomes. Furthermore, this implies that ideally more in-depth consultation and systematization of the primary studies is required.

## References


1. International Council of Nurses. Definition of nursing. 2004. Available on http://www.icn.ch/who-we-are/icn-definition-of-nursing/.
2. American Nursing Association. What is nursing? 2010. Retrieved from http://www.nursingworld.org/EspeciallyForYou/What-is-Nursing.
3. Fitzpatrick JJ. Making a commitment to go public. Appl Nurs Res. 2004;17(4):223.
4. Roy C, Jones D. Nursing knowledge development and clinical practice. New York: Springer Publishing; 2007.
5. British Medical Association. Safe handover: safe patients. Guidance on clinical handover for clinicians and managers. Available at: bma.ac.uk.
6. Cohen MD, Hilligoss PB. The published literature on handoffs in hospitals: deficiencies identified in an extensive review. Qual Saf Health Care. 2010;19(6):493–7.
7. O'Connell B, MacDonald K, Kelly C. Nursing handover: it's time for a change. Contemp Nurse. 2008;30:2–11.
8. Bomba DT, Prakash R. A description of handover in Australian Public Hospital. Aust Health Rev. 2005;29(1):68–79.
9. Bost N, Crilly J, Wallis M, Patterson E, Chaboyer W. Clinical handover of patients arriving by ambulance to the emergency department–a literature review. Int Emerg Nurs. 2010;18(4):210–20.
10. Jensen SM, Lippert A, Østergaard D. Handover of patients: a topical review of ambulance crew to emergency department handover. Acta Anaesthesiol Scand. 2013;57(8):964–70.
11. Arora VM, Manjarrez E, Dressler DD, Basaviah P, Halasyamani L, Kripalani S. Hospitalist handoffs: a systematic review and task force recommendations. J Hosp Med. 2009;4(7):433–40.
12. Smeulers M, Lucas C, Vermeulen H. Effectiveness of different nursing handover styles for ensuring continuity of information in hospitalised patients. Cochrane Database Syst Rev. 2014;6:CD009979.
13. Gordon M, Findley R. Educational interventions to improve handover in health care: a systematic review. Med Educ. 2011;45(11):1081–9.
14. Riesenberg LA, Leisch J, Cunningham JM. Nursing handoffs: a systematic review of the literature. Am J Nurs. 2010;110(4):24–34.
15. Abraham J, Kannampallil T, Patel VL. A systematic review of the literature on the evaluation of handoff tools: implications for research and practice. J Am Med Inform Assoc. 2014;21:154–62.
16. Staggers N, Blaz JW. Research on nursing handoffs for medical and surgical settings: an integrative review. J Adv Nurs. 2013;69(2):247–62.
17. Collins SA, Stein DM, Vawdrey DK, Stetson PD, Bakken S. Content overlap in nurse and physician handoff artifacts and the potential role of electronic health records: a systematic review. J Biomed Inform. 2011;44(4):704–12.
18. Flemming D, Hübner U. How to improve change of shift handovers and collaborative grounding and what role does the electronic patient record system play? Results of a systematic literature review. Int J Med Inform. 2013;82(7):580–92.
19. Polit DF, Beck CT. Is there gender bias in nursing research? Res Nurs Health. 2008;31(5):417–27.
20. Foster S, Manser T. The effects of patient handoff characteristics on subsequent care: a systematic review and areas for future research. Acad Med. 2012;87(8):1105–24.
21. Kitson AL, Muntlin Athlin Å, Elliott J, Cant ML. What's my line? A narrative review and synthesis of the literature on Registered Nurses' communication behaviours between shifts. J Adv Nurs. 2014;70(6):1228–42.
22. Button D, Harrington A, Belan I. E-learning & information communication technology (ICT) in nursing education: a review of the literature. Nurse Educ Today. 2013. doi:10.1016/j.nedt.2013.05.002. S0260-6917(13)00165-2.

23. Finke A, Bicans J. E-learning systems content and architecture evolution. Paper presented at 16th international conference on information and software technologies. Kaunas, Lithuania 2010. p. 311–39.
24. European Union. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52001DC0172. Last accessed on 17 Aug 2015.
25. Ganino G. Immagini per la didattica. Metodologie e Tecnologie dell'audiovisivo digitale. Roma, Italy: Anicia; 2009.
26. American Nurses Association. Nursing: scope and standards of practice. 2nd ed. Silver Spring: American Nurses Association; 2010. Nursebooks.org, Silverspring, MD
27. Carroll C, Booth A, Papaioannou D, Sutton A, Wong R. UK health-care professionals' experience of on-line learning techniques: a systematic review of qualitative data. J Contin Educ Health Prof. 2009;29(4):235–41. doi:10.1002/chp.20041.
28. Lahti M, Hätönen H, Välimäki M. Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: a systematic review and meta-analysis. Int J Nurs Stud. 2013. doi:10.1016/j.ijnurstu.2012.12.017. S0020-7489(12)00459-2.
29. Petty J. Interactive, technology-enhanced self-regulated learning tools in healthcare education: a literature review. Nurse Educ Today. 2013;33(1):53–9. doi:10.1016/j.nedt.2012.06.008.
30. Rowe M, Frantz J, Bozalek V. The role of blended learning in the clinical education of healthcare students: a systematic review. Med Teach. 2012;34(4):e216–21. doi:10.3109/0142159X.2012.642831.
31. Feng JY, Chang YT, Chang HY, Erdley WS, Lin CH, Chang YJ. Systematic review of effectiveness of situated e-learning on medical and nursing education. Worldviews Evid Based Nurs. 2013;10(3):174–83. doi:10.1111/wvn.12005.
32. Cook DA, Garside S, Levinson AJ, Dupras DM, Montori VM. What do we mean by web-based learning? A systematic review of the variability of interventions. Med Educ. 2010;44(8):765–74. doi:10.1111/j.1365-2923.2010.03723.x.
33. Frehywot S, Vovides Y, Talib Z, Mikhail N, Ross H, Wohltjen H, Bedada S, Korhumel K, Koumare AK, Scott J. E-learning in medical education in resource constrained low- and middle-income countries. Hum Resour Health. 2013;11(1):4. doi:10.1186/1478-4491-11-4.
34. Mancuso JM. Perceptions of distance education among nursing faculty members in North America. Nurs Health Sci. 2009;11(2):194–205. doi:10.1111/j.1442-2018.2009.00456.x.
35. Booth A, Carroll C, Papaioannou D, Sutton A, Wong R. Applying findings from a systematic review of workplace-based e-learning: implications for health information professionals. Health Info Libr J. 2009;26(1):4–21. doi:10.1111/j.1471-1842.2008.00834.x.
36. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. Acad Med. 2010;85(5):909–22. doi:10.1097/ACM.0b013e3181d6c319.
37. Keyte D, Richardson C. Re-thinking pain educational strategies: pain a new model using e-learning and PBL. Nurse Educ Today. 2011;31(2):117–21. doi:10.1016/j.nedt.2010.05.001.
38. Hansen MM. Versatile, immersive, creative and dynamic virtual 3-D healthcare learning environments: a review of the literature. J Med Internet Res. 2008;10(3), e26. doi:10.2196/jmir.1051.
39. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg. 2012;99(10):1322–30. doi:10.1002/bjs.8819.
40. Wolbrink TA, Burns JP. Internet-based learning and applications for critical care medicine. J Intensive Care Med. 2012;27(5):322–32. doi:10.1177/0885066611429539.
41. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA. 2008;300(10):1181–96. doi:10.1001/jama.300.10.1181.
42. Du S, Liu Z, Liu S, Yin H, Xu G, Zhang H, Wang A. Web-based distance learning for nurse education: a systematic review. Int Nurs Rev. 2013;60(2):167–77. doi:10.1111/inr.12015.

43. Cook DA, Levinson AJ, Garside S. Time and learning efficiency in internet-based learning: a systematic review and meta-analysis. Adv Health Sci Educ Theory Pract. 2010;15(5):755–70. doi:10.1007/s10459-010-9231-x.
44. Wilkinson A, While AE, Roberts J. Measurement of information and communication technology experience and attitudes to e-learning of students in the healthcare professions: integrative review. J Adv Nurs. 2009;65(4):755–72. doi:10.1111/j.1365-2648.2008.04924.x.
45. Patterson BJ, Krouse AM, Roy L. Student outcomes of distance learning in nursing education: an integrative review. Comput Inform Nurs. 2012;30(9):475–88. doi:10.1097/NXN.0b013e3182573ad4.
46. Bloomfield JG, While AE, Roberts JD. Using computer assisted learning for clinical skills education in nursing: integrative review. J Adv Nurs. 2008;63(3):222–35. doi:10.1111/j.1365-2648.2008.04653.x.

# Case Study in Orthopedics

# 18

Maciej Płaszewski and Josette Bettany-Saltikov

**Abstract**

The aim of this chapter is to present a case study of why and how an umbrella review was conducted and reported within the context of the conservative management of adolescents with idiopathic scoliosis. We present and discuss the findings of an umbrella review of systematic reviews regarding both the screening and the effectiveness of nonsurgical interventions for adolescents with idiopathic scoliosis. The mandate for school-based screening programs for adolescents with scoliosis is a highly contentious and strongly debated issue. As a result of numerous controversial and discordant recommendations presented over a number of years by different societies and organizations that have been based on poor-quality systematic reviews, patients, therapists, insurance providers, policy makers, and researchers remain uncertain and confused. Nonsurgical interventions for adolescents with idiopathic scoliosis, including scoliosis-specific exercise interventions and bracing, are also controversial. We evaluated the validity as well as the methodological quality of all existing systematic reviews on this topic. We summarized, appraised, analyzed, and synthetized all available studies meeting the minimal criteria for a systematic review. The methodological quality of the majority of the included 27 systematic reviews, as assessed with the AMSTAR risk of bias tool, was disappointingly low. More, good-quality primary and secondary studies of

M. Płaszewski, PhD, MEd, PT (✉)
Institute of Physiotherapy, Biała Podlaska, University of Physical Education, Marymoncka 34, Warsaw 00-968, Poland
e-mail: maciej.plaszewski@awf-bp.edu.pl

J. Bettany-Saltikov, PhD, MSc, MCSP, PGC, THLE
School of Health and Social Care, University of Teesside, Middlesbrough, UK
e-mail: j.b.saltikov@tees.ac.uk



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_18

higher level designs are urgently needed in this topic. We demonstrated the role of an umbrella review, as a research tool, to prevent misleading information and erroneous guidance for stakeholders, resulting from poorly conducted systematic reviews.

## 18.1 Introduction

### 18.1.1 General Assumptions

The volume of knowledge currently generated in medical journals worldwide is phenomenal [1, 2]. At the same time the science of evidence synthesis has also developed rapidly [3–5]. Furthermore the importance of high-quality bodies of evidence together with the evolution of the evidence based from primary clinical trials to high-quality systematic reviews has significantly strengthened the quality of evidence currently available within most scientific and medical fields [6]. This overwhelming amount of knowledge and information both in quantitative and in qualitative areas has resulted in the need for a novel method of research – a systematic and umbrella review of systematic reviews.

This handbook elaborates on the methodology, developments, and challenges, as well as the dilemmas regarding the conduct and reporting of methodological issues concerning umbrella reviews. The aim of this chapter is to present a case study of why and how an umbrella review was conducted and reported within the context of a specific clinical research question. In this chapter we present a case study on the conservative management of adolescents with idiopathic scoliosis (AIS). We demonstrate how a careful methodological analysis of current existing systematic reviews can provide health-care professionals as well as service users engaged in the care of patients with AIS with new evidence and information regarding the credibility of published secondary research papers. We present and discuss the findings of an umbrella review of systematic reviews regarding both the screening for AIS and the effectiveness of nonsurgical interventions in adolescents with AIS.

### 18.1.2 Why Was this Umbrella Review Needed?

Screening programs are legitimate, provided that early treatment interventions for patients diagnosed with AIS as a result of screening are effective [7] – on the other hand, patients with false-positive diagnoses may undergo unnecessary and potentially harmful treatments. In the case of AIS, screening and nonsurgical treatment methods are both the subject of controversial opinions and published papers with discordant results from systematic reviews. These have resulted in contradictory recommendations and statements that have frequently lead to opinion-based practice guidelines being developed (Tables 18.1 and 18.2). The selection of this case was also prompted

by the description of the specific role of the umbrella review in informing evidence-based decision-making, highlighted by Cornell and Laine in their editorial [8]:

> (…) practice guidelines often require answers to not only one but several linked questions. A common example concerns screening interventions for which direct clinical trials that randomly assign participants to a screening or control group are lacking. In such cases, guideline developers typically seek answers to a sequence of interrelated questions: How prevalent is the condition? What is its natural history? Is a good screening test available? What are the benefits (and harms) of early detection and treatment? (…)

We performed an umbrella review to find answers to research questions regarding the reliability (for quantitative questions) and credibility (for qualitative questions) of all reviews addressing the screening as well as the nonsurgical management of a single condition – AIS, to inform policy makers and all engaged in the diagnosis, prevention, therapy, and education of people with this condition.

Systematic reviews, including those in the orthopedics [9], surgery [10], and physiotherapy [11] fields, vary in terms of their quality. Thus, another reason for this umbrella review was to assess the quality of the evidence from secondary studies, regarding the conservative management of AIS.

We aimed at evaluating the validity as well as the methodological quality of all existing systematic reviews on this topic. Our goal was also to examine which papers labeled as "systematic reviews" were in fact opinion pieces or narrative reviews.

### 18.1.3 Terminology and Methodological Issues: The Report as a Case Study within this Methods Handbook

Before we discuss the case study, we would like to address a number of methodological issues as well as issues concerning terminology characteristic for the comparatively new, developing, and evolving type of research synthesis methods: the umbrella review, which we believe are important in the context of the case report that we present.

Currently different terminologies are used to describe systematically developed reviews of systematic reviews (and, in some instances, also of primary studies). These terms include complex systematic reviews [12], overview of reviews [13, 14], systematic umbrella review [15], overview of systematic reviews [16, 17], umbrella systematic review [18, 19], analysis of systematic reviews [20], metaepidemiologic study [21], systematic review of systematic reviews [22], systematic review of systematic reviews and meta-analyses [23] or systematic review of meta-analyses [24], systematic map of systematic reviews [25], and even the term "survey" of reviews [26] that has been used. We will be using the term "umbrella review" throughout this chapter, even when referring to publications whose authors have used different terms for studies of this type.

The term "umbrella review" also has a number of different usages and understandings [13, 27–30]. We conducted an "umbrella systematic review of systematic reviews." We did not include primary studies for analysis. We conducted an umbrella

review, as elaborated within the Cochrane Handbook [13] and the Joanna Briggs Institute's (JBI) [29] guidelines. We also followed the guidelines described by Ioannidis [27] and Grant and Booth [30].

Umbrella reviews of systematic reviews are needed not only to summarize large bodies of evidence [27, 28]. Different types of research questions are addressed within umbrella reviews. These have included the quality of reporting methods [31], financial conflicts of interest of review authors [20], reporting the quality of search methods in systematic reviews [26], reporting publication bias [21], as well as the handling of missing outcome data [23]. To our best knowledge, we conducted the first comprehensive, explicit, and systematic overview of systematic reviews addressing screening and nonsurgical interventions for adolescents with idiopathic scoliosis.

The rigor for the selection of systematic reviews for inclusion within umbrella reviews in terms of the design and type of systematic reviews is increasing. Umbrella review of meta-analyses [32], umbrella review of systematic reviews and meta-analyses of observational studies and randomized trials [33], and umbrella review of meta-analyses of observational studies [34] have all been published. Although rare and unusual, quantitative data syntheses (meta-analysis) are also possible within umbrella reviews [35, 36]. In our case study, we present the more typical umbrella review, with a narrative describing the evidence from the systematic reviews of any types of primary studies that were included.

## 18.2 Methodology and Reporting

In contrast to systematic reviews of primary studies, umbrella reviews lack dedicated, published reporting guidelines. We followed both the PROSPERO protocol, which is universal for systematic reviews of primary studies and overviews of systematic reviews, and the PRISMA reporting guidelines [37], which though originally developed for systematic reviews of primary studies is also applicable to umbrella reviews [38].

### 18.2.1 The Case Study on Adolescent Idiopathic Scoliosis: Report of the Umbrella Review

To address the whole subject matter regarding the two interrelated issues – the screening programs and the treatment methods – we divided our umbrella review process into two parts. The first part discusses the systematic reviews on school screening for AIS. The second part presents the systematic reviews on nonsurgical interventions for AIS. Such approaches have recently been used in both a published combined report of an "umbrella review of systematic reviews" and "a systematic review of primary studies regarding conservative interventions for tendinopathy" [16]. Similar methodologies have also been reported in an umbrella review of systematic reviews and guideline documents addressing diet and physical activity interventions and policies [39].

Protocol registration and resource publications. This chapter is based on our study "Effects of nonsurgical management for patients with idiopathic scoliosis. An overview of systematic reviews" that has been registered at PROSPERO (CRD42013003538) and has previously been presented in two reports, regarding screening [40] and nonsurgical interventions [41].

## 18.2.2 Background

### 18.2.2.1 Description of the Health Problem

The prevalence of AIS is estimated at 2–4 % of children aged 10–16. The ratio of girls to boys ranges from 1:1for spinal lateral deviations (curve angles) below 10° to over 10:1 for curves exceeding 30°. The risk of progression of untreated scoliosis depends both on the skeletal maturity and the size of the curve. The tendency for progression is also more frequent among girls [42–45]. The deformity may have lasting consequences and can be accompanied with pulmonary complications (life endangering in patients with very severe curves), pain symptoms, difficulties in participation, and psychological disorders [42–46].

### 18.2.2.2 Screening and Nonsurgical Management

As a prevalent condition, with individual variations, affecting both the physical and psychosocial functioning in the long term, AIS is considered by numerous clinicians, researchers, and authors to be an "important healthcare problem meeting the requirements for a screening program" [47, 48].

Nonsurgical interventions for the treatment of adolescents with AIS typically constitute a variety of physical modalities: braces of various types and modes of application, scoliosis-specific exercises, as well as a number of diverse physical therapy modalities such as electrical stimulation, manual therapy, and different types of osteopathic and chiropractic interventions [49–51].

### 18.2.2.3 National and International Guidelines and Recommendations

**School screening** The mandate for school-based screening programs for adolescents with scoliosis is a highly controversial, strongly debated issue [52–54]. As a result of controversial and discordant recommendations, in the USA, screening has remained mandatory in some US States (Arkansas, Alabama, California, Florida, Pennsylvania, Texas, Utah) is recommended in Minnesota, is not required in Montana and Oregon, and was repealed in Indiana and Maryland. Table 18.1 summarizes the recommendations and position statements of different institutions. The table also classifies the available documents as either opinion-based statements or systematically developed practice guidelines.

**Conservative treatment** Nonsurgical interventions for adolescents with AIS are subject to similar controversies. To illustrate the polarity, examples of opinions from experts can be seen in Table 18.2.

**Table 18.1** Summary of current recommendations regarding school screening for scoliosis

| Developer/initiative (year)[a] | Recommendation | Type |
|---|---|---|
| Canadian Task Force on Preventive Health Care (1994) | "insufficient evidence (…) to indicate that screening for idiopathic scoliosis in adolescents is either effective or ineffective in improving the outcome" | EB |
| National Health and Medical Research Council, Australia (2002) | "Good evidence to recommend against screening" | EB |
| National Self-Detection Program for Scoliosis[b] (Spine Society of Australia, current as of April 2014) | Two-tier self-detection program for girls, replacing school screening programs | N |
| Italian guidelines[c] (2005) | "School screening programs (…) should be conducted"; "scientific evidence": E2 – fair scientific consensus | CB/EB |
| SRS-AAOS-POSNA-AAP[d] position statement (2008) | "Societies recognize the benefits that can be provided by effective clinical screening programs" and "do not support any formal recommendations against scoliosis screening, given the available literature" | CB |
| Society for Spinal Orthopaedic and Rehabilitation Treatment, SOSORT (2012) | School screening useful for clinical purposes; several improvements to the programs postulated | CB/EB |
| Ministry of Health Malaysia, HTA Section (2009) | "Screening for scoliosis among school children is recommended only for high risk group such as girls at 12 years or age"; fair level of evidence to suggest that school scoliosis screening program was able to detect scoliosis at a younger age and with smaller Cobb angle; [and] reduce the frequency of surgical treatment; evidence to suggest its cost-effectiveness | EB |
| UK National Screening Committee[e] (2012) | "Screening should not be offered," "systematic population screening programme not recommended," "(…) there is little evidence that screening would be necessary to pick up children needing surgery. (…) many children would be subjected to unnecessary X-rays and treatment, which may themselves be harmful. This could cause unnecessary stress to children and their families" | EB |
| Institute for Clinical Systems Improvement (2013) | Recommendations from USPSTF (2004) and SRS-AAOS-POSNA-AAP (2008) reported; level III service: "(…) the evidence is currently incomplete (…); providing these services is left to the judgment of individual medical groups, clinicians, and their patients"; the SRS-AAOS-POSNA-AAP position statement evaluated as "low quality evidence" | EB |

**Table 18.1** (continued)

| Developer/initiative (year)[a] | Recommendation | Type |
|---|---|---|
| Scoliosis Research Society International Task Force on Scoliosis Screening (2013) | "(…) an expert panel supports scoliosis screening in 4 of the 5 domains (…) of the World Health Organisation criteria for a valid screening procedure" | CB/EB |
| US Preventive Services Task Force/Agency for Healthcare Research and Quality (2004, 2012) | "Do not screen for idiopathic scoliosis"; recommendation against; grade D | EB (2004) |
| US Preventive Services Task Force/Agency for Healthcare Research and Quality (2014) | Idiopathic scoliosis not among recommendations | n/a |
| "Bright Futures" initiative[f] (current as of April 2015) | The Bright Futures/AAP periodicity schedule does not include screening for scoliosis | N |
| American Academy of Family Physicians (2015) | "The AAFP recommends against the routine screening of asymptomatic adolescents for idiopathic scoliosis (2004)"; grade D (USPSTF classification, prior to 2007)[f] | EB (USPSTF, 2004) |

*EB* evidence based, *CB* consensus based, *N* narrative describing a program or recommendations, *n/a* not applicable

[a]For references see resource publication [40]

[b]Endorsed by the Paediatrics and Child Health Division of the Royal Australasian College of Physicians

[c]Endorsed and approved by many Italian professional bodies, mandated by the Italian Ministry of Health

[d]Scoliosis Research Society, American Academy of Orthopedic Surgeons, Pediatric Orthopedic Society of North America, American Academy of Pediatrics; AAP endorsed the position statement; however, AAP also leads the "Bright Futures" initiative

[e]Approved by the British Orthopaedic Association, British Scoliosis Society, Institute of Child Health, Royal College of General Practitioners, Royal College of Surgeons, Scoliosis Association

[f]Launched under the leadership of the Maternal and Child Health Bureau of the Health Resources and Services Administration, led by AAP and partnered by numerous agencies, groups, and organizations [http://brightfutures.aap.org]

## 18.2.3 Methods

### 18.2.3.1 Inclusion and Exclusion Criteria

To deliver the most comprehensive evidence synthesis possible, we summarized, appraised, analyzed, and synthetized all available studies meeting the minimal criteria for a systematic review. Papers were considered as systematically developed reviews if they reported on methods to search, identify and select studies, and critically appraised relevant evidence [55]. We considered all systematic reviews with

**Table 18.2** Opinions regarding nonsurgical interventions for adolescents with idiopathic scoliosis

| |
|---|
| *Negative comments*: |
| "Time and common sense prevent me from discussing any other treatment modality than bracing' |
| "Treatment options for patients with scoliosis range from the unproven or harmful to the beneficial" |
| "Physical therapy, chiropractic care, biofeedback and electric stimulation have not been shown to alter the natural history of scoliosis" |
| "Patients should be aware of the absence of evidence for these [physiotherapy] treatments" |
| *Positive statements*: |
| "Bracing and spinal surgery have been proven to alter the natural history of curve progression" |
| "Exercise-based therapies, alone or in combination with orthopaedic approaches, are a logical approach to improve and maintain flexibility and function in patients at risk for pain, pulmonary dysfunction, and progression" |
| "The triad of out-patient physiotherapy, intensive in-patient rehabilitation and bracing has proven effective in conservative scoliosis treatment in central Europe" |

For references see resource publication [41]

meta-analyses as well as qualitative systematic reviews with descriptive syntheses of findings from individual studies. To address research questions other than effectiveness – risk of adverse effects, prevalence, and test accuracy – and to allow for the analysis of the methodological rigor of the systematic reviews that were included, we did not limit the inclusion criteria to systematic reviews of randomized controlled trials, but considered systematic reviews of any types of primary studies, including those of different types of non-randomized studies.

#### 18.2.3.2 Search Strategy

We prioritized databases and other resources and began searching the potentially more productive databases (this approach termed or known as "purposive searching" [5] is described as being more efficient and less time-consuming than the typical comprehensive search strategies applied within systematic reviews of primary studies) [56]. These databases included the following types: databases of systematic reviews, databases with separate indexing of systematic reviews, guideline registries, general bibliographic databases, and then websites of relevant institutions. Search strategies, keywords, and limits are detailed in the resource publications [40, 41], and their detailed reporting exceeds the volume of this chapter. To document the process of searching, Table 18.3 has been included and shows databases and other resources that were searched in the "screening" umbrella review [40] together with the order of the search. In the "interventions" review [41], we performed a similar search, but with details specific to the different research question.

#### 18.2.3.3 Study Selection

We independently conducted the searches as well as the initial selection of studies by their title and/or abstract. Full papers were then examined for eligibility. Disagreements were resolved by discussion. The combined PRISMA search flow for the selection of included studies for the two parts of the umbrella review is shown in Fig. 18.1.

**Table 18.3** Databases searched in the screening umbrella review together with the "purposive" [5] order of searching beginning with the potentially more productive databases [40]

| |
|---|
| *Databases of systematic reviews, guideline registries and databases with separate indexing of systematic reviews and guidelines:* |
| The Centre for Reviews and Dissemination databases – DARE, HTA, NHSEED, Cochrane Database of Systematic Reviews (CDSR), Joanna Briggs Institute, Campbell Library, Cochrane Effective Practice and Organisation of Care (EPOC) Group, the AHRQ databases and resource lists from USPSTF, AHRQ Evidence-based Practice Centers (EPC Reports) and National Guideline Clearinghouse, PEDro, INAHTA, and TRIP |
| *Websites of institutions*: |
| USPSTF, CTFPHC, NHMRC, UK Screening Portal/UK NSC Policy Database, Scottish Intercollegiate Guidelines Network (SIGN), National Institute for Health and Care Excellence, UK (NICE) |
| *General bibliographic databases*: |
| MEDLINE through PubMed, Web of Science, and SportDiscus through EBSCO and Google Scholar |
| *Gray literature*: |
| Registered protocols, reviews in progress, guidelines in development, and registered titles: |
| PROSPERO, CDSR, the USPSTF registry of the topics in progress, the CTFPHC protocols, HSR Project Database, NICE, AHRQ EPC Reports database (for the EPC Reports in Progress), HSRProj Database, the NHMRC website and the Systematic Review Data Repository (SRDR) database, the Conference Proceedings Citation Index – Science from the Web of Knowledge |

For abbreviations not explained here, see Tables 18.1 and18.5

#### 18.2.3.4 Scope of the Systematic Reviews

Inclusion and exclusion criteria, formulated to the two umbrella reviews according to the PEO (problem/population–exposure/issue–outcome) and PICO (problem/population–intervention–comparator/control–outcome) principles, respectively, are summarized in Table 18.4.

#### 18.2.3.5 Types of Outcomes

We analyzed and summarized both patient relevant and surrogate (or intermediate) outcomes of both the screening of patients with AIS as well as the effectiveness of nonsurgical interventions, taking into account qualitative and other not "numeric" issues. For us this is especially relevant as regards AIS – surrogate outcomes (e.g., curve angle, angle of trunk rotation, curve progression) are frequent in the available systematic reviews whereas patient-oriented ones (e.g., quality of life, body image) are less popular and frequent.

#### 18.2.3.6 Data Extraction, Methodological Quality, and Level of Evidence Assessment

##### Data Extraction

We independently extracted the data, using predefined data extraction forms. Discrepancies were resolved through discussion.

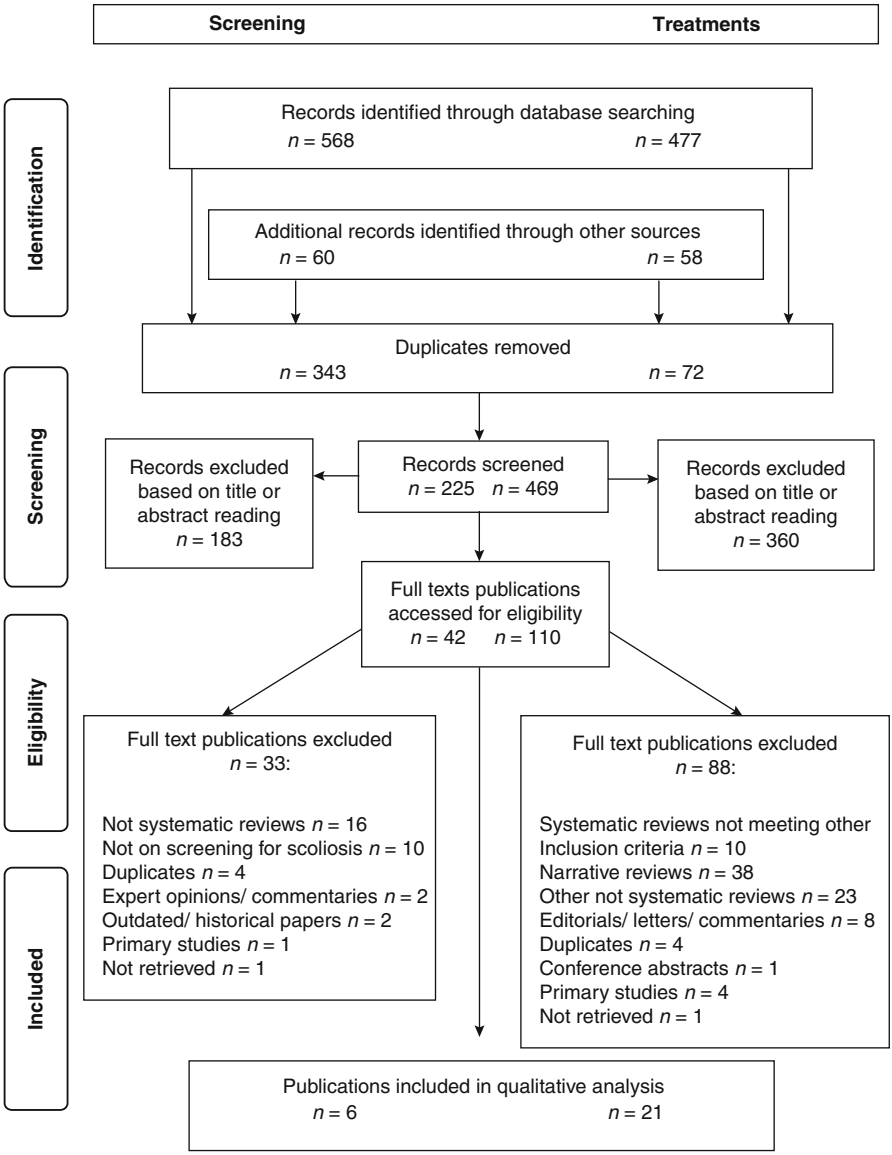



**Fig. 18.1** Compiled PRISMA flow diagrams for the selection of included systematic reviews for screening (*left*) and treatment methods (*right*)

## Methodological Quality

We used the "Assessment of Multiple Systematic Reviews" (AMSTAR) risk of bias tool [57] to assess the methodological quality of included reviews. The AMSTAR comprises 11 items addressing criteria relating to the assessment of methodological rigor (Table 18.1). The items are scored "yes," "no," "cannot answer," or "not

**Table 18.4** Criteria for inclusion and exclusion in the screening and intervention reviews

| | Inclusion criteria | Exclusion criteria |
|---|---|---|
| *Screening review:* | | |
| Population | Schoolchildren, both girls and boys, with no geographical or other (e.g., societal, racial, cultural) restrictions, within the growth spurt associated with the risk of development of AIS, typically 10–12 years of age; however, no strict age criteria were defined | Papers including other populations of children, e.g., with comorbidities, such as Duchenne muscular dystrophy |
| Exposure or issue | Any reviews addressing "school screening for AIS" | Papers addressing other screening programs that did not exclusively address screening tests for adolescent idiopathic scoliosis, e.g., general health examinations; screening programs for other types of scoliosis were also excluded (e.g., adult scoliosis) |
| Outcomes | Primary outcome: any recommendation that stated "to recommend screening or not to recommend screening, i.e., a "yes" or "no" with regard to the authors' recommendations; also any secondary outcomes | Not specified |
| *Intervention review:* | | |
| Population | Systematic reviews addressing adolescents of both genders with AIS, diagnosed and managed between the ages of 10–18 years, with no restriction as to bone age (Risser sign), with mild, moderate, and/or severe AIS (11–24°, 25–44°, and 45° Cobb and greater, respectively) | Reviews on-early-onset (infantile or juvenile) scoliosis, reporting on scoliosis secondary to other conditions, e.g., Duchene dystrophy, cerebral palsy, spinal cord injury, neurofibromatosis |
| Interventions | Nonsurgical interventions applied as a sole treatment or as combinations of different nonsurgical interventions: braces of any type (both rigid and soft) and mode of application (any number of hours a day or nighttime); any approach (s) or "school" of scoliosis-specific exercise treatment, regardless of the severity of the deformity, both as a single intervention and as part of a group of different complex interventions, e.g., supplementing brace treatment (add-on treatment); chiropractic; manual therapy; electrical stimulation; general conditioning (usual) exercises; any other nonsurgical interventions | Reviews on generalized and non-curve-specific exercises or other physiotherapeutic interventions administered to patients with AIS for other reasons, e.g., respiratory physiotherapy, spinal stabilization exercises, or electrical stimulation due to low back pain or leg pain; pre- or postoperative physiotherapeutic management of AIS patients; natural history or observation ("watchful waiting") as a form of therapy; reviews on screening, diagnostics, prognosis, economic analysis, or other research questions other than nonsurgical interventions |

(continued)

**Table 18.4** (continued)

| | Inclusion criteria | Exclusion criteria |
|---|---|---|
| Comparative interventions | Bracing, or scoliosis-specific exercises versus scoliosis-specific exercises plus other interventions, or different forms of these interventions (e.g., different modes of exercises or different types of braces); natural history or observation; other forms of nonsurgical interventions applied for scoliosis curve correction, e.g., chiropractic, manual therapy, electrical stimulation | Not specified |
| Outcomes | All short- and long-term outcomes that addressed the effectiveness and adverse effects of nonsurgical interventions; both patient-centered (e.g., pain, quality of life, depression, sense of stigmatization) and surrogate, secondary, or intermediate outcomes (e.g., curve progression, angle of trunk rotation, jaw deformity); the number of surgeries or the number needed to treat to avoid one surgery (need for surgery) as a criterion of failure of the nonsurgical interventions | Not specified |

applicable." The maximum score is 11. Scores 0–4, 5–8, and 9–11 indicate low-, moderate-, and high-quality reviews, respectively [58]. We conducted the appraisal independently. Exceptions were the Cochrane reviews [59, 60] that were included and coauthored by JB-S, when MP and a collaborator performed the independent appraisals. Assessments were conducted using guidelines for scoring AMSTAR questions [57, 58]. Disagreements were resolved by discussion.

### Levels of Evidence

We assessed the level of evidence from each included review, considering the types of studies included, using the new Oxford Centre for Evidence Based Medicine (OCEBM) [61, 62], the JBI [63, 64] classifications, and, for the screening reviews, additionally, the improved National Health and Medical Research Council (NHMRC) hierarchy [65]. We decided to apply more than one classification because different classifications vary as regards their content and characteristics, and this allowed us to assess the included papers more comprehensively. The NHMRC document proposed the only hierarchy corresponding to the characteristics of screening reviews, while the OCEBM and JBI classifications are more suitable for intervention reviews.

## 18.2.4 Results

### 18.2.4.1 Description of Included Reviews/Quantity of Research Available

For the screening review, six articles met the criteria for inclusion within the analysis from a total of 224 papers (see Fig. 18.1 and Table 18.1): two quantitative systematic reviews, one of which included a meta-analysis and four systematic analyses of evidence which were part of or supplementing recommendation documents. For the non-surgical interventions, from a total of 469 titles or titles and abstracts of papers, 21 papers met the criteria for inclusion: 18 systematic reviews addressed the effectiveness of different interventions, 1 review evaluated usual physical activity, and 2 systematic reviews addressed the side effects in braced patients. Overall the reviews addressed numerous, patient-centered, and surrogate short- and long-term outcomes.

### 18.2.4.2 Methodological Quality of Included Reviews/Quality Assessment

Overall the quality of systematic reviews regarding screening ranged from the comparatively recent (2009) moderate-quality (AMSTAR score 6) (Fong et al. [66] and Sabirin et al. [67]) through to the outdated (2002) moderate-quality NHMRC review [68] to the poor-quality recent (2011) UK NSC [46] and the outdated 2004 USPSTF review [69], that is nonetheless still used for current and recent recommendations [52, 70]. The quality of the intervention reviews ranged from low methodological quality to high quality. Only two of the included reviews were of high quality [61, 62], while three were of moderate quality [72–74], and the remaining 16 reviews were found to be of low methodological quality [75–90] (Table 18.5).

### 18.2.4.3 Levels of Evidence of the Included Reviews

The six reviews relating to screening that we analyzed matched neither the improved NHMRC nor the new OCEBM levels of evidence hierarchy, with the exception of Fong et al. [66], which can be classified as a level 3 evidence in the OCEBM classification (Table 18.2). The levels of evidence from the reviews on interventions ranged from 1+ to 4, with some reviews not matching the OCEBM and the JBI hierarchies. The classification of the levels of evidence depended mainly on the type of included individual studies and also on the specific level of evidence hierarchy applied (Table 18.6).

## 18.2.5 Discussion

Detailed narrative characteristics of the content of the reviews that were included in this umbrella reviews are out with both the volume and the scope of this chapter. Therefore our report was limited to a short description of the quantity and quality of the evidence from the systematic reviews that were included and the characteristics

**Table 18.5** AMSTAR ratings for reviews included in the quality analysis

| Paper (year) [reference] | AMSTAR questions[a] | | | | | | | | | | | Total Yes | Overall quality[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | | |
| *Screening*: | | | | | | | | | | | | | |
| NHMRC (2002) [68] | Y | CA | Y | Y | N | Y | Y | N | NA | N | N | 5 | Moderate |
| USPSTF (2004) [69] | Y | CA | N | N | N | Y | N | N | NA | N | N | 2 | Low |
| Negrini et al. (2005) [71] | Y | CA | N | Y | N | N | N | N | NA | N | N | 2 | Low |
| Sabirin et al., MaHTAS (2010) [67] | Y | Y | Y | Y | Y | Y | CA | N | NA | N | N | 6 | Moderate |
| Fong et al. (2010) [66] | Y | Y | Y | N | N | Y | N | N | Y | Y | N | 6 | Moderate |
| UK NSC (2011) [47] | Y | CA | Y | Y | N | Y | N | N | NA | N | N | 4 | Low |
| *Exercise treatments*: | | | | | | | | | | | | | |
| Negrini et al. (2003) [75] | N | CA | Y | N | N | Y | Y | Y | N | N | N | 4 | Low |
| Negrini et al. (2008) [74] | Y | CA | Y | N | N | Y | Y | Y | Y | N | N | 6 | Moderate |
| Fusco et al. (2011) [76] | Y | CA | N | N | N | Y | N | N | N | N | N | 2 | Low |
| Mordecai and Dabke (2012) [77] | N | CA | N | N | N | Y | N | N | N | N | N | 1 | Low |
| Romano et al. (2012) [60] | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | N | 9 | High |
| *Manual therapy*: | | | | | | | | | | | | | |
| Romano and Negrini (2008) [78] | N | CA | Y | N | N | Y | N | N | N | N | N | 2 | Low |
| Gleberzon et al. (2012) [79] | N | CA | Y | N | N | Y | Y | N | NA | N | N | 3 | Low |
| McKennedy et al. (2013) [80] | N | Y | N | N | N | Y | Y | Y | N | N | N | 4 | Low |
| Posadzki et al. (2013) [72] | N | Y | Y | N | N | Y | Y | Y | Y | N | Y | 7 | Moderate |
| *Bracing*: | | | | | | | | | | | | | |
| Dolan and Weinstein (2007) [81] | N | N | Y | N | N | Y | N | N | N | N | N | 2 | Low |
| Negrini et al. (2010) [59] | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | N | 9 | High |

**Table 18.5** (continued)

| Paper (year) [reference] | AMSTAR questions[a] | | | | | | | | | | | Total Yes | Overall quality[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | | |
| Maruyama et al. (2011) [82] | N | CA | N | N | N | Y | Y | Y | N | N | N | 3 | Low |
| Davies et al. (2011) [83] | N | CA | CA | N | N | Y | N | N | Y | N | N | 2 | Low |
| Sanders et al. (2012) [84] | N | CA | N | N | N | N | N | N | N | N | N | 0 | Low |
| *Different combinations of nonsurgical interventions*: | | | | | | | | | | | | | |
| Focarile et al. (1991) [85] | N | Y | N | N | N | Y | Y | Y | N | N | N | 4 | Low |
| Rowe et al. (1997) [86] | N | CA | N | N | Y | N | N | N | Y | N | N | 2 | Low |
| Lenssinck et al. (2005) [73] | N | Y | N | N | N | Y | Y | Y | Y | N | N | 5 | Moderate |
| Weiss and Goodall (2008) [87] | N | N | Y | N | N | N | N | N | N | N | N | 1 | Low |
| *Usual physical activity:* | | | | | | | | | | | | | |
| Green et al. (2009) [88] | N | Y | Y | Y | N | Y | N | N | N | N | N | 4 | Low |
| *Adverse effects*: | | | | | | | | | | | | | |
| Li et al. (2008) [89] | N | CA | Y | N | N | Y | N | N | Y | N | N | 3 | Low |
| Saccucci et al. (2011) [90] | N | Y | Y | N | N | N | N | N | N | N | N | 2 | Low |

*Y* yes, *N* no, *CA* cannot answer, *NA* not applicable, *NHMRC* National Health and Medical Research Council, Australia, *USPSTF* US Preventive Services Task Force, *MaHTAS* Health Technology Assessment Section, Ministry of Health Malaysia, *UK NSC* UK National Screening Committee

[a]Questions [55, 56]: "1. Was an a priori design provided?, 2. Was there duplicate study selection and data extraction?, 3. Was a comprehensive literature search performed?, 4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?, 5. Was a list of studies (included and excluded) provided? 6. Were the characteristics of the included studies provided?, 7. Was the scientific quality of the included studies assessed and documented?, 8. Was the scientific quality of the included studies used appropriately in formulating conclusions?, 9. Were the methods used to combine the findings of studies appropriate?, 10. Was the likelihood of publication bias assessed?, 11. Were potential conflicts of interest included?"

[b]Review quality scores [55, 56]: 0–4 low quality, 5–8 moderate quality, and 9–11 high quality

of the evidence available from the most recent and methodologically sound reviews. All relevant detailed information, comprising all the included papers as well as all the excluded papers, can be found in the original publications [40, 41].

#### 18.2.5.1 Brief Summary of Evidence from Included Reviews

The evidence from included higher quality reviews is summarized in Table 18.2 according to the type of management (screening and nonsurgical interventions) and

**Table 18.6** Evidence from higher quality, more recent systematic reviews on screening and non-surgical interventions in AIS

| Title (year) [reference] | Findings/conclusions | Level of evidence [OCEBM/JBI] | AMSTAR score[b]/overall quality |
|---|---|---|---|
| Screening tests and programs: | | | |
| A meta-analysis of the clinical effectiveness of school scoliosis screening (2010) [66] | Only 17 % of the primary studies included within the meta-analysis of retrospective cohort studies found screening to be ineffective; the authors advocate for school screening, but recommended that the forward bend test should not be used alone within screening programs and that large, retrospective cohort studies are needed | 3[a] | 6/moderate |
| Exercise treatments: | | | |
| Exercises for adolescent idiopathic scoliosis (2012) [60], Cochrane review | "Due to a lack of high quality RCTs in this area, there is no evidence for or against exercises, so hardly any recommendations can be given"; "no major risks of the intervention have been reported (…), and no side effects were cited in the considered studies" | 1/1a | 9/high |
| Exercises reduce the progression rate of adolescent idiopathic scoliosis: results of a comprehensive systematic review of the literature (2008) [74] | "Exercises can be recommended according to level-1b evidence with the aim of reducing scoliosis progression"; "it is impossible to state anything regarding the kind of exercises .. [or] ..kind of auto-correction to be performed" | 3/1b | 5/moderate |
| Manual therapy: | | | |
| Osteopathic manipulative treatment for pediatric conditions: a systematic review (2013) [72] | Findings from the AIS RCT: no evidence to support OMT as an effective treatment of mild AIS; the study assessed as high-quality RCT; "more robust RCTs are needed (…). Until such data are available, OMT cannot be regarded as effective therapy for paediatric conditions, and osteopaths should not claim otherwise" | 1/1a | 7/moderate |
| Bracing: | | | |
| Braces for idiopathic scoliosis in adolescents (2010) [59], Cochrane review | Very low quality of evidence in favor of bracing in terms of curve progression; low evidence in favor of hard bracing vs elastic bracing; serious side effects not documented in the included studies | 1/1a | 9/high3 |

**Table 18.6** (continued)

| Title (year) [reference] | Findings/conclusions | Level of evidence [OCEBM/JBI] | AMSTAR score[b]/overall quality |
|---|---|---|---|
| Different combinations of nonsurgical interventions: | | | |
| Effect of bracing and other conservative interventions in the treatment of idiopathic scoliosis in adolescents: a systematic review of clinical trials (2005) [73] | "Effectiveness of bracing and exercises is promising but not yet established"; limited evidence for the effectiveness of braces vs no treatment and vs electrical stimulation (ES); bracing, exercises, or ES as add-on treatment – additional effect cannot be justified; no difference for ES vs no treatment, bracing vs exercises, different types of bracing | 1/1b | 5/moderate |

[a]Matched the OCEBM classification only

[b]Details are in Table 18.5

in the order of descending levels of evidence. The table does not include any studies on "usual physical activity" and adverse events [88–90] nor some highly cited older systematic reviews [81, 86] as all those reviews were of low or very low quality.

#### 18.2.5.2 Screening

The screening reviews were heterogeneous, both with regard to the research questions asked and the methodology used for their development. The reviews which supported the recommendation for school screening as well as those recommending against screening are based on different papers selected for inclusion and thus on different evidences or assumptions of the absence of evidence (Fig. 18.2). Conclusions were based on different criteria as follows: the set of criteria for appraising – feasibility, effectiveness, and appropriateness of screening programs, accuracy of screening tests, treatment effectiveness as a criterion justifying the need for screening, and cost-effectiveness. Three of the systematic reviews were found to be of moderate quality; Fong et al. 's systematic review of retrospective cohort studies with a meta-analysis [66] as well as MaHTAS systematic review by Sabirin et al. (2010) [67] supported screening under certain conditions. The NHMRC document [68], which included a recommendation against screening, was also found to be of moderate quality.

#### 18.2.5.3 Scoliosis-Specific Exercises (SSE)

The most recent of the available reviews was the rigorous Cochrane review [60]. This provided no convincing evidence from RCTs for or against these interventions in terms of curve progression as a primary outcome and no evidence of risks or side effects from performing scoliosis-specific exercises. A moderate-quality review by Negrini et al. [74] recommended the use of SSE exercises based on primary studies classified by the authors as level 1b evidence.

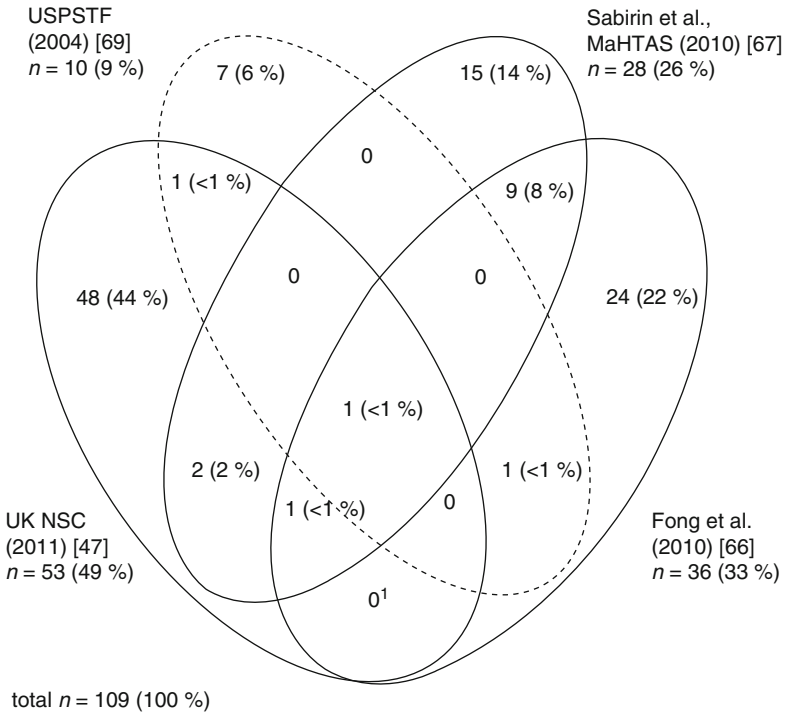



**Fig. 18.2** Example Venn diagram showing overlaps across included systematic reviews: overlaps of papers included in four screening systematic reviews. *Numbers* and *percentages inside the ellipses* show the overlap of studies included in the four reviews. *Percentages outside the ellipses* illustrate the percentage of all 109 studies included in each of the four reviews. *n* number of papers included in the reviews, *USPSTF* the US Preventive Services Task Force, *MaHTAS* Health Technology Assessment Section, Ministry of Health Malaysia, *UK NSC* the UK National Screening Committee; this review also included the systematic review by Fong et al.

#### 18.2.5.4 Manual Therapies

A recent good-quality systematic review of controlled studies [72] found one high-quality RCT showing no evidence to support osteopathic manual therapy as an effective treatment for mild AIS.

#### 18.2.5.5 Bracing

A rigorous Cochrane review [59] found very low-quality evidence supporting the effectiveness of bracing in reducing curve progression and low-quality evidence favoring hard braces as compared to soft braces. In an earlier, moderate-quality systematic review of prospective controlled trials, Lennsinck et al. [73] concluded that due to the low-power, weak methodological quality, and clinical heterogeneity of the included studies, drawing firm conclusions was impossible. However the effectiveness of bracing and SSE treatments in reducing curve progression appeared to the authors to be promising.

#### 18.2.5.6 Quality Analyses

The methodological quality of the majority (19 out of 27; see Table 18.5) of the systematic reviews that were retrieved was disappointingly low (Table 18.5), regardless of the limitations of the primary studies included in the reviews. Common errors included the following: no second independent reviewer and blind study selection and/or data extraction, no lists of included and excluded studies, no comprehensive search for evidence, and, perhaps most importantly, no quality assessment of included studies conducted. This is crucial as if the studies in question were of poor quality. then we should not be basing clinical practice or recommendations on the results of these reviews.

The screening reviews also differed significantly with regard to the databases selected and other resources searched. Moreover, significant heterogeneity was found within the reviews as follows: different research designs were considered (prospective trials and retrospective observational studies, systematic reviews, editorials), which were, except in the Fong et al. meta-analysis [66], analyzed separately.

In some of the reviews, the level of evidence hierarchy classification (categories of studies) was reported as a quality assessment. Further, a number of excluded reviews (listed in detail, with reasons for exclusion, as supplementary files to the resource publications) were called "systematic" but actually comprised only a structured and systematic literature search and then presented as a narrative discussion of a few papers of diverse designs. The only intervention systematic review with a meta-analysis by Rowe et al. [86] was seriously flawed methodologically (Table 18.6). Further and crucially the patient group was not homogenous and did not differentiate between juvenile and adolescent IS.

#### 18.2.5.7 Limitations of the Study

Firm conclusions cannot be drawn from this umbrella review as it cannot clearly be established from the individual systematic reviews included that the interventions tested may differ significantly from each other within similar papers even if they have the same label (e.g., bracing – there are considerable differences in the construction, biomechanical principles of action, as well as the length of time worn [49]). The same applies to scoliosis-specific exercise treatment with at least six different schools of thought and approaches available [50].

Secondly, we were not able to firmly distinguish the methodological quality of the process for conducting systematic reviews from the quality of reporting of the reviews that we included and analyzed. The AMSTAR tool does not clearly distinguish between the two, and we did not utilize any measure of the reporting quality of the included reviews (such as PRISMA). It is also important to consider that this overview will need an update every few years to reflect recently published work. This could be undertaken by using another (or modified) appraisal tool rather than the AMSTAR tool (as the appraisal tool itself may influence the findings from an overview of systematic reviews [91]). The reliability and validity of the findings of umbrella reviews conducted with the use of different versions of this tool requires further studies [92].

### 18.2.6 Conclusions: What this Study Adds

#### 18.2.6.1 Methodological Considerations

When many systematic reviews exist about a given topic, it is critical that the analysis of the methodological quality (i.e., the rigor for the development) of the systematic reviews and – consequently – their credibility are fully assessed as well. Systematic reviews should not only be considered as the base for an umbrella review that summarizes and synthetizes the findings from the currently available systematic reviews [29]. In this scenario an umbrella review's role is not only to summarize or synthetize but also to critically appraise, analyze, and assess the limitations of the systematic reviews that are currently available [13]. The role of an umbrella review is very important as poorly conducted systematic reviews may mislead and provide erroneous guidance for stakeholders – patients, therapists, insurance providers, policy makers, and researchers.

#### 18.2.6.2 Conclusions of Case Study Issue

In this chapter we have explained why an umbrella review undertaking a best evidence synthesis approach was urgently needed for the school screening and nonsurgical treatment of adolescents with idiopathic scoliosis. We have described how we performed the two reviews and illustrated how umbrella reviews may be useful in informing end users in avoiding the misinterpretation of the available evidence from reviews of various quality and credibility.

#### 18.2.6.3 Implications for Practice

A recent Cochrane Collaboration's analysis of the process of guideline development has revealed the underutilization of systematic reviews and meta-analyses in developing practice guidelines [93]. The results of our study can aid policy makers and guideline developers in producing better evidence-informed, up-to-date guidelines, both for the screening of schoolchildren in the risk groups of the development of AIS and for the nonsurgical treatment of adolescents with idiopathic scoliosis.

#### 18.2.6.4 Conclusions and Implications for Research

In conclusion, good-quality primary studies of higher level designs are urgently needed in the areas of screening and conservative methods for the treatment of adolescents with idiopathic scoliosis. Further developments in the conduct of systematic reviews, especially – in the case of the subject of this chapter – using multiple types of studies in systematic reviews [94], will hopefully facilitate finding the right answers to the complex and diverse research questions currently found in research, practice, and policy, through conducting more valid and reliable umbrella reviews based on the greater trustworthiness currently found within systematic reviews.

## 18.3 Additional Details

We would like to thank Dr. Igor Cieśliński for his contribution to the AMSTAR assessment of the included Cochrane reviews. MP registered the protocol, conceived, and designed the experiments and prepared data extraction tables. MP and JB-S performed the experiments, analyzed the data, and contributed to the writing of the manuscript. The review followed the protocol with the exception of some details of database searching (additional specialty websites and guideline registries, instead of AMED, CINAHL, and EMBASE databases searches were conducted), as the "productivity scheme" of searching was preferred to the comprehensive search strategy [5, 57].

## References

1. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7:e1000326.
2. Shadish WR, Lecy JD. The meta-analytic big bang. Res Synth Methods. 2015. doi:10.1002/jrsm.1132.
3. Moher D, Tetzlaff J, Taricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med. 2007;4:e78.
4. Chandler J, Hopewell S. Cochrane methods–twenty years' experience in developing systematic review methods. Syst Rev. 2013;2:76.
5. Petticrew M. Time to rethink the systematic review catechism? Moving from 'what works' to 'what happens'. Syst Rev. 2015;4:36.
6. Gartlehner G, Dobrescu A, Swinson Evans T, Thaler K, Nussbaumer B, Sommer I, Lohr KN. Comparison of effects as evidence evolves from single trials to high-quality bodies of evidence. Research White Paper. (Prepared by the RTI–UNC Evidence-based Practice Center under Contract No. 290-2012-00008-I.) AHRQ Publication No. 15-EHC010-EF. Rockville: Agency for Healthcare Research and Quality; 2015.
7. Barratt A, Glasziou P, Cumming RC, Raffle A, Hicks N, Muir Gray JA, Guyatt GH, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature XVII. How to use guidelines and recommendations about screening. JAMA. 1999;281:2029–34.
8. Cornell JE, Laine C. The science and art of deduction: complex systematic overviews (editorial). Ann Intern Med. 2008;148:786–8.
9. Garnier JL, Kellam PJ. Reporting and methodological quality of systematic reviews in the orthopaedic literature. J Bone Joint Surg Am. 2013;95:e771–7.
10. Adie S, Ma D, Harris IA, Naylor JM, Craig JC. Quality of conduct and reporting of meta-analyses of surgical interventions. Ann Surg. 2015;261:685–94.
11. Gianola S, Gasparini M, Agostini M, Castellini G, Corbetta D, Gozzer P, Li LC, Sirtori V, Taricco M, Tetzlaff JM, Turolla A, Moher D, Moja L. Survey of the reporting characteristics of systematic reviews in rehabilitation. Phys Ther. 2013;93:1456–66.
12. Whitlock E, Lin J, Chou R, et al. Using existing systematic reviews in complex systematic reviews. Ann Intern Med. 2008;148:776–82.
13. Becker LA, Oxman AD. Chapter 22. Overviews of reviews. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011. Available from www.cochrane-handbook.org.
14. Biondi-Zoccai G, Lotrionte M, Landoni G, Modena MG. The rough guide to systematic reviews and metaanalyses. HSR Proc Intensive Care Cardiovasc Anesth. 2011;3:161–73.

15. Safron M, Cislak A, Gaspar T, Luszczynska A. Effects of school-based interventions targeting obesity-related behaviors and body weight change: a systematic umbrella review. Behav Med. 2011;37(1):15–25.
16. Long L, Briscoe S, Cooper C, Hyde C, Crathorne L. What is the clinical effectiveness and cost-effectiveness of conservative interventions for tendinopathy? An overview of systematic reviews of clinical effectiveness and systematic review of economic evaluations. Health Technol Assess. 2015;19(8):1–134.
17. Li L, Tian J, Tian H, Sun R, Liu R, Liu Y, Yang K. Quality and transparency of overviews of systematic reviews. J Evid Based Med. 2012;5:166–73.
18. Domecq JP, Prutsky G, Wang Z, Elraiyah T, Brito JP, Mauck K, Lababidi MH, Leppin A, Fidahussein S, Prokop LJ, Montori VM, Murad MH. Drugs commonly associated with weight change: umbrella systematic review and meta-analysis (Protocol). Syst Rev. 2012;1:44.
19. Bidonde J, Jean Bush A, Bath B, Milosavljevic S. Exercise for adults with fibromyalgia: an umbrella systematic review with synthesis of best evidence. Curr Rheumatol Rev. 2014;10:45–79.
20. Dunn AG, Arachi D, Hudgins J, Tsafnat G, Coiera E, Bourgeois FT. Financial conflicts of interest and conclusions about neuraminidase inhibitors for influenza. An analysis of systematic reviews. Ann Intern Med. 2014;161:513–8.
21. Onishi A, Furukawa TA. Publication bias is underreported in systematic reviews published in high-impact-factor journals: metaepidemiologic study. J Clin Epidemiol. 2014;67:1320–6.
22. Caird J, Sutcliffe K, Kwan I, Dickson K, Thomas J. Mediating policy-relevant evidence at speed: are systematic reviews of systematic reviews a useful approach? Evid Policy J Res Debate Pract. 2015;11(1):81–97. -01-26T00:00:00.
23. Spineli LM, Pandis N, Salanti G. Reporting and handling missing outcome data in mental health: a systematic review of Cochrane systematic reviews and meta-analyses. Res Synth Methods. 2015. doi:10.1002/jrsm.1131.
24. Osnabrugge RL, Head SJ, Zijlstra F, Ten Berg JM, Hunink MG, Kappetein AP, et al. A systematic review and critical assessment of 11 discordant meta-analyses on reduced-function CYP2C19 genotype and risk of adverse clinical outcomes in clopidogrel users. Genet Med. 2015;17:3–11.
25. Mejàre IA, Klingberg G, Mowafi FK, Stecksén-Blicks C, Twetman SH, Tranæus SH. A systematic map of systematic reviews in pediatric dentistry – what do we really know? PLoS One. 2015;10(2):e0117537.
26. Mullins MM, DeLuca JB, Crepaz N, Lynes CM. Reporting quality of search methods in systematic reviews of HIV behavioral interventions (2000–2010): are the searches clearly explained, systematic and reproducible? Res Synth Methods. 2014;5:116–30.
27. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ. 2009;181:488–93.
28. Pieper D, Antoine S-L, Morfeld J, Mathes T, Eikermann M. Methodological approaches in conducting overviews: current state in HTA agencies. Res Synth Methods. 2014;5:187–99.
29. The Joanna Briggs Institute. Joanna Briggs Institute reviewers' manual: 2014 edition/ Supplement. Methodology for JBI umbrella reviews. Adelaide, Australia: The Joanna Briggs Institute; 2014.
30. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associates methodologies. Health Info Libr J. 2009;26:91–108.
31. Hutton B, Salanti G, Chaimani A, Caldwell DM, Schmid C, Thorlund K, Mills E, Catalá-López F, Turner L, Altman DG, Moher D. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. PLoS One. 2014;9(3):e92508.
32. Ioannidis JPA, Zhou Y, Chang CQ, Schully SD, Khoury MJ, Freedman AN. Potential increased risk of cancer from commonly used medications: an umbrella review of meta-analyses. Ann Oncol. 2014;25:16–23.
33. Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JPA. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomized trials. BMJ. 2014;348:g2035.

34. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JPA. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ. 2015;350:g7607.
35. Katerndahl DA, Lawler WR. Variability in meta-analytic results concerning the value of cholesterol reduction in coronary heart disease: a meta-meta-analysis. Am J Epidemiol. 1998;149:429–41.
36. Hemming K, Bowater RJ, Lilford RJ. Pooling systematic reviews of systematic reviews: a Bayesian panoramic meta-analysis. Stat Med. 2012;31:201–16.
37. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(6):e1000097.
38. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7:e49667.
39. Horodyska K, Luszczyska A, van der Berg M, Hendriksen M, Roos G, De Bourdeaudhuij I, Brug J. Good practice characteristics of diet and physical activity interventions and policies: an umbrella review. BMC Public Health. 2015;15:19.
40. Płaszewski M, Bettany-Saltikov J. Are current scoliosis school screening recommendations evidence-based and up to date? A best evidence synthesis umbrella review. Eur Spine J. 2014;23:2572–83.
41. Płaszewski M, Bettany-Saltikov J. Non-surgical interventions for adolescents with idiopathic scoliosis: an overview of systematic reviews. PLoS One. 2014;9:e110254.
42. Weinstein SL, Dolan LA, Cheng JCY, Danielson A, Morcuende JA. Adolescent idiopathic scoliosis. Lancet. 2008;371:1527–37.
43. Altaf F, Gibson A, Dannawi Z, Noordeen H. Adolescent idiopathic scoliosis. BMJ. 2013;346:f2508.
44. Hresko MT. Idiopathic scoliosis in adolescents. N Engl J Med. 2013;368:834–41.
45. Stokes OM, Luk KD. The current status of bracing for patients with adolescent idiopathic scoliosis. Bone Joint J. 2013;95-B:1308–16.
46. Tones M, Moss N, Polly DW. A review of quality of life and psychosocial issues in scoliosis. Spine. 2006;31:3027–38.
47. [No authors listed]. Screening for adolescent scoliosis. External review against programme appraisal criteria for the UK National Screening Committee (UK NSC). Bazian Ltd.; 2011. http://www.screening.nhs.uk/scoliosis. Accessed 11 June 2013.
48. Labelle H, Richards SB, De Kleuver M, Grivas TB, Luk KD, Wong HK, Thometz J, Beauséjour M, Turgeon J, Fong DY. Screening for adolescent idiopathic scoliosis: an information statement by the scoliosis research society international task force. Scoliosis. 2013;8:17.
49. Zaina F, De Mauroy JC, Grivas T, Hresko MT, Kotwicki T, et al. Bracing for scoliosis in 2014: state of the art. Eur J Phys Rehabil Med. 2014;50:93–110.
50. Bettany-Saltikov J, Parent E, Romano M, Villagrasa M, Negrini S. Physiotherapeutic scoliosis-specific exercises for adolescents with idiopathic scoliosis. Eur J Phys Rehabil Med. 2014; 50:111–21.
51. Romano M, Negrini S. Manual therapy as a conservative treatment for adolescent idiopathic scoliosis. Scoliosis. 2008;3:2. http://www.scoliosisjournal.com/content/3/1/2.
52. Plaszewski M, Nowobilski R, Kowalski P, Cieslinski M. Screening for scoliosis: different countries' perspectives and evidence-based health care. Int J Rehabil Res. 2012;35:13–9.
53. Linker B. A dangerous curve. The role of history in America's scoliosis screening programs. Am J Public Health. 2012;102:1106–16.
54. Fong DYT, Luk KDK, Cheung KMC, Wong Y-W. Scoliosis screening for school children. In: McPehrson B, Driscoll CJ, editors. School screening systems. New York: Nova Science Publishers Inc.; 2014. p. 197–217.
55. [No authors listed]. Glossary of terms in The Cochrane Collaboration. Version 4.2.5. London: The Cochrane Collaboration; 2005.
56. Booth A. How much searching is enough? Comprehensive versus optimal retrieval for technology assessment. Int J Technol Assess Health Care. 2010;26:431–5.

57. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
58. Popovic I, Windsor B, Jordan V, Showell M, Shea B, Farquhar CM. Methodological quality of systematic reviews in subfertility: a comparison of two different approaches. PLoS One. 2012;7(12):e50403.
59. Negrini S, Minozzi S, Bettany-Saltikov J, Zaina F, Chockalingam N, et al. Braces for idiopathic scoliosis in adolescents. Cochrane Database Syst Rev. 2010;(1):CD006850.
60. Romano M, Minozzi S, Bettany-Saltikov J, Zaina F, Chockalingam N, et al. Exercises for adolescent idiopathic scoliosis. Cochrane Database Syst Rev. 2012;8:CD007837.
61. OCEBM Levels of Evidence Working Group. The Oxford 2011 levels of evidence. Oxford Centre for Evidence-Based Medicine. Available http://www.cebm.net/index.aspx?o=5653. Accessed 15 Mar 2014.
62. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, Moschetti I, Phillips B, Thornton H. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) levels of evidence (Background Document). Oxford Centre for Evidence-Based Medicine. Available: http://www.cebm.net/index.aspx?o=5653. Accessed 15 Mar 2014.
63. The Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Working Party. New JBI levels of evidence. The Joanna Briggs Institute 2014. Available: http://joannabriggs.org/assets/docs/approach/JBI-Levels-of-evidence_2014.pdf. Accessed 15 Mar 2014.
64. The Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Working Party. Supporting document for the Joanna Briggs Institute levels of evidence and grades of recommendation. The Joanna Briggs Institute; 2014. Available: http://joannabriggs.org/assets/docs/approach/Levels-of-Evidence-SupportingDocuments.pdf. Accessed 15 Mar 2014.
65. National Health and Medical Research Council. NHMRC levels of evidence and grades of recommendations for developers of guidelines. NHMRC; 2009. http://www.nhmrc.gov.au/guidelines/resources-guideline-developers. Accessed 30 May 2013.
66. Fong DY, Lee CF, Cheung KM, Cheng JC, Ng BK, Lam TP, Mak KH, Yip PS, Luk KD. A meta-analysis of the clinical effectiveness of school scoliosis screening. Spine. 2010;35:1061–71.
67. Sabirin J, Bakri R, Buang SN, Abdullah AT, Shapie A. School scoliosis screening programme – a systematic review. Med J Malaysia. 2010;65:261–7.
68. National Health and Medical Research Council. Child health screening and surveillance: a critical review of the evidence. Centre for Community Child Health, Royal Children's Hospital Melbourne, Australia; 2002.
69. [No authors listed]. Screening for idiopathic scoliosis in adolescents: a brief evidence update for the U.S. Preventive Services Task Force. 2004. AHRQ Pub. No. 05-0568-B. http://www.uspreventiveservicestaskforce.org/3rduspstf/scoliosis/scolioup.pdf.
70. American Academy of Family Physicians. Summary of recommendations for clinical preventive services. 2015. Available from http://www.aafp.org/dam/AAFP/documents/patient_care/clinical_recommendations/cps-recommendations.pdf. Last accessed 1 Apr 2015.
71. Negrini S, Aulisa L, Ferraro C, Fraschini P, Masiero S, Simonazzi P, Tedeschi C, Venturin A. Italian guidelines on rehabilitation treatment of adolescents with scoliosis or other spinal deformities. Eura Medicophys. 2005;41:183–201.
72. Posadzki P, Lee MS, Ernst E. Osteopathic manipulative treatment for pediatric conditions: a systematic review. Pediatrics. 2013;132:140–52.
73. Lenssinck ML, Frijlink AC, Berger MY, Bierman-Zeinstra SM, Verkerk K, Verhagen AP. Effect of bracing and other conservative interventions in the treatment of idiopathic scoliosis in adolescents: a systematic review of clinical trials. Phys Ther. 2005;12:1329–39.
74. Negrini S, Fusco C, Minozzi S, Atanasio S, Zaina F, Romano M. Exercises reduce the progression rate of adolescent idiopathic scoliosis: results of a comprehensive systematic review of the literature. Disabil Rehabil. 2008;30:772–85.
75. Negrini S, Antonini G, Carabalona R, Minozzi S. Physical exercises as a treatment for adolescent idiopathic scoliosis. A systematic review. Pediatr Rehabil. 2003;3–4:227–35.

76. Fusco C, Zaina F, Atanasio S, Romano M, Negrini A, Negrini S. Physical exercises in the treatment of adolescent idiopathic scoliosis: an updated systematic review. Physiother Theory Pract. 2011;1:80–114.
77. Mordecai SC, Dabke HV. Efficacy of exercise therapy for the treatment of adolescent idiopathic scoliosis: a review of the literature. Eur Spine J. 2012;21:382–9.
78. Romano M, Negrini S. Manual therapy as a conservative treatment for adolescent idiopathic scoliosis. Scoliosis. 2008;3:2.
79. Gleberzon BJ, Arts J, Mei A, McManus EL. The use of spinal manipulative therapy for pediatric health conditions: a systematic review of the literature. J Can Chiropr Assoc. 2012;56:128–41.
80. McKennedy K, Sinclair Elder A, Elder C, Hutchins A. Myofascial release as a treatment for orthopaedic conditions: a systematic review. J Athl Train. 2013;48:522–7.
81. Dolan LA, Weinstein SL. Surgical rates after observation and bracing for adolescent idiopathic scoliosis: an evidence-based review. Spine. 2007;32:S91–100.
82. Maruyama T, Grivas TB, Kaspiris A. Effectiveness and outcomes of brace treatment: a systematic review. Physiother Theory Pract. 2011;27:26–42.
83. Davies E, Norvell D, Hermsmeyer J. Efficacy of bracing versus observation in the treatment of idiopathic scoliosis. Evid Based Spine Care J. 2011;2:25–34.
84. Sanders JO, Newton PO, Browne RH, Herring AJ. Bracing in adolescent idiopathic scoliosis, surrogate outcomes, and the number needed to treat. J Pediatr Orthop. 2012;32 Suppl 2:S153–7.
85. Focarile FA, Bonaldi A, Giarolo MA, Ferrari U, Zilioli E, Ottaviani C. Effectiveness of non-surgical treatment for idiopathic scoliosis. Overview of available evidence. Spine. 1991;16:395–401.
86. Rowe DE, Bernstein SM, Riddick MF, Adler F, Emans JB, Gardner-Bonneau D. A meta-analysis of the efficacy of non-operative treatments for idiopathic scoliosis. J Bone Joint Surg Am. 1997;79:664–74.
87. Weiss HR, Goodall D. The treatment of adolescent idiopathic scoliosis (AIS) according to present evidence. A systematic review. Eur J Phys Rehabil Med. 2008;44:177–93.
88. Green BN, Johnson CD, Moreau W. Is physical activity contraindicated for individuals with scoliosis? A systematic literature review. J Chiropr Med. 2009;8:25–37.
89. Li XF, Li H, Liu ZD, Dai LY. Low bone mineral status in adolescent idiopathic scoliosis. Eur Spine J. 2008;17:1431–40.
90. Saccucci M, Tettamanti L, Mummolo S, Polimeni A, Festa F, et al. Scoliosis and dental occlusion: a review of the literature. Scoliosis. 2011;6:15. Available: http://www.scoliosisjournal.com/content/6/1/15.
91. Pieper D, Mathes T, Eikermann M. Impact of choice of quality appraisal tool for systematic reviews in overviews. J Evid Based Med. 2014;7:72–8.
92. Pieper D, Buechter RB, Li L, Prediger B, Eikerman M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. J Clin Epidemiol. 2015;68:574–83.
93. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA, the Cochrane IPD Meta-analysis Methods Group. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. BMJ. 2015;350:h1088.
94. Peinemann F, Tushabe DA, Kleijnen J. Using multiple types of studies in systematic reviews of health care interventions – a systematic review. PLoS One. 2013;8:e85035.

# Case Study in Prevention

# 19

Leandro Fórnias Machado de Rezende,
Juan Pablo Rey-López, Olinda do Carmo Luiz,
and Jose Eluf-Neto

**Abstract**

In this chapter, we highlight our specific experience in conducting and reporting an overview of systematic reviews about sedentary behavior and health outcomes. In this overview, we aimed to cover all types of sedentary behavior, health outcomes, and age groups, taking into account the methodological quality of the systematic reviews. We hope to contribute to the discussion of methodological aspects of overviews of systematic reviews for prevention and health, especially on emerging determinants of health, where there is little conceptual or methodological uniformity between studies.

## 19.1 Introduction

Prevention is defined as actions aimed at eradicating, eliminating, or minimizing the impact of disease and disability [1]. Epidemiology is one of the central pillars upon which public health is based and epidemiology in turn has two central actions which help achieve this purpose. These are: (1) identify causes related to health outcomes and (2) intervene to prevent it [2].

Today, noncommunicable diseases are the main causes of death worldwide (68 % of all deaths – 38 million), killing more than all other causes together, particularly among low- and middle-income countries. Among these, cardiovascular disease, cancer, respiratory disease, and diabetes are the main specific causes of death [3]. The central focus of noncommunicable disease prevention is centered in

L.F.M. de Rezende, BSc, MSc (✉) • J.P. Rey-López • O.d. C. Luiz • J. Eluf-Neto
Departamento de Medicina Preventiva, Faculdade de Medicina da Universidade de São Paulo, Av. Dr. Arnaldo 455, 1° andar., Sao Paulo, SP 01246-903, Brazil
e-mail: lerezende@usp.br



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_19

four modifiable behavioral risk factors (causes): tobacco use, physical inactivity, harmful use of alcohol, and unhealthy diet. The Global Burden of Disease study estimated that these four risk factors collectively accounted for almost half of global deaths [4].

Physical inactivity, defined as less than 150 min of moderate to vigorous physical activity per week, causes 9% of premature mortality (5.3 million) worldwide [5]. Additionally, it is strongly related to coronary heart disease, type 2 diabetes, breast and colon cancer, stroke, depression, high blood pressure, metabolic syndrome, and falling [5]. This body of evidence has been accumulated – since the 1950s – based mainly on self-reported measures that defined physical inactivity as lack of moderate to vigorous physical activity (MVPA).

However, even among the physically active, prolonged periods of sitting time can be accumulated. For example, in a sample of American adults, MVPA represents only 3 % of an active adult waking time, whereas light-intensity activity and sedentary activity represent 39 and 58 %, respectively. In this sense, the MVPA approach neglects a large portion of adult waking time [6], and such phenomena have instigated the emergence of a new paradigm in the field of physical activity.

At the beginning of the twenty-first century, epidemiological studies started to investigate the health consequences of spending prolonged periods of sitting time, independently of MVPA practice, questioning the optimal daily movement for health – 150 min of MVPA per week for adults [7]. This element of research interest was defined as "sedentary behavior."

Sedentary behavior has been defined as time spent engaged in sitting or lying down activities that require an energy expenditure of 1.0–1.5 metabolic equivalents of tasks (METs). Nonetheless, nowadays, this definition makes accurate assessment of sedentary behavior extremely complex. For example, an individual may expend less than 1.5 METs in a standing position, while another may spend more than 1.5 METs while sitting. Furthermore, we still do not know the biological mechanisms that explain the negative consequences attributed to sedentary behaviors. Is the inactivity itself or other unhealthy behaviors associated with some sedentary behaviors? Thus, there are gaps in the knowledge, together with complexities of assessment arising from the variety of domains from which data are provided. These include work, education, transport, environment (e.g., indoor/outdoor, built environment), type of behavior (e.g., screen-based/not screen-based), posture (e.g., sitting, lying), social environment (e.g., alone, with friends, family), and time (e.g., time of day, year) [8].

In such a complex scenario, hundreds of epidemiologically and clinically based studies have investigated the association between sedentary behavior (predominantly using questionnaires) and multiple health outcomes. Many systematic reviews have been published with the aim of synthesizing the etiological evidence gathered from sedentary behavior studies. However, these reviews were focused on specific types of sedentary behavior (e.g., television viewing), age groups (children, adolescents, adults), or health outcomes (e.g., all-cause mortality), making it difficult to appreciate of the current state of the art and gaps in understanding.

## 19.2 Methods: Step by Step

In light of this knowledge gap, we performed an overview of systematic reviews with the aim of synthesizing the current observational evidence that relates sedentary behavior and health outcomes while being aware of the many methodological problems of this new research area. To achieve this aim, we restricted our search to systematic reviews of sedentary behavior that included a health outcome [9].

### 19.2.1 Designing the Research Question

To perform an overview of systematic reviews, a well-defined research question is needed, as with original studies and systematic reviews. Therefore, we strived to follow currently used strategies such as PICOS (population, intervention, control group, and outcome) or PEOD (population, exposure, outcome, study design) that could be adapted and used for overview of systematic reviews as well.

In our case, sedentary behavior studies were predominantly observational, and we opted to restrict our overview to these designs. Hence, we used PEOD to formulate our overview research question:

- Population: no restriction. As we were interested in the detrimental impact of sedentary behaviors on health across the lifespan, we did not restrict our research question by population (children, adolescents, adults, and older adults).
- Exposure: sedentary behavior. We used the sedentary behavior research network definition (time spent engaged in sitting or lying down activities that require an energy expenditure of 1.0–1.5 METs). A well-defined exposure is extremely important for an overview, especially in our study because sedentary behavior studies were still recent, and confusion over the term "physical inactivity" (lack of moderate to vigorous physical inactivity) was still ongoing.
- Outcomes: no restriction. As we were interested in how the scientific literature understands the detrimental impact of sedentary behaviors on health, we did not restrict our research question by outcome.
- Design: observational studies (ecological, cross-sectional, case-control, and cohort studies).

### 19.2.2 Search Strategy

Like the design of a research question, the definition of the search strategy is also very similar to that for systematic reviews. Therefore, a thorough check of the previous database selection procedure and other additional efforts are warranted to identify the main available literature. To select databases, it is important to know the main sources of information and their coverage. For instance, in our study, we used the main biomedical databases such as Medline, Embase, PsycINFO, and Web of

**Table 19.1** Main databases and their vocabulary for reviews

| Databases | Vocabulary |
|---|---|
| Medline | MesH terms |
| Embase | Emtree |
| Lilacs | DeCS |
| Web of Science | MesH terms |
| PsychINFO | APA terms |
| SportsDiscus | MesH terms |
| CINAHL | CINAHL terms (same as MesH terms) |

Science. These databases cover hundreds of journals worldwide, including those most widely cited which give us confidence that systematic reviews regarding sedentary behavior and health would be found. We did not include the Cochrane Library (CENTRAL), another important biomedical database, because of its focus on randomized controlled trial (RCT) studies and systematic reviews of RCTs, a study design not included in our overview.

To select the keywords, it was important to localize the specific vocabulary for each of the databases selected for the overview as shown in Table 19.1.

Although the sedentary behavior literature has identified many types of sedentary behavior, we aimed to find all published systematic reviews using the following keywords: sedentary behavior, sedentary lifestyles, sedentary time, sitting time, television viewing, driving, screen-based, video game, computer, and screen time, always respecting the specific vocabulary of each database (Table 19.1). To restrict our search to systematic reviews or meta-analysis only, we used filters, when available, or related keywords. Detailed information is shown in Supplementary File 1 in Rezende et al. [9].

Despite the well-designed search strategy, it is important to include other search strategy methods in order to provide additional records that may not be identified in the first selection stage. Additional search strategy methods, such as reading references from included articles and contact with experts, have been extensively used in systematic reviews.

To increase the number of systematic reviews in our overview, we searched additional sources such as references from the selected systematic reviews and contacted members of the Sedentary Behaviour Research Network (professors, researchers, and students).

Before starting to read the potential eligible records, it can be useful to import records to reference management software (e.g., EndNote, Zotero, Mendeley) in order to remove duplicate database results. We removed duplicated records in our study using the EndNote Web (Thomson Reuters, Carlsbad, CA, USA)

### 19.2.3 Study Selection

As with systematic reviews, to guarantee transparency and reduce errors, overviews of systematic reviews should select eligible studies by having two independent

reviewers to read titles, abstracts, and after that, entire papers. It is also recommended that both reviewers work in the field under investigation, and the opinion of a third reviewer should be available when needed.

In our overview, we followed all the steps described above. All eligible articles were evaluated by two independent reviewers, who examined all of the observational evidence. Disagreements between the two reviewers were settled by a third reviewer. The participation of several researchers is an important methodological issue because it allowed us to increase the validity of our conclusions. To illustrate this point, authors produced some minor errors during both the selection and appraisal of the best evidence available; these errors were rapidly identified and corrected by the reviewers.

To be included in the overview, as described in the research question section, articles had to be systematic reviews, with or without a meta-analysis, which examined the relationship between sedentary behavior and health outcomes using exclusively observational studies. The concept of sedentary behavior used in our overview was broad (every behavior produced during waking hours that elicits a low energy expenditure in a reclining or sitting position). Importantly, some authors have mistakenly defined sedentary behavior as not meeting the physical activity guidelines (150 min of moderate to vigorous physical activity) designed for health promotion. We excluded all those studies that inappropriately classified sedentary behavior as being physically inactive. We were fully aware that our selection of studies was limited to observational studies. This was an important limitation because, theoretically, with this type of design, multiple types of bias may exist, and thus, only preliminary conclusions could be obtained. Unfortunately, today there are very few randomized controlled trials related to sedentary behavior and health outcomes. Moreover, given that many diseases develop after a very prolonged exposure, knowledge provided by short-term interventions is suggestive but clearly insufficient to draw a definite conclusion. Finally, we also excluded studies based on interventions to reduce sedentary behavior, determinants/correlates of sedentary behavior, the tracking of sedentary behavior, and different methods for measuring sedentary behavior.

### 19.2.4 Data Extraction

As in the study selection step, it is also important to have two independent researchers to extract information from the selected studies. Regarding the relevant information/statistics, it is important to prepare a form in advance with research question items (PICO or PEOD) as well as other relevant information that might be important for gathering detailed and best available evidence.

We extracted information on author(s), year of publication, age group, type of sedentary behavior, type of health outcome measure, whether a meta-analysis was conducted or not, quality assessment of the original studies, eligibility criteria, and whether physical activity was included as a covariate or not.

### 19.2.5 Methodological Quality and Evidence Level Assessment

In order to ensure good methodological quality of a systematic review, the reviewers should analyze all the available evidence for a specific research question. They should report a transparent search strategy, the eligibility criteria used, data extraction, and synthesis methods. In addition, sometimes systematic reviews aiming to answer similar research questions find different results, and it is important to identify the source of such divergence in order to understand the current state of the art and knowledge gaps. To this end, tools aimed at assessing the methodological quality of systematic reviews (e.g., AMSTAR [10, 11], Robis [12]) have recently been designed. These generally focus on content related to the methodological quality of systematic reviews, for example, are the following items included: "a priori" design, details of duplicate study selection/data extraction, a comprehensive literature search, status of publication as inclusion criteria (i.e., gray or unpublished literature), a list of studies included/excluded, list of characteristics of included studies, assessment and documentation and scientific quality, appropriate formulation of conclusions (based on methodological rigor and scientific quality of the studies), assessment of suitability of methods of combining studies (homogeneity test, effect model used, and sensitivity analysis), assessment of publication bias (graphic and/or statistical test), and conflict of interest statement.

To evaluate the scientific quality of the identified reviews, we used the Assessing the Methodological Quality of Systematic Reviews (AMSTAR) tool [10, 11]. AMSTAR contains 11 items that examine the methodological quality of one systematic review. A detailed description of this instrument has been published elsewhere [10, 11]. Briefly, AMSTAR is based on a total score (range 0–11). If the question is affirmative, we assigned 1 point. Otherwise, 0 points were scored in the case of a negative or nonconclusive response. A high-quality systematic review requires at least 9 points, a moderate at least 5, and lower than 5 is considered as low quality. In our overview, we decided to show only the final score (rating) in a continuous scale because it offers a clearer interpretation of the quality achieved.

There are no current statements determining how to evaluate the evidence level of findings. In our overview, we established several steps to determine the observational evidence of association between sedentary behavior and different health outcomes. Firstly, we selected the best systematic reviews according to the rating obtained from AMSTAR. From the selected systematic reviews, we obtained level of evidence obtained for each health outcome: strong, moderate, insufficient, or no evidence. Secondly, we maintained the conclusions of each systematic review if it took into account the methodological quality and several covariates (especially physical activity) of the included studies. However, we decreased the level of the evidence (e.g., from strong to moderate) when reviews did not took into account the above conditions.

### 19.2.6 Registration and Ethical Aspects

As this was an overview of systematic reviews, registration of our study on databases such as PROSPERO was not appropriate. Registration in PROSPERO is

recommended for any systematic review of healthcare interventions and preventive or diagnostic strategies with a clinical outcome, but not reviews or overviews. There is also no ethical approval requirement for overviews of systematic reviews.

## 19.3 Results: Synthesizing the Available Evidence

A systematization of the results was carried out in order to understand the relationship between sedentary behavior and health and identify possibilities for preventive actions. A second concern was consensus and controversies about the relationship between sedentary behaviors and health outcomes, together with efforts to guide the proposals of new researchers, particularly those concerned with prevention.

During the study selection, our first finding was a large number of articles, and a large number of systematic reviews, confirming the need for a broad systematization of knowledge. A first approach indicated that some systematic reviews did not recognize the difference between sedentary behavior and physical inactivity, which is an essential conceptual distinction for understanding the potential impacts of sedentary behavior on health. As described in the methods section, these systematic reviews were therefore excluded, and the criteria for inclusion and exclusion were upgraded.

List of items defined by the two researchers were compared, and differences were arbitrated by a third investigator. This information was then systematized in the form of a flowchart, as recommended by the international guidelines and is reproduced below (Fig. 19.1).

Once identified, the final number of reviews that would be included in the overview, the contents of the articles were examined and the resulting differences and similarities were described.

From these, we assessed the methodological quality of the reviews based on AMSTAR (see Supplementary File 4 in Rezende et al. [9]). Among children and adolescents, 46% (6 reviews) of the reviews scored ≥6 points. Among adults, 62% (5 reviews) scored ≥6 points, whereas the 43% of the reviews with unspecified ages (3 reviews) had a total score of ≥6 points. The score obtained in the methodological quality assessment was used to identify the best available systematic reviews aimed at answering specific research questions (by age group, outcome, and sedentary behavior measure).

Due to our initial concern with prevention, and the differences between studies, our analysis and the systematization according to outcomes were more prominent and detailed (Table 19.2). Thus, taking into account each age group, we analyzed the relationship between sedentary behavior and health outcomes (including all-cause mortality, cancer, type 2 diabetes, metabolic syndrome, cardiovascular risks, obesity) gathering the number of reviews, how reviews considered exposure measurement, the covariates treated (Supplementary file 2 and 3 in Rezende et al. [9]), and the conclusion of the reviews on the association between sedentary behavior and health outcomes (Table 19.2).

Through this analysis, we were able to verify that health policies aimed at reducing the time spent in screen-based activities (e.g., watching television) in children

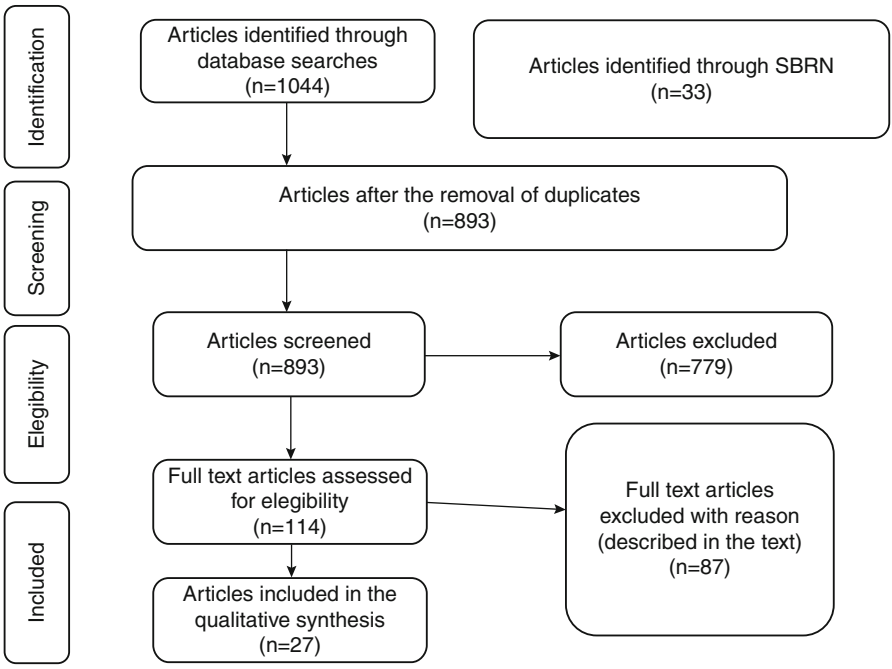



**Fig. 19.1** Preferred reporting items for systematic review flow diagram of the studies included in our overview [9]

and adolescents can safely reduce obesity. They may also reduce (moderate evidence) blood pressure and total cholesterol, reduce social problems, and increase self-esteem and physical activity. Further well-designed studies are needed to confirm these data.

In adults, health policies to reduce screen time, television time, and time sitting could theoretically reduce mortality from all causes, cardiovascular disease, type 2 diabetes, and metabolic syndrome. With less certainty, we can hypothesize that a reduction in sedentary behavior (e.g., occupational sitting) will reduce the incidence of ovarian, colon, and endometrial cancers.

There are inconclusive findings about the association between sedentary behavior and cancer mortality: incidence of breast, colorectal, and ovarian cancer; depressive symptoms; musculoskeletal disorders; and health-related behaviors in adults. In children, there are uncertainties in the association between sedentary behavior and metabolic syndrome, mental health, musculoskeletal disorders, bone mass, and motor dysfunction. These outcomes, therefore, need to be further explored in future studies. In conclusion, our overview proved to be a powerful tool to organize the more general knowledge about outcomes related to sedentary behavior, even in a complex field with little conceptual and methodological uniformity. On the other hand, this method did not allow the assessment of individual studies or the most frequent limitations of these studies. On the other hand, this method did not allow

**Table 19.2** Level of scientific evidence for associations between sedentary behaviors and health outcomes, by age group and type of sedentary behavior

| Outcomes | Children/adolescents | Adults |
|---|---|---|
| *Mortality* | | |
| All-cause mortality | No evidence | Strong evidence[a,b,c] |
| CVD mortality | No evidence | Strong evidence[a,b,c] |
| Cancer mortality | No evidence | No evidence[a,c,d] |
| *Cardiovascular diseases* | No evidence | Strong evidence[a,b,c]<br>Insufficient evidence[d] |
| *Cancer* | | |
| Breast | No evidence | Insufficient evidence[d] |
| Colorectal | No evidence | Insufficient evidence[d] |
| Colon | | Moderate evidence[f] |
| Endometrial | No evidence | Moderate evidence[f] |
| Ovarian | No evidence | Moderate evidence[f]<br>Insufficient evidence[d] |
| Prostate | No evidence | Insufficient evidence[f] |
| *Type 2 diabetes* | No evidence | Strong evidence[a,b]<br>Moderate evidence[c]<br>Insufficient evidence[d] |
| *Metabolic syndrome* | Insufficient evidence[a,b,c] | Strong evidence[a,b,c,e] |
| *Individual cardiovascular risk factors* | | |
| Blood pressure | Moderate evidence[a] | Insufficient evidence[a] |
| Total cholesterol | Moderate evidence[a] | Insufficient evidence[a] |
| HbA1 | Insufficient evidence[a,b,c] | Insufficient evidence[a] |
| Fasting insulin | Insufficient evidence[a,b,c] | Insufficient evidence[b] |
| Insulin resistance | Insufficient evidence[a,b,c] | Insufficient evidence[b] |
| Leptin | No evidence | Insufficient evidence[a] |
| Fibrinogen | No evidence | Insufficient evidence[a] |
| C-peptide | No evidence | Insufficient evidence[a] |
| *Obesity* | Strong evidence[a,b] | Insufficient evidence[a,d] |
| *Mental health* | | |
| Self-esteem | Moderate evidence[a,b] | No evidence |
| Depressive symptoms | No evidence | Insufficient evidence[c] |
| Postnatal depression | No evidence | Insufficient evidence[c,d] |
| Cognitive aspects | Insufficient evidence[a] | No evidence |
| *Musculoskeletal* | Insufficient evidence[b] | Insufficient evidence[d] |
| *Other behaviors (PA, diet, alcohol consumption)* | Insufficient evidence[a] | Insufficient evidence[a] |
| Social behavior problems | Moderate evidence[a,b] | No evidence |
| *Other health outcomes* | | |
| Bone mass | Insufficient evidence[c] | No evidence |
| Motor dysfunction | Insufficient evidence[a] | No evidence |
| Physical fitness | Moderate evidence[a,b] | No evidence |

(continued)

**Table 19.2** (continued)

| Outcomes | Children/adolescents | Adults |
|---|---|---|
| Academic achievement | Moderate evidence[a,b] | No evidence |
| Symptomatic gallstone disease | No evidence | Insufficient evidence[e] |

[a]Television viewing
[b]Screen time
[c]Total sitting time
[d]Occupational sitting time
[e]Objectively measured sedentary time
[f]Unspecified

the assessment of individual studies or the most frequent limitations of these studies. This made it impossible to establish a more solid knowledge base of the current evidence. Finally, based on this specific experience in conducting and reporting an overview of systematic reviews, we also suggest directions for future original and systematic review studies. For example, epidemiological studies with longitudinal design and better sedentary behavior and covariate measurements are still required. For systematic reviews, detailed information regarding the quality assessment of the original studies, study selection and data extraction procedures, inclusion of gray literature, complete listing of included and excluded studies, and evaluation of publication bias is still needed.

## References

1. Porta M. A dictionary of epidemiology. 5th ed. New York: Oxford University Press; 2008.
2. Galea S. An argument for a consequentialist epidemiology. Am J Epidemiol. 2013;178(8):1185–91. doi:10.1093/aje/kwt172.
3. World Health Organization. Global status report on noncommunicable diseases 2014. Geneva: WHO Press; 2014.
4. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. Lancet. 2015;385(9963):117–71. doi:10.1016/s0140-6736(14)61682-2.
5. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012;380(9838):219–29. doi:10.1016/s0140-6736(12)61031-9.
6. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey Data 2003-2004, 2005-2006. Atlanta: Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS); 2009–2010.
7. Katzmarzyk PT. Physical activity, sedentary behavior, and health: paradigm paralysis or paradigm shift? Diabetes. 2010;59(11):2717–25. doi:10.2337/db10-0822.
8. Chastin SFM, Schwarz U, Skelton DA. Development of a consensus taxonomy of sedentary behaviors (SIT): report of Delphi round 1. PLoS One. 2013;8(12):e82313. doi:10.1371/journal.pone.0082313.
9. Rezende LF, Lopes MR, Rey-Lopez JP, Matsudo VK, Luiz OC. Sedentary behavior and health outcomes: an overview of systematic reviews. PLoS One. 2014;9(8):e105620. doi:10.1371/journal.pone.0105620. http://127.0.0.1:8081/plosone/article?id=info:doi/10.1371/journal.pone.0105620.

10. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10. doi:10.1186/1471-2288-7-10.
11. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009;62(10):1013–20. doi:10.1016/j.jclinepi.2008.10.009.
12. Developing ROBIS – a new tool to assess the risk of bias in systematic reviews. http://gtr.rcuk.ac.uk/project/F489B8DD-D7C2-41B5-B54D-9F3CD722133E. Accessed 3 Mar 2015.

# Case Study of Public Health

# 20

Clare Bambra and Marcia Gibson

**Abstract**

Umbrella reviews are an established method of locating, appraising, and synthesising systematic review-level evidence. Umbrella review methodology is though only just beginning to emerge as a well-used technique in public health research. This chapter therefore summarises some of the first umbrella reviews conducted in the field of public health with a thematic focus on the social determinants of health and how interventions might affect health inequalities. The chapter discusses some of the cross-cutting methodological and thematic lessons learned from this body of work and concludes by suggesting new directions for umbrella reviews within the field.

## 20.1 Introduction

Umbrella reviews are an established method of locating, appraising, and synthesising systematic review-level evidence [1]. They use systematic review methodology to locate and evaluate published systematic reviews – most usually of interventions. Umbrella reviews are therefore able to present the overarching findings of such systematic reviews (usually considered to be the highest level of evidence) and can also extract data from the best quality studies within them if desired [2]. They therefore represent an effective way of rapidly reviewing a broad

C. Bambra (✉)
Centre for Health and Inequalities Research, Durham University, Durham DH13LE, UK
e-mail: clare.bambra@durham.ac.uk

M. Gibson
MRC Social and Population Health Unit, University of Glasgow, Glasgow, UK



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_20

evidence base. This can be particularly useful for policymakers or practitioners in public health who may require a quick answer to a question, or a quick overview of a field (and the gaps in it).

Umbrella review methodology is though only just beginning to emerge as a well-used technique in public health research. For example, the first protocol for a Cochrane Public Health Group "overview of reviews" was registered in January 2015 [3], despite such overviews being advocated by the wider Cochrane Collaboration since at least 2008 [1]. This partly reflects the smaller number of public health systematic reviews within the Cochrane database but also the fact that systematic reviews more generally are still relatively new within public health (at least compared to health care or clinical research). This chapter therefore summarises some of the first umbrella reviews in public health (in which the authors of this chapter have been involved). They are methodologically innovative as they pioneered and expanded the use of umbrella systematic review methodology into public health research.

Methodologically, systematic reviews within public health research can differ quite significantly from those in other areas of health care or psychology. For example, a higher proportion of public health research is conducted using observational designs (given the population scale) and experimental studies are consequently less common; the traditional evidence hierarchy has been challenged by some public health reviewers with alternative models suggested instead [4]; overarching meta-analysis is unusual in public health reviews due to the heterogeneity of included study designs or outcomes [5]; qualitative research has also begun to be synthesised within public health reviews [6]; the nature of public health interventions can be very broad – potentially challenging traditional ways of framing the review question [7]; database searches for public health systematic reviews may also need to be broader than in other areas given the multiple areas of research (e.g. education, geography, psychology, social policy, health care, etc.) that might be relevant to a public health topic [8]. Consequently these differences are also reflected in the breadth and inclusion criteria of umbrella reviews conducted in public health to date – something that is apparent in our case studies.

The six case studies of umbrella reviews of public health interventions that we summarise here focus on:

1. Transport and health [9]
2. Housing and health [10]
3. Health-care service quality [11]
4. Equity in health-care services [12]
5. Workplace health [13]
6. The wider determinants of health [2]

They all have in common a concern with how interventions might affect health inequalities by addressing the social determinants of health. It is therefore worth providing a little context on these two issues so that the reviews can be understood in their appropriate conceptual context.

## 20.2 Health Inequalities and the Social Determinants of Health

The term "health inequality" is usually used to refer to the systematic differences in health which exist between socio-economic groups (in terms of income, education, or occupational class) or socio-economic areas (e.g. low-income areas). Socio-economic and spatial inequalities in health are not restricted to differences between the most privileged groups/areas and the most disadvantaged; health inequalities exist across the entire social gradient [14]. There are four levels of interventions to tackle inequalities:

1. Strengthening individuals (person-based strategies to improve the health of disadvantaged individuals)
2. Strengthening communities (improving the health of disadvantaged communities and local areas by building social cohesion and mutual support)
3. Improving living and school environments (reducing exposure to health-damaging material and psychosocial environments across the whole population)
4. Promoting healthy macro policy (improving the macroeconomic, cultural, and environmental context which influences the standard of living achieved by the whole population) [15]

These interventions are further underpinned by one of three different approaches to health inequality [16]:

1. Disadvantage (improving the absolute position of the most disadvantaged individuals and groups)
2. Gap (reducing the relative gap between the best and worst-off groups)
3. Gradient (reducing the entire social gradient)

Interventions are thus either targeted (such as individual level interventions which are underpinned by health as disadvantage) or universal (such as living and school conditions interventions which potentially influence the entire social gradient in health) [17].

The social determinants of health are the conditions in which people work and live – what have been referred to as the fundamental causes of health inequalities [14]. The main social determinants of health are widely considered to be:

1. Access to essential goods and services (specifically water, sanitation, and food)
2. Housing and the living environment
3. Access to health care
4. Working conditions
5. Unemployment [18]

Access to clean water and hygienic sanitation systems are the most basic prerequisites for good public health. Agricultural policies affect the quality, quantity,

price, and availability of food, all of which are important for public health. Physical housing conditions and the cost of housing are both linked with public health. The wider living environment – such as pollution levels, transport infrastructure, access to green space, crime and safety, or place-based stigma – is also recognised as potentially important for individual-, household-, and area-level health. Access to good quality, affordable, and timely health care is a fundamental determinant of health, particularly in terms of the treatment of pre-existing conditions. Physical and psychosocial working conditions are a major cause of ill-health in the working age population and, because of the steep social gradient in conditions, are an important factor behind social inequalities in health. Unemployment is associated with an increased likelihood of morbidity and mortality as a result of the material (e.g. wage loss and resulting changes in access to essential goods and services) and/or psychosocial effects of unemployment (e.g. stigma, isolation, and loss of self-worth). Lower socio-economic classes are disproportionately at risk of unemployment.

## 20.3 Case Studies in Health Inequalities and the Social Determinants of Health

### 20.3.1 Case Study 1: Transport and Health [9]

**Background to the review** Transport is an important determinant of health and there is a well-established association between socio-economic status (SES) and risk of road accidents. Effective traffic calming interventions such as 20 mph zones and limits may therefore improve health and reduce health inequalities.

**Review objective** To identify systematic reviews of the effects of 20 mph zones (including speed limits and road humps) and 20 mph limits on health and SES inequalities in health amongst adults and children.

**Study inclusion criteria** *Population*: Children and adults, all ages.

*Intervention*: 20 mph zones and limits. 20 mph limits consist of simply changing the speed limit to 20 mph using signage, whereas zones include additional traffic calming measures in whole areas in addition to changing the speed limit. Such traffic calming measures may, for example, include installation of road humps or mini-roundabouts.

*Context*: Any country, any location, English language only, and publications from 1990.

*Outcomes*: Health and SES inequality outcomes. Health inequalities were defined as differences by income, education, or occupational class, including area measures, e.g. area-level deprivation. Primary outcome measures included morbidity, health behaviours (especially physical activity such as walking and cycling), mortality, accidents, and injuries. Where additional data was provided, secondary outcomes included cost-effectiveness, public acceptance of schemes, and perceptions of safety.

*Study design(s)*: Systematic reviews of quantitative evaluation studies. Publications had to meet the two mandatory criteria of Database of Abstracts of Reviews of Effects (DARE): (a) that there is a defined review question (with definition of at least two of the participants, interventions, outcomes, or study designs) and (b) that the search strategy included at least one named database, in conjunction with either reference checking, hand-searching, citation searching, or contact with authors in the field.

**Search strategy** 12 databases were searched from 1990 to September 2013, Campbell Collaboration, Cochrane Library (includes Cochrane Database of Systematic Reviews [CDSR], Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, DARE, Health Technology Assessment Database, NHS Economic Evaluation Database, and About the Cochrane Collaboration), EMBASE, PsycINFO, Centre for Review and Dissemination, Database of Promoting Health Effectiveness Reviews (DoPHER), Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre), SafetyLit, Transport Research Information Service (TRIS), PROSPERO, MEDLINE, and Applied Social Sciences Index and Abstracts (ASSIA). Grey literature was also searched as well as the following websites: ROSPA, NICE, and Department for Transport.

**Data extraction and quality appraisal** Screening, data extraction, and quality appraisal of included studies were carried out by one reviewer and checked by a second. The methodological quality of each systematic review was appraised using adapted DARE criteria (http://www.crd.york.ac.uk/CRDWeb/AboutDare.asp). The criteria were as follows: (1) is there a well-defined question; (2) is there a defined search strategy; (3) are inclusion/exclusion criteria stated; (4) are study designs and number of studies clearly stated; (5) have the primary studies been quality assessed; (6) have the studies been appropriately synthesised; (7) has more than one author been involved in each stage of the review process. Based on these criteria, included reviews were categorised as low (met 0–3 criteria), medium (4–5), or high (6–7) quality.

**Synthesis and analysis** Narrative synthesis.

**Results** Five systematic reviews were included. There were no reviews that focused exclusively on 20 mph zones or limits, but within these five reviews, there were a total of ten unique studies on 20 mph zones ($n=8$) or limits ($n=2$). Four of the systematic reviews were high quality and one was rated medium quality. The studies focused on accidents and injuries, traffic speed and volume, perceptions of safety, and physical activity. None of the studies, however, examined SES inequalities in these outcomes. Overall, they provide convincing evidence that these measures are effective in reducing accidents and injuries, traffic speed, and volume, as well as improving perceptions of safety in two of the studies. There was also evidence that such interventions are potentially cost-effective. There was no evidence of the effects on SES inequalities in these outcomes.

**Conclusion** 20 mph zones and limits are effective means of improving public health via reduced accidents and injuries. Whilst there was no direct evidence on the effects of interventions on health inequalities, targeting such interventions in deprived areas may be beneficial.

### 20.3.2 Case Study 2: Housing and Health [10]

**Background to the review** Housing and neighbourhood conditions are widely acknowledged to be important social determinants of health, through three main pathways: (1) internal housing conditions, (2) area characteristics, and (3) housing tenure. Poor housing conditions disproportionately affect lower socio-economic groups. Housing or neighbourhood interventions which target these pathways may improve health and health inequalities.

**Review objective** To identify systematic reviews of housing and neighbourhood interventions which target internal housing conditions, area characteristics, or housing tenure and measure impacts on health and health inequalities.

**Study inclusion criteria** *Population*: Adult participants or the general population.

*Interventions*: Interventions aimed at altering housing or neighbourhood conditions which collected data on health or well-being outcomes.

*Context*: OECD countries (North America, Europe, Australasia, Japan). English language from 2000 to 2007.

*Outcomes*: Health and health inequality outcomes. Physical and mental health outcomes, including youth behavioural problems, morbidity, mortality, violence, injuries, and health and safety risks. Secondary outcomes included crime and social disorder, community cohesion, and economic outcomes.

*Study design(s)*: Systematic reviews of quantitative evaluation studies. Systematic reviews had to meet the two mandatory criteria of DARE.

**Search strategy** We searched 6 electronic databases including the CRD Wider Public Health database (2000–2002), the Cochrane Database of Systematic Reviews (2000–2007), the Criminal Justice Abstracts database (2000–2007), DARE (2002–2007), the Campbell Collaboration Database (2002–2007), and EPPI-Centre (2002–2007). Bibliographies and relevant websites were searched, and experts were contacted. Through expert contacts, we identified one review conducted outwith the search time frame, which was included because it represented a major contribution to the evidence base.

**Data extraction and quality appraisal** All titles and abstracts were independently screened by two reviewers, and relevant reviews were retrieved and assessed for inclusion. Data relating to the review methods (search strategy, inclusion criteria, synthesis) were extracted along with information about the intervention, participants, outcomes, results (including number of studies and study design) authors'

conclusions and research recommendations. Each systematic review was critically appraised by one reviewer and checked by another using a checklist list adapted from DARE criteria (see prior section).

**Synthesis and analysis** Narrative synthesis by pathway of effect.

**Results** Five reviews met the criteria for inclusion. Four of the reviews were judged to be high quality, and one was medium quality. Impacts on health inequalities were not measured directly. However, all of the included interventions were aimed at people of lower SES.

*Area characteristics*: Two reviews found that relocating families living in high poverty areas to more affluent areas has the potential to improve health but evidence is inconclusive due to methodological issues. One review of area-based regeneration was inconclusive, with positive and negative health impacts reported.

*Internal housing conditions*: There is compelling evidence for positive effects on warmth and energy efficiency interventions targeted at vulnerable individuals. However, the health impacts of area-level internal housing improvement interventions are as yet unclear.

*Housing tenure*: No reviews of these interventions were identified. This remains an important area for further research and potentially new evidence syntheses.

One further review included interventions aimed at several of the pathways linking housing to health, reporting that many of the studies found positive impacts on health. However, neither interventions nor outcomes were specified, hampering interpretation of the findings.

**Conclusion** Targeted warmth and energy efficiency interventions show positive impacts on a range of health measures. There was less robust evidence for positive effects of residential mobility programmes and a gap in the evidence base around housing tenure.

### 20.3.3 Case Study 3: Health-Care Service Quality [11]

**Background to the review** Health systems in high-income countries are coming under unprecedented pressure from several directions: pressure on costs, expenditure, and ideological pressure. In some countries these pressures are being used to justify renewed calls to undertake major reforms to the financing and delivery of health care. This is part of a longer trend in high-income countries of the marketisation and privatisation of health-care provision since the mid-1980s. The implications of these changes for the effectiveness of health-care systems need to be examined, particularly in relation to their effects on quality of care.

**Review objective** To review the systematic review-level evidence base on the effects of organisational and financial health system interventions on quality of health care.

**Study inclusion criteria** *Population*: Adults and children of all ages.

*Interventions*: Organisational and financial health-care system interventions were defined as: (1) system financing, (2) funding allocations, (3) direct purchasing arrangements, (4) organisation of service provision, and (5) health and social care system integration.

*Context*: Limited to the health systems of 15 high-income countries used by the Commonwealth Fund: Australia, Canada, Denmark, France, Germany, Iceland, Italy, Japan, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States.

*Outcomes*: Quality of care was defined in terms of (1) professional performance, (2) efficient treatment and care, (3) clinical outcomes, (4) person-centred care, (5) holistic care, and (6) patient satisfaction.

*Study design(s)*: Only systematic reviews of intervention studies with quantitative outcomes (experimental and observational) were included. Reviews were defined as "systematic" if they met the two mandatory criteria of DARE.

**Search strategy** Seven electronic databases were searched for English language studies from start to January 2013 – ASSIA, Campbell Collaboration Database (CDSR), DARE, EPPI-Centre; Medline; and PROSPERO. Citation follow-up was conducted on the bibliographies of included studies.

**Data extraction and quality appraisal** Screening, data extraction, and quality appraisal of included studies were carried out by two independent reviewers. The methodological quality of each systematic review was appraised using adapted DARE criteria (see prior sections).

**Synthesis and analysis** Narrative synthesis by intervention type.

**Results** Nineteen reviews met all criteria and were included in the synthesis. Nine of the nineteen reviews were of a high quality (mostly Cochrane reviews), three were of moderate quality, and seven were low quality. This umbrella review has identified only a small systematic review-level evidence base and substantial evidence gaps around certain interventions, most notably on changes to resource allocation systems (something also noted in our companion review of equity).

**Paying providers** The eight reviews of paying providers to promote quality were largely inconclusive.

**Purchasing and provision** The five reviews provided that no conclusive evidence on the outcomes of various forms of purchaser-provider split is particularly striking. The findings suggest that structural changes, such as the creation of new purchasing organisations, have very little impact on patients or frontline providers, and any changes that do occur are short lived. Furthermore, such arrangements seem to give rise to increased transaction costs that are not compensated for by cost savings.

**Integration of services** In contrast, there was some evidence from six reviews that greater integration of services can benefit patients, although much seems to depend upon the approach taken.

**Funding allocation** No systematic reviews examined the effects of funding allocation reforms.

**Direct purchasing arrangements** No systematic reviews examined the effects of funding allocation reforms on quality of care.

**Conclusion** The evidence base suggests that the privatisation and marketisation of health-care systems does not improve quality and that most financial and organisational system-level reforms have either inconclusive or negative effects.

### 20.3.4 Case Study 4: Equity in Health-Care Services [12]

**Background to the review** Over the last 25 years, the health-care systems of most high-income countries have experienced extensive – usually market-based – organisational and financial reforms. The impact of these system changes on health equity has been hotly debated. Examining evidence from systematic reviews of the effects of health-care system organisational and financial reforms will add empirical information to this debate, identify any evidence gaps, and help policy development.

**Review objective** To conduct an umbrella review of the evidence of the effects of organisational and financial health system interventions on equity of health care.

**Study inclusion criteria** *Population*: Adults and children of all ages.

*Interventions*: Organisational and financial health-care system interventions were defined as: (1) system financing, (2) funding allocations, (3) direct purchasing arrangements, (4) organisation of service provision, and (5) health and social care system integration.

*Context*: Limited to the health systems of 15 high-income countries used by the Commonwealth Fund: Australia, Canada, Denmark, France, Germany, Iceland, Italy, Japan, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States.

*Outcomes*: Health equity was defined in terms of socio-economic inequalities SES in health-care access and utilisation, health outcomes (e.g. self-rated health, mortality rates, disease prevalence, etc.), or income. SES inequalities were defined in terms of differences in outcomes by SES (income, education, occupational class) or outcomes for the most vulnerable or deprived groups (e.g. unemployed, lone parents, deprived areas, etc.).

*Study design(s)*: Only systematic reviews of intervention studies with quantitative outcomes (experimental and observational) were included. Reviews were

defined as "systematic" if they met the two mandatory criteria of DARE. Reviews were defined as "partially systematic" if two or more of these components of the review question could be inferred from the title or text and the search criteria were fulfilled.

**Search strategy** 7 electronic databases were searched for English language studies from start to January 2013 – ASSIA, Campbell Collaboration Database, CDSR, DARE, EPPI-Centre database of health promotion and public health studies; Medline; and PROSPERO. Citation follow-up was conducted on the bibliographies of included studies.

**Data extraction and quality appraisal** Screening, data extraction, and quality appraisal of included studies were carried out by two independent reviewers. The methodological quality of each systematic review was appraised using adapted DARE criteria (see prior sections).

**Synthesis and analysis** Narrative synthesis by intervention type.

**Results** Nine systematic reviews met all aspects of the inclusion criteria and were included in the synthesis. Only three of the nine reviews were of a high quality and only four were considered to be fully systematic.

**General system financing** The four systematic reviews identified suggest that increased use of private insurance has negative health equity impacts. In contrast, there is evidence from the United States that increased use of free-care programmes has positive health equity outcomes. The effects of US-managed care programmes are inconclusive.

**Direct purchasing** The single review of increased user fees and out of pocket payments found a negative impact on health equity.

**Organisation of services** In terms of the marketisation and privatisation of healthcare services, two of the three relevant reviews (including the better quality one) found that such reforms were negative for health equity, whilst the other review was inconclusive.

**Health and social care integration** The evidence on the equity effects of integrated partnerships between health and social services is inconclusive.

**Resource allocation** There were no relevant studies located that related to resource allocation reforms.

**Conclusion** The systematic review-level evidence base suggests that financial and organisational health care system reforms have had either inconclusive or negative impacts on health equity both in terms of access relative to need and in terms of health outcomes.

### 20.3.5 Case Study 5: Workplace Health [13]

**Background to the review** Although the work environment has long been acknowledged as an important determinant of health and health inequalities, physical working conditions have improved a great deal. However, inequalities remain in the psychosocial work environment and interventions to improve these may improve health and reduce health inequalities.

**Review objective** To systematically review studies reporting the impacts on health and health inequalities of workplace interventions aimed at psychosocial working conditions delivered at an organisational level.

**Study inclusion criteria** *Population*: Adult participants (16+) or the general population.

*Interventions*: Any change to the psychosocial work environment which focused on the organisational (rather than individual) level.

*Context*: Developed countries (North America, Europe, Australasia, Japan), reviews from 2000 to 2007.

*Outcomes*:

*Health*: Disease prevalence, general physical and psychological health measures, sickness absence, accident-related injury, and health behaviours.

*Well-being*: Physical and mental well-being, work/life balance, and quality of life. Health inequalities: differences in health or well-being by socio-economic status or demographic characteristics.

*Study design(s)*: Systematic reviews of quantitative evaluation studies. Systematic reviews had to meet the two mandatory criteria of DARE.

**Search strategy** The Centre for Reviews and Dissemination (CRD) Wider Public Health (WPH) database (a web-based database of systematic reviews of public health and related interventions) was searched from 2000 to 2002. In addition, CDSR, DARE, the Campbell Collaboration Database, and EPPI-Centre were searched from 2002 to 2007. The Criminal Justice Abstracts database was searched from 2000 to 2007, and hand-searching of relevant journals, bibliographies, and websites was conducted.

**Data extraction and quality appraisal** Two reviewers independently screened all titles and abstracts identified. Data were extracted by one reviewer and checked by a second. Each systematic review was critically appraised using a checklist list adapted from DARE (see prior sections).

**Synthesis and analysis** Narrative synthesis by intervention subtype.

**Results** Seven reviews addressing the health effects of changes to the psychosocial work environment were located: three examined increased employee control and four evaluated the effects of changes to the organisation of work (shift work, privatisation, health and safety legislation). Five of the reviews specifically examined effects on health inequalities. Five of the reviews met all seven of the critical appraisal.

**Employee control** One review of employee discussion groups found no conclusive effects on health. In another review, participatory employee committees were found to have positive impacts on self-rated health, and there was evidence of some effects on health inequalities. A further review of interventions which increased employees' control over work tasks found that mental health worsened when job control decreased. There was some evidence of differential effects on depression.

**Changes to work organisation** Four reviews, examining changes to shift work schedules (2), privatisation (1), and implementation of health and safety legislation (1), were located. Shift work interventions reported improved work/life balance and evidence of improved health outcomes, but little evidence on health inequalities. The review of privatisation reported that decreased job security led to adverse effects on mental health and on some physical health outcomes. Increased enforcement of health and safety regulation was associated with improved rates of fall injuries. There was limited evidence in the latter two reviews of differential effects by gender or occupational class.

**Conclusion** Organisational-level changes to the psychosocial work environment can have important and generally beneficial effects on health. Five reviews which examined differences by socio-economic or demographic group tentatively suggest that organisational workplace interventions may also have stronger effects on men, lower SES groups and ethnic minorities.

### 20.3.6 Case Study 6: Wider Determinants of Health [2]

**Background to the review** It is increasingly recognised that interventions aimed at the wider social determinants of health are necessary to tackle health inequalities. Developing the evidence base about interventions aimed at the social determinants of health requires that we identify existing evidence and highlight gaps in research.

**Review objective** To identify and synthesise existing systematic reviews which report the health impacts of interventions aimed at the wider social determinants of health.

**Study inclusion criteria** *Population*: Adult participants (16+) or the general population.

*Interventions*: Interventions aimed at the outermost layers of Dahlgren and Whitehead's "rainbow" model of social determinants: macroeconomic, cultural, and environmental conditions and living and working conditions (including water and sanitation, agriculture and food, access to health services, unemployment, work conditions, housing, education, and transport).

*Context*: Developed countries (North America, Europe, Australasia, Japan).

*Outcomes*: SES inequalities in health or well-being, overall population health impacts. Also impacts on social determinants of health amongst disadvantaged groups with an existing health condition.

*Study design(s)*: Systematic reviews of quantitative evaluation studies were included if they met the two mandatory DARE criteria.

**Search strategy** The Centre for Reviews and Dissemination Wider Public Health was searched from 2000 to 2002. CDSR, DARE, the Campbell Collaboration Database, and the EPPI-Centre database were searched from 2002 to 2007. The Criminal Justice Abstracts database was searched from 2000 to 2007. A wide range of relevant websites was also searched, as well as bibliographies and four leading journals (American Journal of Public Health, American Journal of Preventive Medicine, Journal of Epidemiology and Community Health, Social Science and Medicine), from January 2002 to April 2007.

**Data extraction and quality appraisal** Screening of titles and abstracts was conducted by two reviewers independently. Data from included reviews was extracted by two reviewers and cross-validated by another. Data from included reviews were only extracted if the primary studies and/or outcomes were relevant to the umbrella review. Quality was assessed by one reviewer and independently checked by a second, using criteria adapted from DARE (see prior sections).

**Synthesis and analysis** Narrative synthesis by social determinant domain and intervention type.

**Results** Thirty systematic reviews were identified, corresponding to the following domains within the "rainbow" model: housing and living environment (9), work environment (7), transport (5), access to health services (4), unemployment and welfare (3), agriculture and food (1), and water and sanitation (1). Twenty-six reviews were high quality and 4 were appraised as medium quality.

**Housing and living environment** There is some evidence for positive effects on health and social outcomes following relocation to less disadvantaged areas. Improvements to internal housing conditions are also associated with small improvements in health. Finding from reviews of fall reduction interventions was inconclusive. Reviews of area-based interventions also reported mixed results. There was little evidence on the effects on health inequalities.

**Work environment** Employee control interventions reported improved health when job control actually increased, and vice versa. Interventions which increased control over shift times had positive impacts on self-reported (particularly mental) health. Privatisation had negative effects on mental health associated with increased job insecurity. Increased health and safety enforcement in the construction industry was associated with a decrease in fall-related injuries. There was some evidence of differential effects.

**Transport** There was strong evidence from three reviews that driver alcohol restrictions, traffic calming, and speed cameras led to reductions in fatal and non-fatal crashes. Impacts of new road building varied according to road type (bypasses

reduced injuries while major new roads did not). The evidence base on interventions promoting walking and cycling was limited. Effects on health inequalities were not reported.

**Unemployment and welfare** Two reviews of interventions to promote employment found little evidence of health impacts and inconclusive evidence of employment impacts. A review of interventions to increase uptake of welfare benefits indicated that there were clear financial benefits. However, there were only short-term improvements in mental health. None of the reviews reported differential impacts, but all were aimed at disadvantaged groups.

**Access to health services** Three reviews of interventions to overcome cultural barriers to health-care access were inconclusive, although the use of lay health workers in low-income countries was associated with an increase in immunisation uptake. A review of rural outreach interventions reported improved health-care access and better self-reported health. All health-care reviews showed some promise in increasing access for disadvantaged groups, but none reported effects by SES or demographic characteristics.

**Agriculture and food** One review of financial incentives to improve diet found positive effects on weight loss and fruit and vegetable consumption. No evidence on differential effects was included in the review.

**Water and sanitation** One review of water fluoridation found no evidence of adverse effects on bone fracture incidence, bone mineral density, or bone strength in developed countries. The review did not report on the effects on health inequalities.

**Conclusion** There is a lack of evidence about the health impacts of interventions aimed at the wider social determinants of health, which is even greater in relation to health inequalities. Those reviews which reported differential impacts found some indications of differential effects by gender, occupational class, and ethnicity. The domains of education, food, water, health service access, and unemployment show the most striking paucity of evidence. Changes to housing conditions are associated with small positive effects on physical and mental health. Workplace interventions appear to have differing effects on different levels of employee. A number of transport interventions seem to deliver reductions in crash injuries. Evidence for the health effects of interventions aimed at unemployment and welfare, and health service access is either absent or inconclusive. Financial incentives show some promise in improving health and health behaviour.

## 20.4 Discussion

Umbrella reviews of systematic reviews of interventions in the field of public health can be particularly useful for giving a broad overview of the evidence in a given field, particularly when the growth in systematic reviews outstrips the ability of the

lay reader/practitioner to keep pace. They are also extremely useful for identifying gaps in the evidence base on a given topic. However, based on our experience of conducting umbrella reviews in public health, we have several observations, which may point the way towards methodological developments or improvements for the future.

As noted in the introduction to this chapter, public health is notable for the relative lack of experimental studies, in part due to the difficulty of evaluating many public health interventions using such study designs. However, as yet we are unaware of any umbrella reviews that only examined systematic reviews of observational studies in the field of public health. This may of course reflect the relative lack of systematic reviews of observational studies, but it is perhaps worth beginning to consider in what ways umbrella review methodology may need to develop in order to accommodate such study designs and, in particular, natural experiments and comparative studies, which are well suited to the evaluation of macro-level policy interventions [19].

The tendency for public health interventions to be particularly broad and to cross multiple disciplines presents particular challenges in developing search strategies for umbrella reviews. It is extremely common for different disciplines to employ different terms for the same concept, outcome, or indeed intervention, and it is unlikely that any one review team will contain expertise across all of the disciplines which might be involved. It may be that methodological progress in public health umbrella reviews will need to focus on developing new ways of developing search strategies. This is also a problem common to systematic reviews in public health too.

Many of the challenges involved in conducting public health umbrella reviews mirror those of systematic reviews, but are magnified by the increased scale on which they operate. Public health interventions tend to be particularly complex. A common issue with public health systematic reviews is that they do not report sufficient detail on intervention content or context, meaning that important information on factors which may modify the impacts of the intervention is lost [20]. This is to some extent unavoidable when reviewers attempt to include data from multiple studies in one review. However, it is magnified still further in any umbrella review. Similarly, encompassing the heterogeneity of public health intervention studies is challenging for the systematic reviewer – multiple interventions, populations, outcomes, and so forth – which again is escalated within an umbrella review. There are particular problems in terms of getting balance between being totally overwhelmed in terms of question breadth and not losing vital nuance in terms of understanding. Finally, many of the systematic reviews failed to adequately describe the results of their included primary studies or the interventions under evaluation or relied on very broad and vague descriptions.

While recommended appraisal criteria assess the quality of the included systematic reviews, they do not take account of the quality of the studies included in the original systematic review. This can have a major impact on the robustness of the reviews' findings and can mean that a well-conducted systematic review which includes studies of low quality or at high risk of bias will score as highly as a systematic review which includes only well-conducted randomised controlled trials. Hence, conclusions or recommendations based on these quality judgements may

give undue weight to studies of low quality. A means of appraising the included primary studies and their influence on the robustness of the reported findings would be a useful contribution to improving the evidence derived from umbrella reviews. Umbrella reviews also need to start incorporating the quality of included systematic reviews in their interpretation of findings.

We have focused our chapter on case studies from a specific area of public health research – health inequalities and the social determinants of health. There are some common topic themes that come out of this body of work too which will be briefly reflected upon here. Firstly, in all case studies, there is a noticeable lack of systematic reviews that examine the effects of public health interventions on health inequalities (as opposed to just public health in general). Secondly, in methodological terms there are many commonalities between the case study umbrella reviews – as shown in Table 20.1. For example, the quality appraisal tools used and the definition of what constitutes a systematic review (as opposed to just a traditional literature review or a structured review) are also shared. This is of course partly due to the fact that the case studies all involve the work of just two research teams based in the Universities of Durham and Glasgow of which we are both members. Another issue with a number of the case studies is that often only a few databases were searched. This is because umbrella review methodology is often employed to be used as a quick way of surveying the research landscape and providing quick evidence-based responses to time-sensitive public health policy or practice-driven questions. Future development of umbrella reviews in this sub-discipline will need to balance off these tensions.

## Conclusion

This chapter has summarised some of the first umbrella reviews conducted in the field of public health with a thematic focus on the social determinants of health and how interventions might affect health inequalities. It has discussed some of the cross-cutting methodological and thematic lessons learned from this body of work. In terms of new directions for umbrella reviews within this field, the case studies suggest a number of areas for potential methodological development of umbrella reviews in the future including: how umbrella review methodology may need to develop in order to accommodate non-experimental designs, new ways of developing search strategies, assessing implementation of interventions within umbrella reviews, and the potential to extend the critical appraisal undertaken by umbrella reviews to include the quality of the studies included in the original systematic review. However, the future development of umbrella review methodology will need to balance off tensions between methodological refinement and maintaining the role of umbrella reviews in providing a summary of the evidence base. The use of umbrella reviews in public health is likely to grow especially since the publication of the first Cochrane Public Health Group review in 2015 [3].

**Table 20.1** Methodological summary of umbrella reviews on health inequalities and the social determinants of health

| Study and topic | Objective | Included study design(s) | Search strategy | Quality appraisal | Synthesis methods |
|---|---|---|---|---|---|
| Bambra et al. (2014) Health-care equity | To conduct an umbrella review of the evidence of the effects of organisational and financial health system interventions on equity of health care | Systematic reviews of intervention studies with quantitative outcomes (experimental and observational) Reviews were defined as "systematic" if they met the two mandatory criteria of Database of Abstracts of Reviews of Effects (DARE) | 7 electronic databases English language Start to January 2013 Citation follow-up was conducted on the bibliographies of included studies | Conducted by two independent reviewers 7-point adapted DARE Quality criteria checklist[a] Low (met 0–3 criteria), medium (4–5), or high (6–7) quality | Narrative synthesis by intervention subtype |
| Footman et al. (2014) Health-care quality | To conduct an umbrella review of the evidence of the effects of organisational and financial health system interventions on health-care quality | Systematic reviews of intervention studies with quantitative outcomes (experimental and observational) Reviews were defined as "systematic" if they met the two mandatory criteria of Database of Abstracts of Reviews of Effects (DARE) | 7 electronic databases English language Start to January 2013 Citation follow-up was conducted on the bibliographies of included studies | Conducted by two independent reviewers 7-point adapted DARE Quality criteria checklist[a] Low (met 0–3 criteria), medium (4–5), or high (6–7) quality | Narrative synthesis by intervention subtype |

(continued)

**Table 20.1** (continued)

| Study and topic | Objective | Included study design(s) | Search strategy | Quality appraisal | Synthesis methods |
|---|---|---|---|---|---|
| Cairns et al. (2014) Transport and health | To identify systematic reviews of the effects of 20mph zones and limits on health and inequalities in health amongst adults and children | Systematic reviews of intervention studies with quantitative outcomes Reviews were defined as "systematic" if they met the two mandatory criteria of Database of Abstracts of Reviews of Effects (DARE) | 12 electronic databases English language 1990 to September 2013 Grey literature and relevant websites | Conducted by one with checking by second 7-point adapted DARE Quality criteria checklist[a] Low (met 0–3 criteria), medium (4–5), or high (6–7) quality | Narrative synthesis |
| Gibson et al. (2014) Pathways linking housing to health | To identify systematic reviews of housing and neighbourhood interventions which target internal housing conditions, area characteristics, or housing tenure and measure impacts on health and health inequalities | Systematic reviews of intervention studies with quantitative outcomes Reviews were defined as "systematic" if they met the two mandatory criteria of Database of Abstracts of Reviews of Effects (DARE) | 6 electronic databases. English language 2000–2007 Relevant websites, bibliographies, and journals. Experts contacted | Conducted by one with checking by second 7-point adapted DARE Quality criteria checklist[a] Low (met 0–3 criteria), medium (4–5), or high (6–7) quality | Narrative synthesis by pathway of effect |

| | | | | | |
|---|---|---|---|---|---|
| Bambra et al. (2009) Workplace interventions and health | To systematically review studies reporting the impacts on health and health inequalities of workplace interventions aimed at psychosocial working conditions delivered at an organisational level | Systematic reviews of intervention studies with quantitative outcomes Reviews were defined as "systematic" if they met the two mandatory criteria of Database of Abstracts of Reviews of Effects (DARE) | 6 electronic databases English language 2000–2007 Relevant websites, bibliographies, and journals. Experts contacted | Conducted by one with checking by second 7-point adapted DARE Quality criteria checklist[a] Low (met 0–3 criteria), medium (4–5), or high (6–7) quality | Narrative synthesis by intervention subtype |
| Bambra et al. (2010) Health impacts of interventions aimed at social determinants of health | To systematically review studies reporting the impacts on health and health inequalities of workplace interventions aimed at psychosocial working conditions delivered at an organisational level | Systematic reviews of intervention studies with quantitative outcomes Reviews were defined as "systematic" if they met the two mandatory criteria of Database of Abstracts of Reviews of Effects (DARE) | 6 electronic databases English language 2000–2007 Relevant websites, bibliographies, and journals. Experts contacted | Conducted by one with checking by second 7-point adapted DARE Quality criteria checklist[a] Low (met 0–3 criteria), medium (4–5), or high (6–7) quality | Narrative synthesis by social determinant domain and intervention type |

[a]Criteria: (1) is there a well-defined question; (2) is there a defined search strategy; (3) are inclusion/exclusion criteria stated; (4) are study designs and number of studies clearly stated; (5) have the primary studies been quality assessed; (6) have the studies been appropriately synthesised; (7) has more than one author been involved in each stage of the review process?

## References


1. Becker L, Oxman A. Overviews of reviews. In: Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions. British Medical Journal, London. 2008; p. 607–32.
2. Bambra C, Gibson M, Sowden A, Wright K, Whitehead M, Petticrew M. Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews. J Epidemiol Community Health. 2010;64:284–91.
3. Baker P, Dobbins M, Soares J, Francis D, Weightman A, Costello J. Public health interventions for increasing physical activity in children, adolescents and adults: an overview of systematic reviews. Cochrane Library. 2015. doi:10.1002/14651858.CD011454.
4. Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health. British Medical Journal, London; 2003;57:527–9.
5. Joyce K, Pabayo R, Critchley JA, Bambra C. Flexible working conditions and their effects on employee health and wellbeing. Cochrane Library. 2010;2.
6. Perkins N, Smith KE, Hunter DJ, Bambra C, Joyce KE. 'What counts is what works'? New labour and partnerships in public health. Policy Polit. 2010;38:101–17.
7. Bambra C. Reviewing the evidence: reflections from experience. Evid Policy. 2005;1:241–53.
8. Ogilvie D, Hamilton V, Egan M, et al. Systematic reviews of health effects of social interventions: 1. Finding the evidence: how far should you go? J Epidemiol Community Health. 2005;59:804–8.
9. Cairns J, Warren J, Garthwaite K, Greig G, Bambra C. Go slow: an umbrella review of the effects of 20mph zones and limits on health and health inequalities. J Public Health. 2014. doi:10.1093/pubmed/fdu067.
10. Gibson M, Petticrew M, Bambra C, Sowden A, Wright K, Whitehead M. Housing and health inequalities: a synthesis of systematic reviews of interventions aimed at different pathways linking housing and health. Health Place. 2011;17:175–84.
11. Footman K, Garthwaite K, Bambra C, McKee M. Quality check: does it matter for quality how you organise and pay for health care? a review of the international evidence. Int J Health Serv. 2014;44:479–505.
12. Bambra C, Garthwaite K, Hunter D. All things being equal: does it matter for equity how you organise and pay for health care? a review of the international evidence. Int J Health Serv. 2014;44:457–72.
13. Bambra C, Gibson M, Sowden A, Wright K, Whitehead M, Petticrew M. Working for health? Evidence from systematic reviews on the health effects of changes to the psychosocial work environment. Prev Med. 2009;48:454–61.
14. Marmot M. Introduction. In: Marmot M, Wilkinson RG, editors. Social determinants of health. Oxford: Oxford University Press; 2006. p. 1–5.
15. Whitehead M. A typology of actions to tackle social inequalities in health. J Epidemiol Community Health. 2007;61:473–8.
16. Graham H, Kelly MP. Health inequalities: concepts, frameworks and policy. London: Health Development Agency; 2004. Available at: http://www.nice.org.uk/aboutnice/whoweare/aboutthehda/evidencebase/keypapers/papersthatinformandsupporttheevidencebase/health_inequalities_concepts_frameworks_and_policy__briefing_paper.jsp. Accessed 24 Mar 2015.
17. Bambra C, Hillier F, Cairns J, Kasim A, Moore H, Summerbell C. How effective are interventions at reducing socioeconomic inequalities in obesity among children and adults? Two systematic reviews. Public Health Res. 2015;3(1).
18. Bambra C. Work, worklessness and the political economy of health. Oxford/New York: Oxford University Press; Oxford 2011.
19. Barr B, Bambra C, Smith KE. For the good of the cause: generating evidence to inform social policies that reduce health inequalities. In: Smith KE, Hill S, Bambra C, editors. Health inequalities: critical perspectives. Oxford: Oxford University Press; 2015.
20. Egan M, Bambra C, Petticrew M, Whitehead M. Reviewing evidence on complex social interventions: appraising implementation in systematic reviews of the health effects of organisational-level workplace interventions. J Epidemiol Community Health. 2009;61:4–11.

# Part IV

# Implementation and Further Research

# Moving from Evidence to Action

# 21

Fabrizio D'Ascenzo, Claudio Moretti, Christian Templin, and Fiorenzo Gaita

**Abstract**

Decision making should always be based on the best evidence while also taking into account stakeholder values and resources. The traditional hierarchy of evidence puts at the uppermost level a large randomized trial or a meta-analysis of homogeneous randomized trials. This approach is atomically correct but fails to capture the complexity and comprehensiveness of evidence sources. Umbrella reviews, overviews of reviews, and meta-epidemiologic studies offer a novel tool to summarize and appraise clinical evidence at a level which is even more general than that of meta-analyses. Knowledge is an essential prerequisite of effective action, but if such exercises in evidence synthesis are to be truly meaningful, their impact on decision making must actually translate into specific actions. To enable this, while empowering all stakeholders, it is crucial to search appropriately for umbrella reviews, to correctly appraise them, and to correctly grade them in terms of pragmatic impact.

*Come, let us go down and there confuse their language, so that they may not understand one another's speech.*

Genesis 11, 1–9 (Holy Bible)

F. D'Ascenzo, MD (✉) • C. Moretti • F. Gaita
Dipartimento di Scienze Mediche, Divisione di Cardiologia, Città della Salute e della Scienza, Corso Bramante 88-90, Turin 10126, Italy
e-mail: fabrizio.dascenzo@gmail.com

C. Templin
Department of Cardiology, University Heart Center, University Hospital Zürich, Zuerich, Switzerland



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_21

## 21.1 Introduction

The translation from paper (or notably in 2015 from smartphone) of evidence to everyday clinical practice represents a challenge for busy physicians around the world, and like in the episode of the Tower of Babel, this may lead to complete misunderstanding between researchers and physicians.

This comes even truer for complex pieces of evidence, as meta-analysis of randomized controlled trials (RCTs), of observational studies, or of umbrella reviews.

Umbrella reviews, actually, may be seen as a summary of previous evidence, including also meta-analysis already performed. This may be particular relevant, because common "medical readers" are often crowded by many information, which may be misleading or potentially biased at different levels. The role of umbrella review is consequently essential, aiming to describe and to correctly grade different levels of evidence.

In the present chapter we will aim to describe: when search for umbrella review is needed, how to correctly appraise them, and how to correctly grade them.

## 21.2 When Search for Umbrella Review Is Needed

When do we need to add more to pairwise/network meta-analysis? For physicians taking practical decisions in everyday work, contrasting results coming from randomized evidence, that is, the one with the highest level of grading, may be hard to face. Basically medical doctors should search for umbrella review in cases of contrasting results for different meta-analyses or presence of low-grade evidence. For example, in a recurrent way, newspapers claim alarms for risk of cancer induced by different drugs, which is a question often asked by patients, especially for those assuming chronic therapy [1].

In this field a large number of observational studies have been conducted, as subanalysis stemming from randomized controlled trials. Despite the large amount of evidence, discording data have been proposed [2–6], even according to different meta-analyses. In this setting, an umbrella review was conducted by Ioannidis et al.

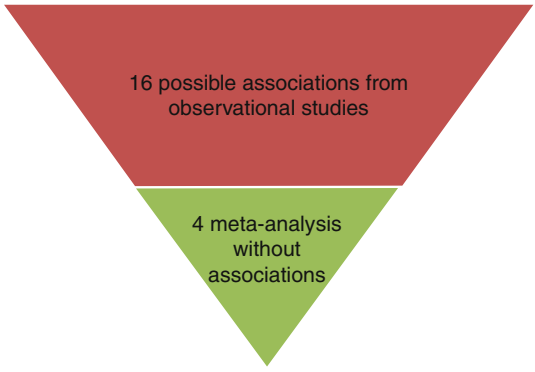



**Fig. 21.1** "Degrading" of evidence

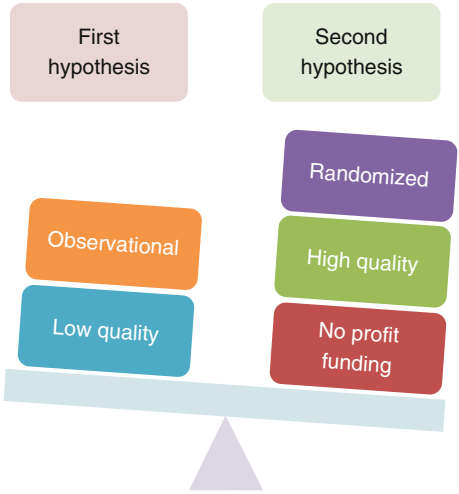



**Fig. 21.2** Weight of evidence

[7] aiming to grade the available evidence or more correctly to "degrade" it. Actually as shown in Fig. 21.1, the number of possible associations shown with observational studies decreased and disappeared after evaluation with meta-analysis.

But how often does this happen? That is, how often meta-analyses are not concordant among them? Or in contrast with randomized controlled trials? A classic example is the paper of Biondi-Zoccai et al. [8] which showed that on a relatively limited topic (i.e., use of acetylcysteine to prevent contrast-associated nephropathy) of the ten systematic reviews analyzed, five advocated use of acetylcysteine, while the others suggested further research on this setting. Poignantly the paper also presented two of the three major issues which should be considered when evaluating these issues, especially in cases of contrasting results (Fig. 21.2):

- Presence or not of randomization
- Quality of data
- Presence or not of funding

Meta-analyses derived by observational studies, although widely published, are often prone to many confounding factors. Actually even if combining results from multivariate analysis, they are still not of the same level of randomized evidence, that is, combination of data with as few as possible confounders. A practical example is represented by the paper of Vlaar et al. [9] which tried to pool all available evidence about multivessel revascularization in patients presenting with ST segment elevation myocardial infarction. Despite being published on a relevant journal and by a research group with great experience in this field, the main message of increased harm for complete revascularization was overestimated due to presence of observational data, while data derived from randomized controlled trials (RCTs) showed neutral effect [10]. This result was confirmed after inclusion of more RCTs [11], with a result potentially impacting everyday management of these patients in the cath lab.

Quality of data represents another burning issue in this field. Apart from inclusion or not of randomized evidence only, assessment of prospective of retrospective design, of multicenter or not study, and of presence and relevance of all the bias related to conduction of the study should be always reported. The PRISMA group reported different scales according to kind of study [12] and suggested to conduct sensitivity analysis for low vs. good quality of included evidence.

Last but not least, the association between presence of funding and probability of positive results of the evaluated intervention has been widely described [13, 14]: this does not mean, however, that readers should not trust these results, but this simply represents another issue to be evaluated.

## 21.3 How to Correctly Appraise

Umbrella reviews, sometimes, may be apparently difficult to be understood. The recent paper of the group of Ioannidis [15] on use of vitamin D concluded "despite a few hundred systematic reviews and meta-analyses, highly convincing evidence of a clear role of vitamin D does not exist for any outcome, but associations with a selection of outcomes are probable," a sentence which may appear hard to be translated in real life. On the other hand, for other topics they appear to convey a clear and useful message, for example, in this paper appraising usefulness of drugs between men and women [16] "Overall, for the majority of drugs sex does not appear to be a factor that has to be taken into consideration when choosing a drug treatment."

Basically, as for all literature [17], at least two levels of interpretations may be advocated for umbrella review (Fig. 21.3). The first is the one regarding institutional health system or economic board of medical companies, which aims to evaluate and to accurately appraise the exact level of evidence of a single intervention, in order to correctly plan financial plans or exposure. For them, for example, the first paper

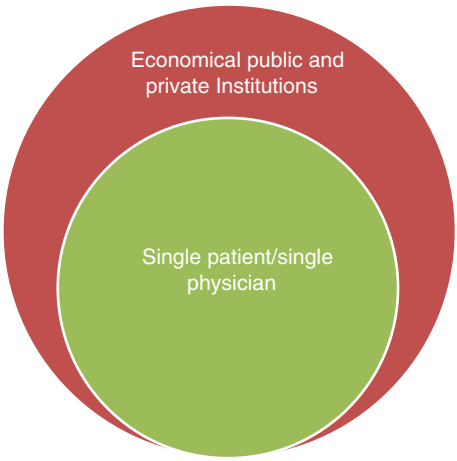



**Fig. 21.3** Different levels to appraise an umbrella review

about vitamin D may be of relevance, describing an area with limited evidence and with potential interest, while for single physician or patient, it may be less useful. On the other setting, if read carefully, the second example may be of worth also in clinical practice, aiming to furnish a practical tool, that is, the drug a physician gives to his/her patient will work the same, independently from gender.

## 21.4 How to Correctly Grade

As for every piece of evidence, also umbrella review should be carefully evaluated. They rely, probably more than any other single paper, on quality of included meta-analysis and of included studies. For common readers, the following points should always be checked:

- Was the search strategy explicitly explained in the text? And performed by at least two researchers?
- How was data extraction performed?
- Did the included meta-analysis follow the simple rule of use of random effect for moderate to high heterogeneity? This has been largely advocated by the Cochrane Collaboration, as use of fixed effect (i.e., considering all studies evaluating patients from the same population) in studies with high heterogeneity may lead to an increased risk of "false-positive" results, that is, of significant results not correctly evaluated [18]. As widely described, use of fixed effect in case of high heterogeneity is misleading [19].

In cases of meta-analysis with use of data derived both from RCTs and observational studies, three rules should be followed:

- First, sensitivity analysis according to kind of studies should also be reported.
- Second logarithmic transformation of the measure of risk (like hazard ratio) should have been performed. This is fundamental, as the normal scale is not symmetric, starting from 0 as lower benefit, with 1 as neutral effect, and leading to infinite. On the contrary, the log transformation makes this scale symmetric, because the log of 0 is minus infinity, that of 1 is 0, and that of infinity is infinity [20, 21].
- Third in case of observational studies, only results derived from multivariate analysis should be reported. Actually observational studies are prone to many biases, which may partly (and only partly) be reduced by multivariate analysis, which however cannot adjust for unknown confounders [22, 23].
- If the umbrella review deals with this topic, critically appraising and evaluating them, it should be trusted as a precious aim for taking decision in terms of interventions and/or clinical exams. If these aspects have not been evaluated, it should be weighted as a relevant piece of evidence, but with internal point of weakness.

## References


1. ACE inhibitors may increase risk of recurrence in breast cancer survivors. Available at: http://www.eurekalert.org/pub_releases/2011-04/uoc--aim041911.php. Last accessed on 31 July 2015.
2. Soranna D, Scotti L, Zambon A, et al. Cancer risk associated with use of metformin and sulfonylurea in type 2 diabetes: a meta-analysis. Oncologist. 2012;17:813–22.
3. Taylor ML, Wells BJ, Smolak MJ. Statins and cancer: a meta-analysis of case control studies. Eur J Cancer Prev. 2008;17:259–68.
4. Corrao G, Scotti L, Bagnardi V, et al. Hypertension, antihypertensive therapy and renal-cell cancer: a meta-analysis. Curr Drug Saf. 2007;2:125–33.
5. Yoon C, Yang HS, Jeon I, et al. Use of angiotensin-converting-enzyme inhibitors or angiotensin-receptor blockers and cancer risk: a meta-analysis of observational studies. CMAJ. 2011;183:E1073–84.
6. Tressler CS, Wiseman RL, Dombi TM, et al. Lack of evidence for a link between latanoprost use and malignant melanoma: an analysis of safety databases and a review of the literature. Br J Ophthalmol. 2011;95:1490–5.
7. Ioannidis JP, Zhou Y, Chang CQ, Schully SD, Khoury MJ, Freedman AN. Potential increased risk of cancer from commonly used medications: an umbrella review of meta-analyses. Ann Oncol. 2014;25(1):16–23.
8. Biondi-Zoccai GG, Lotrionte M, Abbate A, Testa L, Remigi E, Burzotta F, Valgimigli M, Romagnoli E, Crea F, Agostoni P. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: case study. BMJ. 2006;332(7535):202–9.
9. Vlaar PJ, Mahmoud KD, Holmes Jr DR, van Valkenhoef G, Hillege HL, van der Horst IC, Zijlstra F, de Smet BJ. Culprit vessel only versus multivessel and staged percutaneous coronary intervention for multivessel disease in patients presenting with ST-segment elevation myocardial infarction: a pairwise and network meta-analysis. J Am Coll Cardiol. 2011;58(7):692–703.
10. Biondi-Zoccai G, Agostoni P, Abbate A, D'Ascenzo F, Modena MG. Potential pitfalls of meta-analyses of observational studies in cardiovascular research. J Am Coll Cardiol. 2012;59(3):292–3.
11. Moretti C, D'Ascenzo F, Quadri G, Omedè P, Montefusco A, Taha S, Cerrato E, Colaci C, Chen SL, Biondi-Zoccai G, Gaita F. Management of multivessel coronary disease in STEMI patients: a systematic review and meta-analysis. Int J Cardiol. 2015;179:552–7.
12. http://www.prisma-statement.org/.
13. Wu W, Tong Y, Zhao Q, Yu G, Wei X, Lu Q. Coffee consumption and bladder cancer: a meta-analysis of observational studies. Sci Rep. 2015;5:9051.
14. Yaphe J, Edman R, Knishkowy B, Herman J. The association between funding by commercial interests and study outcome in randomized controlled drug trials. Fam Pract. 2001;18(6):565–8.
15. Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JP. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ. 2014;348:g2035.
16. Gartlehner G, Chapman A, Strobelberger M, Thaler K. Differences in efficacy and safety of pharmaceutical treatments between men and women: an umbrella review. PLoS One. 2010;5(7):e11895.
17. Eco U. Semeiotica e filosofia del linguaggio. 1994.
18. Incorporating heterogeneity into random-effects models. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011. Available at: http://handbook.cochrane.org/chapter_9/9_5_4_incorporating_heterogeneity_into_random_effects_models.htm. Last accessed on 31 July 2015.

19. Melsen WG, Bootsma MC, Rovers MM, Bonten MJ. The effects of clinical and statistical heterogeneity on the predictive values of results from meta-analyses. Clin Microbiol Infect. 2014;20(2):123–9.
20. Keene ON. The log transformation is special. Stat Med. 1995;14(8):811–9.
21. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011. Available at: www.cochrane-handbook.org. Last accessed on 31 July 2015.
22. Biondi-Zoccai G, Romagnoli E, Agostoni P, Capodanno D, Castagno D, D'Ascenzo F, Sangiorgi G, Modena MG. Are propensity scores really superior to standard multivariable analysis? Contemp Clin Trials. 2011;32(5):731–40.
23. D'Ascenzo F, Cavallero E, Biondi-Zoccai G, Moretti C, Omedè P, Bollati M, Castagno D, Modena MG, Gaita F, Sheiban I. Use and misuse of multivariable approaches in interventional cardiology studies on drug-eluting stents: a systematic review. J Interv Cardiol. 2012;25(6):611–21.

# Avenues for Further Research

# 22

Dawid Pieper, Lun Li, and Roland Brian Büchter

**Abstract**

Overviews of reviews represent a new publication type and a new form of evidence synthesis. They have rapidly gained popularity. The development of their methodology is, however, still in its infancy. We present a bundle of areas where more work is needed in order to make overviews of reviews more valuable and reliable. Firstly, a clear-cut definition of an overview of reviews is needed. Secondly, methods of presenting the results of overviews of reviews need to be further developed. The needs of groups of users should be kept in mind, in this context, including clinicians, patients, and political decision makers. Maintaining a reasonable balance between necessary complexity and an inevitable loss of information from the reviews is a major challenge. A registration of overviews of reviews is called for. All overviews should be gathered in one freely available registry. When registering new overviews of reviews, a special note should be given to the conflicts of interests of the authors. Reporting guidelines for overviews of reviews should be prepared as soon as possible as this area lacks standardization. Furthermore, more attention should be given to different types of overviews of reviews (e.g., comparison of interventions, comparison of populations).

D. Pieper (✉)
Institute for Research in Operative Medicine (IFOM), Evidence-Based Health Services Research, Witten/Herdecke University, Ostmerheimer Str. 200, 51009 Cologne, Germany
e-mail: dawid.pieper@uni-wh.de

L. Li
Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China

R.B. Büchter
Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies,* DOI 10.1007/978-3-319-25655-9_22

## 22.1 Preface and Foreword

This chapter on further avenues for research will provide a showcase to anticipate, to the best of our knowledge, what an overview of reviews will be able to be and do and what further research should focus on.

In the first section we argue for the development of broadly accepted definitions for overviews of reviews. In the context of this chapter, we will use the term overview of reviews in a general sense, with umbrella reviews being a specific type of overviews of reviews focusing on a specific clinical topic (e.g., a drug or a condition), while meta-epidemiological studies deal with a nonclinical topic (e.g., mainly methodological issues such as funding issues or small study effects). The next section discusses potential users of overviews of reviews with different interests. This is followed by a section dealing with issues of registering overviews of reviews and incorporating all of them into one database. Methodological issues are considered in the final sections, including the reporting of overviews, the issue of quality assessment of overviews, conflicts of interests in overviews, and the presentation of results. A conclusion summarizes the chapter at the end.

## 22.2 Standardized Terms and Definition(s) Needed

Talking about a topic without a standardized term and a clear understanding or definition is always problematic, and this is also true for overviews of reviews. There is a wide choice of synonyms such as umbrella review, scoping review, meta-review, review of reviews, review of systematic reviews, overview of systematic reviews, systematic review of systematic reviews, and so forth [1]. All of these examples, among others, can be found in the literature. It should be mentioned that the term overview has also been used as a synonym for systematic reviews.

It remains unclear whether the use of different names reflects different methods or whether differences are just semantic. For example, there might be good reasons to differentiate between overviews of reviews and overviews of systematic reviews. The latter will include only systematic reviews, while the first might also include non-systematic reviews (i.e., narrative reviews). Although in general systematic reviews should be preferred, it is important to acknowledge that systematic reviews are the standard when investigating effectiveness or efficacy (i.e., reviews of interventions), but epidemiological reviews (i.e., reviews of risk factors) may often not be considered systematic, as many of them do not assess the quality of the included studies. This example illustrates that there might be a ground for different names. Thus, what is needed is a general term for overviews of reviews with further possible breakdowns.

There is no broadly accepted definition for overviews of reviews. As also highlighted in prior chapters of this book, we have struggled with this when we defined an overview "as a synthesis of systematic reviews on the same or a similar topic and/or intervention that have been derived through a systematic literature search" [2]. This would, however, exclude overviews of reviews that also search for primary

studies, an approach which is often utilized in the conduct of clinical practice guidelines. There might also be hybrids, comparing two interventions and searching for both reviews: first one might end up with reviews for one intervention, but not for the second intervention, and primary studies will be searched for this intervention instead. Will this still be called an overview of reviews? There is still much room for discussion. However, it is inevitable to try to end up with a clear definition. A clear definition will facilitate discussions and foster further developments of this publication type. It will also facilitate more standardized reporting, which will make it much easier to identify overviews in the literature. Currently, it can be expected that many overviews are not recognized.

Nevertheless, it is not only about finding proper terms for overviews of reviews but also clearly defining them as there is probably not a clear-cut scope, and some questions still need to be answered in this process.

## 22.3 Balancing the Needs of Different Users

We do not know who reads overviews of reviews. However, it is very likely that there is more than one group of users and each group has different needs. In particular, the presentation of the results might be strongly dependent on the target group. For example, it is widely recognized that clinicians have different needs than, for example, consumer representatives or policy makers, while the findings of an evidence synthesis can be of interest to all of them. The presentation of findings from systematic reviews is an emergent research field.

It is supposed that overviews of reviews can be a valuable source for clinicians. However, this might also be very much dependent on the question under study. It has to be taken into account that clinicians are often only interested in research focusing on their subspecialty and their daily practice. Comparing different interventions where different professional groups are responsible for one intervention might be too broad, although informative, for clinicians. For example, an overview of reviews tried to identify effective interventions for the prevention of suicidal behavior [3]. The authors found training of general practitioners (GPs) to recognize and treat depression and suicidality, improving accessibility of care for at-risk people, and restricting access to means of suicide to be effective. These interventions include individual-level and population-level interventions. This might raise the question about the target audience of this overview. There are certainly elements that are relevant for GPs, while other aspects will be of use for policy makers. It might be discussed whether such different interventions working at different levels should be collated in an overview of reviews. For sure, we need to know better about the target audience of overviews of reviews and their needs to this new publication type.

Another (personal) example of an overview that was proposed to a Cochrane group on open versus laparoscopic surgical procedures illustrates the difficulty of conducting broad overviews: laparoscopic surgery has been used for nearly 20 years, and many reviews have been published comparing open with laparoscopic surgery.

Laparoscopic procedures superseded open surgery in many areas, but evidence showing laparoscopic approaches to be beneficial over open surgery is not always present. Therefore, we proposed a broad overview on laparoscopic versus open surgery. The idea was found to be of general interest, but the project was not taken on because of doubts about who would be interested in such a broad overview, since clinicians and patients would still use the evidence for their particular condition in decision making. The argumentation can be fully followed. Nevertheless, from a general or scientific point of view, this might still be an interesting project as much could be learned from an example where the new technology (laparoscopic surgery) is shown not to be superior to the old technology, although, nowadays, laparoscopic surgery is performed regularly without questioning its benefit compared to open surgery.

Overviews of reviews are also used to support health policy decisions. There are examples where agencies for health technology assessment (HTA), such as the Institute for Quality and Efficiency in Health Care (IQWiG, Germany), the Medical Services Advisory Committee (MSAC, Australia), or the Canadian Agency for Drugs and Technologies in Health (CADTH), performed reports, and these reports were overviews of reviews instead of the more common approach of a systematic review of primary studies [4]. The main reason for choosing this approach is the idea that overviews of reviews can be produced more quickly as they rely on already published reviews [5]. Decision makers as the consumers of HTAs are interested in quick decisions, since the timely delivery of HTAs is known to be crucial in order that the gathered evidence also comes into practice (e.g., decisions are failed in correspondence with the HTA) [6].

It might also be questioned whether overviews of reviews can serve as a viable source for patients and support decision making (e.g., treatment decisions). Evidence syntheses in overviews of reviews have to collate a huge amount of data. This can cause complexity and make it difficult to use them in decision making. Nevertheless, it is common that decision aids can also be produced based on secondary research. The patient information produced by IQWiG also relies mainly on systematic reviews, if available [7].

## 22.4 Registration

The prospective registration of clinical trials has become a mandatory research step, today. Registering clinical trials when they start, updating relevant information, and providing summary results while making all of them publicly available have become a prerequisite [8]. The same can be said for systematic reviews and other study designs (e.g., cohort studies). The registration of overviews of reviews allows an assessment of the methodological quality, provides transparency in conducting the overview of reviews, and minimizes the risk of publication bias or selective reporting of outcomes [9]. The registration of an overview of reviews should usually come along with an available protocol.

Currently, Cochrane overviews of reviews need to be registered, following the same registration steps as for Cochrane systematic reviews. At the time of writing

(April 2015), 26 protocols of Cochrane overviews of reviews can be found in the Cochrane Database of Systematic Reviews (CDSR) via the Cochrane Library, although five of them have been withdrawn. In addition, 21 published Cochrane overviews of reviews (with their corresponding protocols) can be found in the CDSR. Overviews of reviews can also be registered in PROSPERO if they meet the standard eligibility criteria (as detailed elsewhere). PROSPERO serves as an international database of prospectively registered systematic reviews in health sciences [10]. Up to now, PROSPERO only includes reviews (i.e., systematic reviews or overviews of reviews) on the effects of interventions. Other reviews (e.g., reviews on risk factors) are expected to be included in PROSPERO in the future. A number of published protocols are not determinable as overviews of reviews are not yet separately listed.

## 22.5 Database of Overviews

Registering overviews of reviews will also make it much easier to identify them. But it would also be worth moving some steps further making the database user friendly as much as possible. It would be beneficial if users could search for topics, diseases, interventions, and so on within such a database. Quality ratings of the included overviews of reviews should be incorporated in this database, too. Beside the advantage of an easier identification, the database could also serve as a starting point for methodological projects on overviews of reviews. Nowadays, researchers are often reluctant to do this sort of work as it is a very time-consuming task to identify overviews.

A good example for such a database is the healthevidence.org database (http://www.healthevidence.org/) commissioned by the McMaster University in Canada. This database collects quality-assessed systematic reviews investigating the effectiveness of public health interventions. Funding for the database is provided by several organizations with an interest in public health.

## 22.6 Methods: One Size Fits All?

Although the term overview of reviews is still ill- defined, it is generally considered to be based on secondary research. However, primary research might also be included in an overview of reviews. Debates about appropriate methods, usefulness or validity, and other factors concentrate on the fact that reviews serve as the basis of an overview of reviews. This is definitely true, but this approach has the major drawback that debates only focus on methodological aspects while losing sight of the topic under study.

There are several reasons why an overview of reviews might be conducted resulting in different objectives. According to the *Cochrane Handbook*, there are five types of overviews of reviews [11]. However, not all of them are suitable for a Cochrane overview of reviews. Simply put, one can think of the PICO (population,

intervention, comparison, outcome) scheme where (at least) one component can vary while the remaining components remain constant. This idea is helpful in order to imagine what kinds of overviews of reviews are possible.

First of all, one can look at different populations or conditions. Thus, the objective of such an overview of reviews is a summary from more than one review of the same intervention for different populations or conditions according to the Cochrane Collaboration. However, as the Cochrane Collaboration focuses primarily on interventions, one could also think of exposures (e.g., risk factors) in general. Our research group has conducted an overview of reviews on the hospital volume-outcome relationship in surgery [12]. We were able to define surgical procedures where the relationship is strongest. This overview of reviews is of a huge importance for health policy decisions. In the German health care system, minimum volume thresholds exist. They define a minimum number of procedures a hospital has to perform per year in order to be able to perform this procedure in the next year. On the one hand, our overview of reviews demonstrated that there were areas with evidence for a hospital volume-outcome relationship where no minimum volume thresholds existed. On the other hand, our overview of reviews identified areas with minimum volume thresholds, although there was no evidence for a hospital volume-outcome relationship.

Another example is an overview of reviews dealing with the efficacy of adjuvant chemotherapy after surgery in cancer [13]. The objective of this overview of reviews was to investigate differences in the efficacy among different cancer types. Such investigations can be useful to identify deviating results that cannot be explained by biological reasons and should therefore undergo further investigations. We refer to an article published on this type of overviews of reviews where more examples and issues for methodological debates can be found [14].

Strongly related to this type of overviews of reviews are overviews of reviews aiming to summarize evidence about adverse effects as they also compare reviews over multiple populations or conditions. The only difference is that the outcome (adverse events) is clearly defined. Adverse effects of biologics were investigated in one of the first Cochrane overviews of reviews [15]. Furthermore, the evidence synthesis was performed applying a network meta-analysis. Anticholinergic adverse effects of antipsychotic drugs were investigated in another review [16]. However, the search for systematic reviews was only performed to identify relevant trials, and data were analyzed at the study level. However, it is also possible to perform the evidence synthesis at the review level [17].

The third type of overviews of reviews focuses on different outcomes for a given population or condition and intervention. This is especially important in the case of multiple reviews dealing with different outcomes (e.g., mortality, quality of life). Such differences are very obvious, but the devil can also be in the detail. For example, two reviews of massage for chronic unspecific low back pain yielded different results, although the reviews performed meta-analyses for pain as the outcome measure of almost the same trials [18, 19]. The main reason for the discordance between them is that data from different assessment points were used. One of the reviews used data from the pretreatment assessment, and the other used data from the

posttreatment assessment. This difference is not apparent when simply reading both reviews due to the reporting.

Probably the most common and important type of overview of reviews is focused on comparing different interventions in a given population. The advantage of such overviews is they allow for indirect comparisons of interventions, especially where head-to-head trials are lacking. For example, several interventions exist for enuresis, and there are Cochrane reviews available for each single intervention, like alarm systems [20], medications [21–23], complementary medicine [24], and behavioral interventions [25]. From a theoretical side of view, a clinical epidemiologist would suggest to conduct a five-arm (four arms for each intervention plus a placebo group) randomized trial. However, such trials are unlikely to be conducted. Nevertheless, one is able to compare the interventions applying network meta-analyses. Although it is more common to conduct network meta-analyses at primary study level, there are also methods available for conducting such analyses at the review level [25], making this approach very interesting when conducting overview of reviews. Further analyses at the review level based on primary study data can also be conducted later, if needed.

Lastly, there is also the possibility to conduct an overview of reviews where all reviews aim to answer the same research question (i.e., the PICO scheme remains constant). The huge amount of reviews and in particular of multiple reviews for one question has been criticized in the past, and initiatives like the prospective registration of systematic review try to tackle this problem. However, there are many examples of multiple reviews which came to different conclusions and where comparisons between the reviews provided interesting insight. Thus, while multiple reviews may be resource wasting in some cases, they can be valuable in others. In the case of multiple high-quality reviews where all reviews have the same results (i.e., replication of findings), this can be taken as a high degree of reliability. This approach is often chosen by HTA agencies [4]. In the case of discordant findings, one should try to investigate the reasons for discordance. There is already an algorithm for the analysis of discordant reviews available [26]. However, more practical applications are needed here.

It seems unlikely that one methodological approach to conduct overviews of reviews will suit all types of questions. Different types of overviews require different methods and each type should be discussed separately. This will be one of the most important steps in the further development of overviews of reviews.

The need for a variety of methods has also been accepted in systematic reviews, and this might provide some guidance. More generally, a systematic review is defined as a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research and to collect and analyze data from the studies that are included in the review [27]. However, there are differences in methods for systematic review depending on whether they focus on interventions, etiology (i.e., risk factors), prognostic factors, diagnostic accuracy, measurement properties, animal studies, and so on. Furthermore, separate reporting guidelines exist for systematic reviews of randomized controlled trials [27], observational studies [28], or qualitative research among others [29].

## 22.7 Reporting and Quality Assessment

Overviews of reviews are designed to compile evidence from multiple reviews into one accessible and usable document and allow the reader a quick overview of reviews relevant to a specific decision. As overviews of reviews try to bundle a huge amount of evidence, researchers should provide a clear and transparent description of how they conducted their overviews of reviews and what they found [30]. However, the reporting of overviews of reviews was found to be poor. Nearly one in four of the overviews of reviews did not specify the literature search; one in three did not report inclusion and exclusion criteria; half of the overview of reviews did not state how and who selected systematic reviews, abstracted information from included systematic reviews, and assessed the methodological quality; while 80 % of the overviews of reviews did not report the quality of evidence evaluation [31].

This can probably be explained by a lack of a well-recognized and comprehensive reporting checklist for overview of reviews. Authors of overview of reviews do not know which reporting checklist should be followed and what should be reported. For systematic reviews, the PRISMA (preferred reporting items for systematic reviews and meta-analyses) checklist is widely accepted by several well-known organizations, and many journals encourage authors to adhere to it [32]. Overviews of reviews follow a similar methodological approach as systematic reviews. However, it is not clear whether PRISMA could also be used as a reporting guideline for overviews of reviews. Due to different purposes and analysis/description methods between these two kinds of publication types, some items seem to be unsuitable for overviews of reviews, while other aspects (e.g., overlapping reviews) are not covered by PRISMA. This needs to be analyzed and validated in the future.

How are reporting items specific to overviews of reviews? This depends probably on the needs of consumers and the corresponding design of overview of reviews (e.g., different types of overviews of reviews). Clinicians do not only want to know the results and conclusions of included reviews, but they also want to know whether their patients are similar to the patients in the included systematic reviews. This might be a reason why authors of overviews of reviews focused on reporting the characteristics, results, and conclusion of the included systematic reviews and ignored reporting the details of their own study. As a result it is difficult to judge the validity of an overview of reviews. Reporting items should ensure that both content-specific and methodological characteristics are reported in sufficient details to allow readers to both judge the appropriateness of the methods and the applicability of the results. This includes: details on the methodology of the overview of reviews (including literature search, study selection, data abstraction, quality assessment, and evidence level evaluation), review characteristics (including patients, interventions/risk factors, control, and outcomes), and the assessment of the methodological quality of the included reviews.

In a previous study, a pilot version of a reporting checklist has been published [31]. But the purpose of this study was originally the investigation of reporting characteristics of overviews of reviews, instead of developing a reporting checklist. However, to the best of our knowledge, the checklist has not been tested in practice,

and no statements with respect to the reproducibility, face validity, content validity, and construct validity can be made. That is why this checklist cannot be recommended for use, unless tested for validity and reliability. This should also be addressed in the future. The pilot version of the reporting checklist can provide potential guidance for reporting overviews of reviews, but there is definitely a need of a reporting checklist using rigorous developing methods in order to improve the reporting of overviews of reviews [33].

Another issue for overviews of reviews is quality assessment. This has two dimensions. The first concerns the methodological quality of systematic reviews included in overviews in order to determine if the potentially eligible reviews meet the minimum requirements based on quality. Although there are several tools for this purpose, only two tools (AMSTAR [a measurement tool to assess systematic reviews] and the OQAQ [*Overview Quality Assessment Questionnaire*]) have been validated as a means to assess the methodological quality of systematic reviews [5, 34]. AMSTAR, which was published in 2007, is growing in popularity, possibly because of its availability but also because it reflects methodological developments that have occurred after the development of the OQAQ in 1991. The two instruments also differ in that the OQAQ gives an overall quality score to each systematic review [35]. A systematic review found AMSTAR, but not R-AMSTAR, to have good measurement properties [36]. A caveat is that validation studies of these tools were restricted to systematic reviews of randomized controlled trials. Thus, further investigations on systematic reviews of mixed and other study designs, including diagnostic accuracy test studies, etiology studies, and prognostic studies, are needed [36].

The second dimension refers to the methodological quality of the overview of reviews themselves. Assessing the methodological quality (risk of bias) of randomized controlled trials contributes to understanding the quality of evidence level produced by randomized controlled trials; assessing the methodological quality of systematic reviews can help readers understand the risk of bias of a wider body of evidence. It is also necessary to assess the methodological quality of overviews of reviews. Currently, there is no such tool available. While if tools for systematic reviews such as AMSTAR could also be used in overviews of reviews, possibly through adaptation, remains to be determined. Considering the large overlap between systematic reviews and overviews of reviews, trying to adapt available tools seems worthwhile (Table 22.1).

## 22.8 Conflicts/Declaration of Interests

Conflicts of interest have become a much debated and studied topic in clinical medicine and research. There are several definitions of conflicts of interests, most of which have in common that a conflict of interest arises when someone has two conflicting goals in his or her research: the primary goal, which is the pursuit of reasonable truth concerning a scientific question, and a secondary goal such as career advancement or profits [37, 38].

**Table 22.1** Similarities and differences between systematic reviews and overviews of reviews

| Feature | Systematic reviews | Overviews of reviews |
|---|---|---|
| Literature searching | As comprehensive as possible | Similar to that for systematic reviews |
| Inclusion criteria | Primary studies | Reviews (including systematic reviews) and/or primary studies |
| Study selection | Independent selection | Similar to that for systematic reviews |
| Data abstraction | Independent abstraction | Similar to that for systematic reviews |
| Data analysis | Qualitative (narrative reviews) and quantitative methods (meta-analysis) | Qualitative (narrative reviews) and quantitative methods (network meta-analysis) |

Conflicts of interest can be financial or nonfinancial in nature. Financial conflicts of interest arise from financial relationships with industry, for example, from research funding, consultancy fees, or lecture fees [37]. Nonfinancial conflicts of interest include intellectual or private conflicts of interest arising from political, academic, or ideological stances or self-interests, for example [39].

A large body of research on conflicts of interest in medicine has been concerned with the comparison of industry-funded to non-industry-funded research, particularly clinical trials. The results of these studies typically showed that industry-funded trials more often had statistically significant results in favor of the funders' interventions, produced larger effect sizes, or came to more positive conclusions than non-industry-funded trials [40–43], although analyses in some areas of research did not find such an association or contradicted others [44, 45]. A number of studies have also examined bias from industry-funded head-to-head trials, again showing that such comparisons tend to produce results in favor of the sponsors' product (sometimes resulting in illogical assertions) [46–48]. Given this large body of research, much of which has found effects of quite a large magnitude with odds ratios in the range of 3–5, there is little doubt that conflicts of interest arising from profit-oriented research are a major issue in biomedical research.

There has also been some research on conflicts of interest in guideline development. However, while there is little doubt that most guideline panels include members with industry affiliations, few studies have examined bias arising from conflicts of interest in guidelines [49]. A recent case study suggested that financial conflicts of interest may have resulted in inappropriate recommendations in a guideline on major depressive disorder [50], while a study on diabetes guidelines did not find an association between financial conflicts of interest and drug recommendations [51].

Research on nonfinancial conflicts of interest is more limited. A recent examination of intellectual conflicts of interest in screening mammography guidelines showed that guideline panels that included radiologists were more likely to recommend routine screening than guidelines developed without radiologists. Furthermore, guidelines with lead authors who had recent publications on breast cancer were more likely to recommend routine screening. Guidelines with a higher proportion of

primary care physicians, on the other hand, were less likely to recommend routine screening [52]. Research on peer review highlights the importance of considering conflicts of interest arising from "in-group favoritism." Two studies showed that author-suggested peer reviewers are more likely to recommend publication [53, 54]. Another study suggested that exclusion of specific peer reviewers upon request and editorial board membership increased the chances of publication [55]. Lastly, a study on data interpretation compared how a group of methodologists interpret a meta-analysis compared to the authors of studies included in the meta-analysis [56]. The results showed that authors who had published studies with statistically significant findings and authors who had published more research on the topic made stronger conclusions based on the meta-analysis they were presented than methodologists. There are limitations to the research on nonfinancial conflicts of interest. For example, in this study context experts not involved in research in the area may have been a more appropriate comparator, since one would expect methodologists to be more skeptical in general and because they have a less practice-oriented perspective. Furthermore, the few available studies on nonfinancial conflicts of interest have been conducted in selected areas, and data on bias resulting from nonfinancial conflicts of interest is sparse. Nevertheless, there can be little doubt that they play just an important role.

Needless to say, the abovementioned aspects of conflicts of interest ought to be considered in overviews just as in any other application of research. That aside, there are some issues which may be more specific to overviews. While there is hardly any empirical data on conflicts of interest in newer types of evidence syntheses so far, there is some reason to believe that such research could provide interesting insights. This is exemplified by a number of studies on conflicts of interest in industry-funded versus independent systematic reviews [57–61]. These studies have generally shown that industry-funded systematic reviews more often make conclusions that are favorable for the sponsor compared to independent reviews. A similar association may be expected for overviews of reviews. In fact, overviews may be even more susceptible to bias arising from secondary interests. The reasons for this are twofold. Firstly, overviews of reviews are highly aggregated fourth-order interpretations of data, which leaves much room for subjectivity. Secondly, a significant number of overviews appear to be conducted by researchers who were also involved in some of the underlying systematic reviews. This is undoubtedly a competing interest, since the authors of the reviews may have already developed strong views about specific interventions and are unlikely to be as critical of their own reviews as they are expected to be of other reviews. Potential for bias from this type of dual (co-)authorship may arise from quality assessment, interpretation of data and conclusions, and (non-)inclusion of or dealing with competing reviews conducted by other authors. Whether this holds true in reality remains open to investigation. In our own analysis, 89 % of Cochrane overviews were affected by dual authorship regarding at least one of the included systematic reviews, and a median of five reviews per overview were affected by dual authorship (unpublished data). This was not considered in most of the overviews, for example, regarding quality assessment of the included reviews, raising questions about independence. However, the

analysis was based on a small sample ($n = 18$) and can only be considered preliminary. Research on larger samples and non-Cochrane overviews could provide further interesting insights into potential conflicts of interests in overviews of reviews. A further research question of interest is how often authorship extends to the level of primary research and how such triple authorship may influence the conduct of overviews. Lastly, many of the research questions that have shed light on to conflicts of interest in primary and secondary research could also be applied to overviews.

## 22.9 Presenting Results

The presentation of results is a crucial point for the further development and acceptability of overviews of reviews. It is known from systematic reviews that forest plots as a graphical representation of meta-analyses are easily understood by the readers. However, it is important to note that meta-analysis might turn out to be difficult in the context of overviews of reviews as primary studies might be included in more than one review. Combining the results of meta-analysis in a meta-meta-analysis would give too much power to these studies [5]. On the other hand, pooling might be acceptable if there is only a small degree of overlap among reviews. A measure of overlap has been developed but has not yet been validated [62]. Another approach might be the exclusion of one or more reviews in order to reduce the amount of overlap.

Furthermore, presenting results of reviews in an overview might be dependent on the type of evidence synthesis in the included reviews. It can be assumed that it is easiest to present results from meta-analysis as the main type of quantitative evidence synthesis due to the fact that it can be presented as a single number, ideally with a confidence interval and/or a p-value and possibly other relevant information such as the number of studies, measures of heterogeneity, type of the model (fixed versus random), and presence of publication bias. However, one should be aware that presenting the results in such a way might be misleading as any information on the quality of evidence is missing. The presentation of results is also far from clear in the case of qualitative (narrative) synthesis where no single number can be presented, making the presentation of the results of a review even more challenging. One possibility to present results in this case is called vote counting. Vote counting simply counts the number of studies indicating an (statistically significant) effect compared to the number of studies that do not. However, this approach is known to be at high risk of misleading as it completely neglects the size of the studies. To overcome this barrier modified vote counting has been described to be an improved approach [63]. Modified vote counting contains data on the number of positive effects (irrespective of statistical significance), the median effect size (range) across all studies, the number of statistically significant effects, and the total number of studies. Although this approach might be still regarded to be error prone, it seems to be feasible in the context of overviews despite its disadvantages. One should also note that this approach relies heavily on reporting in the reviews as a number of items have to be extracted and presented from the primary studies. The use of modified vote counting should be discussed much more in the future.

Presentation gets even more complex when reviews with both quantitative and qualitative synthesis are included. Presenting results in a comparable way for both types, thus giving them equal weight, can prove challenging. This might be of importance since meta-analysis is often mistakenly regarded as a superior form of evidence synthesis. Modified vote counting might provide a solution here. It will be important to find a format to present results while maintaining a reasonable balance between complexity and a potential loss of important findings from the reviews.

**Conclusions**

Each section of this chapter highlighted one or more specific issues where research should focus on in the future to make overviews of reviews more useful. As we have shown, many questions are still not answered, while others have not been questioned at all, so far. Foremost, clear definitions for overviews of reviews are needed to create common ground for further discussions and developments. This should be regarded in the context of different types of overviews of reviews. If research will focus much more on the different subtypes of overviews of reviews, it will be useful to define them clearly as well. The chapter on overviews of reviews in the *Cochrane Handbook* is a good starting point, and it should be read by anyone who is going to conduct an overview of reviews until more guidance is available [11]. However, its main drawback is the focus on interventions while disregarding other objectives and research questions (e.g., on risk factors).

It is also interesting to know who the recipients of overviews of reviews are, as this has implications for methods and in particular the presentation of results. Different target groups (clinicians, patients) have different needs, and these should be addressed in order for overviews to be useful and gain acceptance. Results of overviews of reviews have to be presented in a way that maintains a reasonable balance between complexity and a potential loss of important findings from the reviews. However, clinicians, patients, and decision makers might need information of different levels of complexity. Therefore, it can be expected that different formats of presenting findings will be a future avenue for overviews of reviews.

While conducting new overviews of reviews, it should be ensured that they are registered in a freely available database. Protocols should be available for all overviews of reviews and accessible via the database. The findings of the overview of reviews should also be incorporated into the database, once the overview of reviews is completed. Such a database will serve as an information source for clinicians, patients, and decision makers. In addition, it will facilitate further methodological projects on overviews of reviews. A special interest should also be given to conflicts of interests and their declaration, as they are a potential source of bias. Authors of overviews of reviews should declare their involvement in reviews that will potentially be included in an overview of reviews.

Overall, it can be concluded that overviews of reviews are going to be accepted as a new type of evidence synthesis. However, many questions related to the methodology and their rigor still remain open, and much more work is needed on this. Many of these issues will have a huge impact on their future. International

collaborations and networks consisting of people with experience in overviews of reviews should try to find answers for many of these questions. One of the most important tasks in the next few years should be the development of reporting standards.

## References


1. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7:e49667.
2. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. J Clin Epidemiol. 2012;65:1267–73.
3. van der Feltz-Cornelis CM, Sarchiapone M, Postuvan V, Volker D, Roskar S, Grum AT, et al. Best practice elements of multilevel suicide prevention strategies: a review of systematic reviews. Crisis. 2011;32:319–33.
4. Pieper D, Antoine S-L, Morfeld J-C, Mathes T, Eikermann M. Methodological approaches in conducting overviews: current state in HTA agencies. Res Synth Methods. 2014;5:187–99.
5. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011;11:15.
6. Fronsdal KB, Facey K, Klemp M, Norderhaug IN, Morland B, Rottingen JA. Health technology assessment to optimize health technology utilization: using implementation initiatives and monitoring processes. Int J Technol Assess Health Care. 2010;26:309–16.
7. IQWiG. Allgemeine Methoden 4.1. 2013. [06 Jan 2014]. Available from: https://www.iqwig.de/download/IQWiG_Methoden_Version_4-1.pdf.
8. Wood AJJ. Progress and deficiencies in the registration of clinical trials. N Engl J Med. 2009;360:824–30.
9. Silva V, Grande AJ, Carvalho AP, Martimbianco AL, Riera R. Overview of systematic reviews – a new type of study. Part II. Sao Paulo Med J. 2015;133(3):206–17.
10. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev. 2012;1:2.
11. Becker LA, Oxman AD. Overviews of reviews. In: Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008. p. 607–31.
12. Pieper D, Mathes T, Neugebauer E, Eikermann M. State of evidence on the relationship between high-volume hospitals and outcomes in surgery: a systematic review of systematic reviews. J Am Coll Surg. 2013;216:1015–25.
13. Bowater RJ, Abdelmalik SM, Lilford RJ. Efficacy of adjuvant chemotherapy after surgery when considered over all cancer types: a synthesis of meta-analyses. Ann Surg Oncol. 2012;19:3343–50.
14. Chen YF, Hemming K, Chilton PJ, Gupta KK, Altman DG, Lilford RJ. Scientific hypotheses can be tested by comparing the effects of one treatment over many diseases in a systematic review. J Clin Epidemiol. 2014;67(12):1309–19.
15. Singh JA, Wells GA, Christensen R, Tanjong Ghogomu E, Maxwell L, Macdonald JK, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. Cochrane Database Syst Rev. 2011;(2):CD008794.
16. Ozbilen M, Adams CE. Systematic overview of Cochrane reviews for anticholinergic effects of antipsychotic drugs. J Clin Psychopharmacol. 2009;29(2):141–6.
17. Posadzki P, Watson LK, Ernst E. Adverse effects of herbal medicines: an overview of systematic reviews. Clin Med. 2013;13(1):7–12.
18. Furlan AD, Yazdi F, Tsertsvadze A, Gross A, Van Tulder M, Santaguida L, et al. Complementary and alternative therapies for back pain II. Evid Rep Technol Assess (Full Rep). 2010;194:1–764.
19. van Middelkoop M, Rubinstein SM, Kuijpers T, Verhagen AP, Ostelo R, Koes BW, et al. A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. Eur Spine J. 2011;20(1):19–39.

20. Glazener CM, Evans JH, Peto RE. Alarm interventions for nocturnal enuresis in children. Cochrane Database Syst Rev. 2005;(2):CD002911.
21. Deshpande AV, Caldwell PH, Sureshkumar P. Drugs for nocturnal enuresis in children (other than desmopressin and tricyclics). Cochrane Database Syst Rev. 2012;(12):CD002238.
22. Glazener CM, Evans JH. Desmopressin for nocturnal enuresis in children. Cochrane Database Syst Rev. 2002;(3):CD002112.
23. Glazener CM, Evans JH, Peto RE. Tricyclic and related drugs for nocturnal enuresis in children. Cochrane Database Syst Rev. 2003;(3):CD002117.
24. Huang T, Shu X, Huang YS, Cheuk DK. Complementary and miscellaneous interventions for nocturnal enuresis in children. Cochrane Database Syst Rev. 2011;(12):CD005230.
25. Caldwell PH, Nankivell G, Sureshkumar P. Simple behavioural interventions for nocturnal enuresis in children. Cochrane Database Syst Rev. 2013;(7):CD003637.
26. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. CMAJ. 1997;156(10):1411–6.
27. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and ElaborationPRISMA: explanation and elaboration. Ann Intern Med. 2009;151(4):W-65.
28. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA. 2000;283(15):2008–12.
29. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol. 2012;12:181.
30. Altman DG, Moher D. Declaration of transparency for each research article. BMJ. 2013;347:f4796.
31. Li L, Tian J, Tian H, Sun R, Liu Y, Yang K. Quality and transparency of overviews of systematic reviews. J Evid Based Med. 2012;5(3):166–73.
32. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264–9. W64.
33. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med. 2010;7(2):e1000217.
34. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009;62(10):1013–20.
35. Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, et al. From systematic reviews to clinical recommendations for evidence-based health care: validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J. 2010;4:84–91.
36. Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. J Clin Epidemiol. 2015;68(5):574–83.
37. Lo B, Field MJ. Conflict of interest in medical research, education, and practice. Washington, DC: National Academies Press; 2009. p. 436.
38. Vineis P, Saracci R. Conflicts of interest matter and awareness is needed. J Epidemiol Community Health. 2015;69(10):1018–20.
39. Barbour V, Clark J, Peiperl L, Veitch E, Wong M, Yamey G. Making sense of non-financial competing interests. PLoS Med. 2008;5:E1999.
40. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA. 2003;290:921–8.
41. Bhandari M, Busse JW, Jackowski D, Montori VM, Schünemann H, Sprague S, et al. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. CMAJ. 2004;170:477–80.
42. Killin LOJ, Russ TC, Starr JM, Abrahams S, Della Sala S. The effect of funding sources on donezepil randomised controlled trial outcome: a meta-analysis. BMJ Open. 2014;4:e004083.
43. Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. BMJ (Clin Res Ed). 2002;325:249.

44. Khan NA, Lombeida JI, Singh M, Spencer HJ, Torralba KD. Association of industry funding with the outcome and quality of randomized controlled trials of drug therapy for rheumatoid arthritis. Arthritis Rheum. 2012;64:2059–67.
45. Mugambi MN, Musekiwa A, Lombard M, Young T, Blaauw R. Association between funding source, methodological quality and research outcomes in randomized controlled trials of synbiotics, probiotics and prebiotics added to infant formula: a systematic review. BMC Med Res Methodol. 2013;13:137.
46. Bero L, Oostvogel F, Bacchetti P, Lee K. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. PLoS Med. 2007;4:1001–10.
47. Flacco ME, Manzoli L, Boccia S, Capasso L, Aleksovska K, Rosso A, et al. Head-to-head randomized trials are mostly industry-sponsored and almost always favour the industry sponsor. J Clin Epidemiol. 2015;68(7):811–20.
48. Gartlehner G, Morgan L, Thieda P, Fleg A. The effect of study sponsorship on a systematically evaluated body of evidence of head-to-head trials was modest: secondary analysis of a systematic review. J Clin Epidemiol. 2010;63:117–25.
49. Norris SL, Holmer HK, Ogden LA, Burda BU. Conflict of interest in clinical practice guideline development: a systematic review. PLoS One. 2011;6:e25153.
50. Cosgrove L, Bursztajn HJ, Erlich DR, Wheeler EE, Shaughnessy AF. Conflicts of interest and the quality of recommendations in clinical guidelines. J Eval Clin Pract. 2013;19:674–81.
51. Norris SL, Holmer HK, Ogden LA, Burda BU, Fu R. Conflicts of interest among authors of clinical practice guidelines for glycemic control in type 2 diabetes mellitus. PLoS One. 2013;8:e75284.
52. Norris SL, Burda BU, Holmer HK, Ogden LA, Fu R, Bero L, et al. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol. 2012;65:725–33.
53. Schroter S, Tite L, Hutchings A, Black N. Differences in review quality and recommendations for publication between peer reviewers suggested by authors or by editors. JAMA. 2006;295:314–7.
54. Wager E, Parkin EC, Tamber PS. Are reviewers suggested by authors as good as those chosen by editors? Results of a rater-blinded, retrospective study. BMC Med. 2006;4:61–4.
55. Goldsmith LA, Blalock EN, Bobkova H, Hall RP. Picking your peers. J Invest Dermatol. 2006;126:1429–30.
56. Panagiotou OA, Ioannidis JPA. Primary study authors of significant studies are more likely to believe that a strong association exists in a heterogeneous meta-analysis compared with methodologists. J Clin Epidemiol. 2012;65:740–7.
57. Bes-Rastrollo M, Schulze MB, Ruiz-Canela M, Martinez-Gonzalez MA. Financial conflicts of interest and reporting bias regarding the association between sugar-sweetened beverages and weight gain: a systematic review of systematic reviews. PLoS Med. 2013;10:1–9.
58. Dunn AG, Arachi D, Hudgins J, Tsafnat G, Coiera E, Bourgeois FT. Financial conflicts of interest and conclusions about neuraminidase inhibitors for influenza. Ann Intern Med. 2014;161:513.
59. Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. BMJ (Clin Res Ed). 2006;333:782.
60. Jørgensen AW, Maric KL, Tendal B, Faurschou A, Gøtzsche PC. Industry-supported meta-analyses compared with meta-analyses with non-profit or no support: differences in methodological quality and conclusions. BMC Med Res Methodol. 2008;8:60.
61. Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. BMJ (Clin Res Ed). 2007;335:1202–5.
62. Pieper D, Antoine SL, Mathes T, Neugebauer EAM, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol. 2014;67:368–75.
63. Grimshaw J, McAuley LM, Bero LA, Grilli R, Oxman AD, Ramsay C, et al. Systematic reviews of the effectiveness of quality improvement strategies and programmes. Qual Saf Health Care. 2003;12(4):298–303. Epub 2003/08/05.

# Conclusion

# 23

Giuseppe Biondi-Zoccai

**Abstract**

An ongoing challenge looms for all scholars and decision-makers due to the momentous increase in the breadth of evidence focusing on important decisions. This is typical of the current era dominated by information, access, and participation. Modern information tools to access and exploit this wealth of data, such as online databases, reference management software, and online data repositories, offer a unique opportunity to maximize the scientific and practical yield of available evidence. However, these instruments must be mastered competently, and the resulting pieces of evidence have to be appraised with full awareness of their strengths and weaknesses. Indeed, decision-makers must be able to appropriately use a variety of study designs in alternating settings or for different goals, without limiting their scope to a single one. We cannot narrow our attention to randomized trials or meta-analyses, in as much as we cannot rely solely on case reports. Moving from individuals to studies combining several groups of people and to reviews combining several trials, umbrella reviews and overviews of reviews represent an additional and even more generalizing stage of evidence synthesis. Despite their novelty, their role in the hierarchy of evidence is becoming more and more established. Even in the most challenging settings (e.g., when the evidence base is inconclusive, inconsistent, or incoherent), meta-epidemiologic studies may provide uniquely original and comprehensive perspectives on important issues and may thus indirectly guide both research and practice. Despite the milestones already met in this field of evidence synthesis, further developments are

G. Biondi-Zoccai, MD, MStat
Department of Medico-Surgical Sciences and Biotechnologies,
Sapienza University of Rome, Corso della Repubblica 79, 04100 Latina, Italy

Eleonora Lorillard Spencer Cenci Foundation, Rome, Italy
e-mail: giuseppe.biondizoccai@uniroma1.it



G. Biondi-Zoccai (ed.), *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies*, DOI 10.1007/978-3-319-25655-9_23

eagerly awaited and will surely enable more careful design, conduct, and reporting of umbrella reviews, overviews of reviews, and meta-epidemiologic studies, hopefully leading to more accurate and prolific results.

*I open at the close.*

J. K. Rowling [1]

The information age and the increased participation of multiple players in the creation and expansion of evidence on important clinical, educational, psychological, or social issues represent at the same time a major breakthrough of current times and a formidable challenge [2]. This holds true for many applied fields of human endeavor, from clinical medicine, to biomedical research, education, economics or physics, as well as any field of human knowledge at large. Several attitudes toward this information and access overload are possible, from negligent to passive, but the most productive and constructive way to face the current evidence era is to focus, develop, and apply suitably effective tools for evidence synthesis. The hierarchy of evidence, in clinical medicine as well as in education, economics, or sociology, just to name a few areas of human endeavor, typically goes from individuals to studies including several similarly selected persons and to systematic reviews and pairwise meta-analyses including several homogeneous trials with the above characteristics, up to network meta-analyses with similarly robust features [3, 4]. Should we refrain from further generalizing efforts or go backward with a reductionist stance?

In this book, we have presented, thanks to the contribution of a large international panel of prominent experts, the case for umbrella reviews, overviews of reviews, and meta-epidemiologic studies. We have highlighted that umbrella reviews, including, according to our explicit logic, the more general concept of overviews of reviews and the alternative concept of meta-epidemiologic studies, represent a recent yet uniquely powerful and versatile tool for evidence synthesis. While they have been introduced formally only in the last two decades and their rightful place in the evidence hierarchy is still debated, we believe that in the future they will be recognized more and more often as one of the very uppermost levels in the hierarchy of evidence. Why so?

First, as it clearly transpires from the previous chapters, umbrella reviews, overviews of reviews, and meta-epidemiologic studies are very comprehensive and flexible and can be applied to most fields and topics. Second, they may provide poignantly concise summaries of very complex issues, while simultaneously testing with meta-epidemiologic methods complex effect modifiers which could have been missed or overlooked at more focused levels. Indeed, we may be tempted to see a very strong parallel between the way the human brain works, based on ions moving through cell membranes in synapses, belonging to many different neurons, organized in a complex fashion in the human brain, with the population-level transition from individuals to trials, pairwise meta-analyses, mixed treatment comparisons, and, eventually, umbrella reviews [5]. Accordingly, we are positive that umbrella reviews can offer a remarkably efficient tool to govern such informational complexity in as much as the human brain can control and exploit its complex physiology

encompassing an average of 100 billion neurons connected by means of a total of more than 100 trillion synapses.

Despite these pros, the previous contributions have also made clear that umbrella reviews are no panacea and their validity can be undermined by specific methodological issues (e.g., information bias, ecological fallacy, and regression to the mean), on top of all the limitations of primary studies and reviews being overviewed, which are all too often and surreptitiously carried over to the higher evidence synthesis level [6–8]. Nonetheless, the positive downfalls due to careful reading and application of a soundly designed, conducted, and reported umbrella review clearly outweigh the negative consequences and implications.

The aim of this book was indeed to formally introduce umbrella reviews, highlighting best and recommended practices in designing, conducting, reporting, reading, and applying such research tools. Despite our collective effort and in light of the fact that this field of evidence synthesis is quite novel and this is the first book ever devoted to this topic, we again concede that errors and typos are likely. We hope you will candidly highlight them so that the future editions of this opus will prove better and more impactful. In addition, an important caveat is that this field of research, as similar fields of evidence synthesis, is surely going to evolve rapidly. As previously stated, it is thus not unlikely that this type of book will become outdated in 5 years or less [4]. This is not necessarily sad news for anyone believing in the pros of umbrella reviews, such as ourselves. Indeed, this would be the very piece of evidence that umbrella reviews have further improved and can be used more maturely to guide decision-making.

Finally, we hope that this book has proved useful as well as entertaining within a scholarly framework, as we maintain that entertaining works can also be insightful and profound, as clearly proven by the book we quote at the beginning of this chapter [1].

## References

1. Rowling JK. Harry Potter and the Deathly Hallows. New York: Arthur A. Levine; 2009.
2. Rifkin J. The age of access: the new culture of hypercapitalism, where all of life is a paid-for experience. Los Angeles: Tarcher; 2001.
3. Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice. 3rd ed. New York: McGraw-Hill Professional; 2014.
4. Biondi-Zoccai G, editor. Network meta-analysis: evidence synthesis with mixed treatment comparison. Hauppauge: Nova; 2014.
5. Ramón y Cajal S. Recuerdos de mi Vida. Madrid: Madrid Imprenta y Librería de N. Moya; 1917.
6. Piantadosi S, Byar DP, Green SB. The ecological fallacy. Am J Epidemiol. 1988;127:893–904.
7. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol. 2005;34:215–20.
8. Egger M, Smith GD, Sterne JA. Uses and abuses of meta-analysis. Clin Med. 2001;1:478–84.