Arthroskopie

AGA-Komitee-Hefte

Arthroskopie 2024 · 37:224–231
https://doi.org/10.1007/s00142-024-00681-y
Accepted: 15 April 2024
Published online: 23 April 2024

Study designs, levels of evidence, and scientific bias

Adrian Deichsel[1] · Lukas N. Münch[2] · Brenda Laky[3] · AGA Research Committee

[1] Department of Trauma, Hand and Reconstructive Surgery, University Hospital Münster, Münster, Germany

[2] Department of Sports Orthopaedics, Technical University of Munich, Munich, Germany

[3] Austrian Research Group for Regenerative and Orthopedic Medicine (AURROM), Vienna, Austria

Abstract

Bias should be taken into account when assessing clinical trials. It can occur in various forms in clinical studies and might influence the results in different directions. Bias can occur through the selection of study patients, the investigators, the type of data, and the analysis of the data. Different study types suffer from different potential biases. The aim of this paper is to describe common types of clinical trials and to illustrate their potential biases.

Keywords
Evidence-based medicine · Clinical trial · Randomized controlled trial · Statistics · Research designs

Introduction

The quality and relevance of a research project depends on the choice of the appropriate study design. The study design, in turn, is determined by the respective research question and must be defined before conducting the study. In general, a distinction is made between original studies including basic research, observational studies and interventional studies, and review articles and meta-analyses. In 1981, Dr. David Sackett (McMaster University, Canada) and colleagues laid the groundwork for evidence-based medicine (EBM) [2]. From its inception, the goal of EBM has been to integrate the best available scientific evidence into clinical decision making. In this process, it is necessary to evaluate the available literature on a specific topic to critically assess the validity of the presented data. Therefore, when evaluating clinical trials, classification according to study type into different levels of evidence is an almost omnipresent concept. The classification is made according to the risk that bias is present in the data [3, 4]. Blinded randomized controlled trials (RCTs) have the highest scientific quality, as they are at least subject to relevant bias. In the following, the levels of evidence for different research questions (therapeutic, prognostic, diagnostic, economic) and disciplines have been adapted so that a classification can be made for each clinical study (see ◘ **Fig. 1**).

Original research

Basic research (non-human research)

In this area, a fundamental distinction is made between theoretical and applied basic research. Theoretical basic research primarily comprises method development and material testing, such as the development/improvement of analytical (determination of markers, enzymes, genes) or imaging measurement procedures (computed tomography, magnetic resonance imaging). Applied basic research includes experiments on animals, cells, biochemical and genetic studies, and material development. This experimental research usually involves studying the effects of changing at least one independent variable on the dependent variable. By performing stan-

Scan QR code & read article online

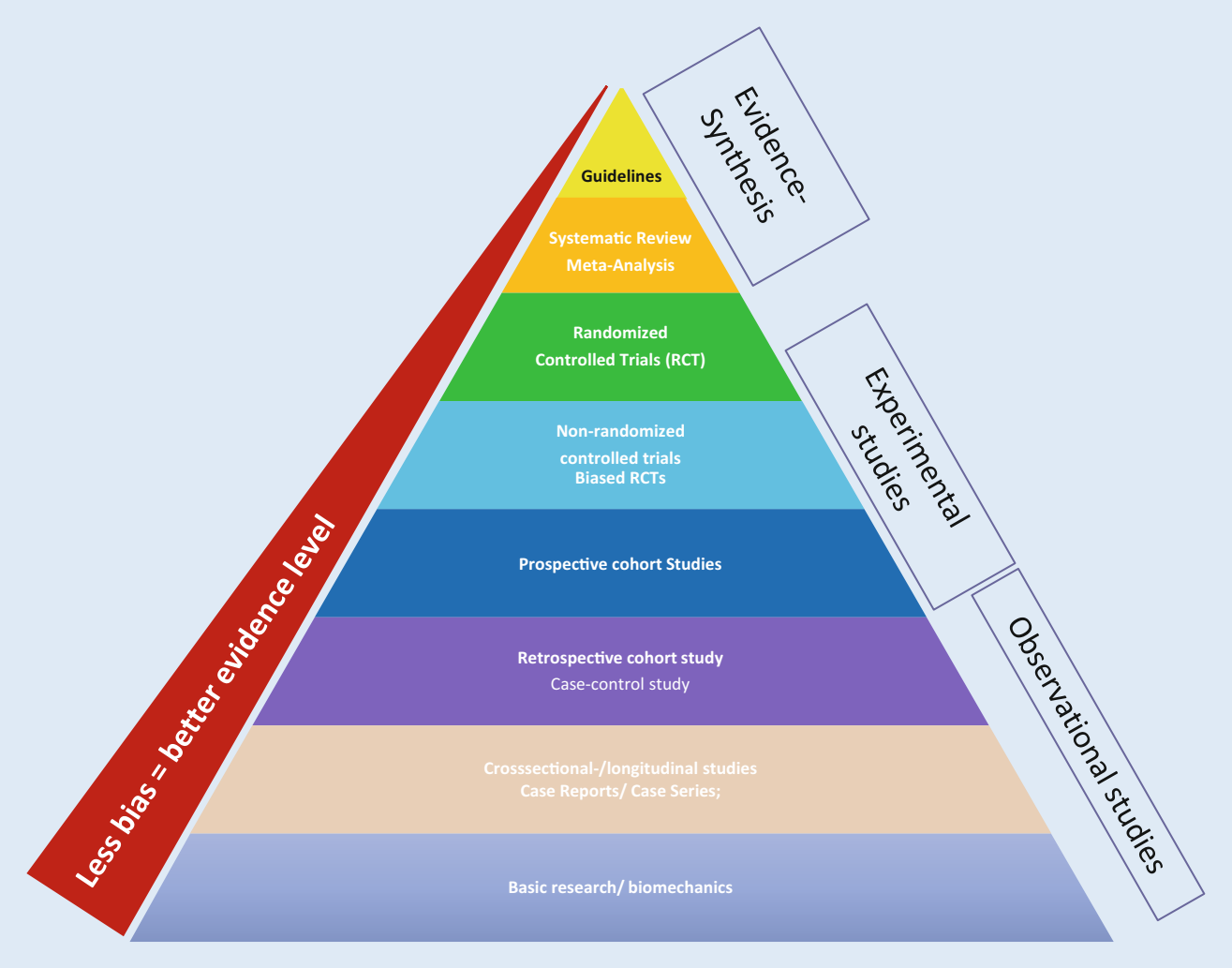

Fig. 1 ▲ Evidence levels of medical studies

dardized experiments, high internal validity can be achieved, with usually limited extrapolation of study results to clinical reality.

Clinical research

In clinical (human) research, a basic distinction is made between observational and interventional studies. Observational studies are characterized by the fact that the scientist cannot control the assignment of the treatment to the subjects, which can lead to certain biases in the study results. These include cohort studies, case–control studies, cross-sectional and longitudinal studies, as well as case series and single-case presentations. Interventional clinical trials, on the other hand, include studies in which the researcher can assign participants directly to an intervention, usually by randomization. These include, in particular, RCTs (**Table 1**).

Observational studies

Observational studies aim to gain insights into possible associations between certain diseases, their treatment, and other factors (e.g., individual life circumstances) of the study participants. Here, study participants are merely observed, so no specific intervention takes place as part of the study. However, observational studies can follow patients after predefined surgeries. Mostly, observational studies involve large groups of participants (healthy, patients/experienced, or even former patients), who are usually evaluated by means of questionnaires. These observational studies can be either retrospective or prospective in nature (**Fig. 2**). Overall, the study types described below can also be summarized as non-randomized studies (NRS).

Cohort studies follow patients based on specific exposures (e.g., a particular surgery), so they distinguish between exposed and unexposed subjects. They can be both prospective (level 2) and retrospective (level 3). In a prospectively designed cohort study, the research question is formulated before the start of the study and patients are followed longitudinally according to the study design. In a retrospective cohort study, the research question is formulated and answered on the basis of existing data (e.g., register, in-hospital database).

Case–control studies examine patients based on a clinical event (e.g., re-rupture of the anterior cruciate ligament). Patients in

Table 1 Observational and interventional clinical studies

Observation (prospective or retrospective)	Intervention (prospective)
Cohort studies Case-control studies Cross-sectional studies Longitudinal studies Case series Case reports	Interventional clinical studies Clinical trial (according to AMG/MPG) ±controlled (e.g., placebo, healthy control, standard therapy …) ±randomized to prevent selection bias ±blinded to prevent information bias ±parallel group ±cross-over

AMG Arzneimittelgesetz (German Medicines Act), *MPG* Medizinproduktegesetz (German Medical Devices Act)

whom the event occurred (cases) are compared to patients without the event (controls) to find possible differences. In the literature, the classifications as case–control study and retrospective cohort study are often incorrectly used as interchangeable. Here, retrospective studies that follow patients based on a therapy (exposure), i.e., cohort studies, are incorrectly referred to as case–control studies ([1]; ◘ **Fig. 3**).

Case series also examine a group of patients on the basis of an event. However, in comparison to the case–control study, there is no control group, so that only descriptive results can be derived from case series. Hypotheses arising from such data must subsequently be tested in further clinical trials of higher levels of evidence. Therefore, case series have the lowest level of evidence compared to the other study designs.

In a *cross-sectional study*, the survey, observation, or measurement is conducted once, so it reflects only a snapshot in time. Cross-sectional studies can be used both descriptively and analytically. In contrast, in a *longitudinal study*, the same survey, observation, or measurement is conducted at multiple timepoints, so it reflects processes of change over time (◘ **Fig. 4**).

Selection bias in non-randomized trials. The key feature of NRS is that the included patients originate from current common clinical practice and are not subject to any study-specific intervention. Regardless of whether data were collected prospectively or retrospectively, the physician decides on the patient's treatment [6]. In addition to the available evidence, the physician's decision is influenced by their own experience, the patient's needs, and other factors. For example, young, athletically active patients with high physical demands may be more likely to be recommended a reconstruction of the anterior cruciate ligament than less active, or older patients. If surgically and conservatively treated patients are now compared to each other with regard to their clinical outcome, these confounding variables ("confounders") between the patient groups can lead to bias in the study results. The conscious or subconscious selection of patients for study groups based on parameters that are not the variable of interest (e.g., intervention) is referred to as "selection bias" [7]. This can lead to overestimation of the effects of an intervention in an NRS. Thus, meta-analyses that include NRS and RCTs frequently show a more pronounced effect of an intervention in NRS compared to RCTs [8, 9]. In particular, small significant effects in observational studies should be interpreted with caution, as they may be influenced by a relevant selection bias.

The most effective way to minimize selection bias is randomization, as described below. However, it is also possible to compensate for confounding variables. A well-known possibility is the implementation of a "matched-pair" analysis. Here, the study groups are composed in such a way that they are similar in previously defined basic variables that are known to have an influence on the results (e.g., age, gender, body mass index, smoking habits) [10]. However, matching can only compensate for known confounding variables. Thus, selection bias can at best be reduced, but not fully eliminated. In addition to matching, selection bias can also be reduced by applying specific statistical methods, e.g., through propensity score matching. As explaining propensity score matching would be outside the scope of this article, we refer to exemplary literature [10, 11]. Regardless of whether specific efforts were made to adjust for confounding variables, each scientific article should have a description of the study groups [12]. This should include, in addition to epidemiological parameters (age, sex, body mass index, smoking status, etc.), possible study-specific confounding variables (e.g., meniscal or cartilage injuries in anterior cruciate ligament reconstruction studies). Thus, the reader can see for himself to what extent the study groups are similar.

Retrospective assessment of preoperative parameters and the recall bias. A main drawback of retrospective studies is the fact that patients find it difficult to adequately reconstruct their medical history. When subjective parameters are queried, this can lead to biases known as "recall bias." For example, Crutchfield and colleagues showed that when patients retrospectively answered patient-centered questionnaires after hip arthroscopy, this recall bias caused patients to rate their preoperative subjective condition significantly worse than it actually was. This then leads to overestimation of the effect of an intervention postoperatively [13]. Similar observations were shown in shoulder and knee surgery [14, 15]. Thus, a prospective survey of outcome parameters, as performed in prospective cohort studies, is preferable.

Outcome parameter collection. Outcome parameters in clinical studies can be assessed in various ways. While objective parameters (e.g., quantification of knee joint stability using the Rolimeter® (Aircast Europa, Neubeuern, Germany)) have to be collected by an investigator, there are different ways to collect subjective patient-centered parameters. In particular, the regularly queried patient-reported outcome measures (PROMs) are suitable for collection using web-tools or by telephone, provided they have been validated for this purpose. However, the manner of collection can have significant influences on the outcome. Acosta et al. showed in their meta-analysis that

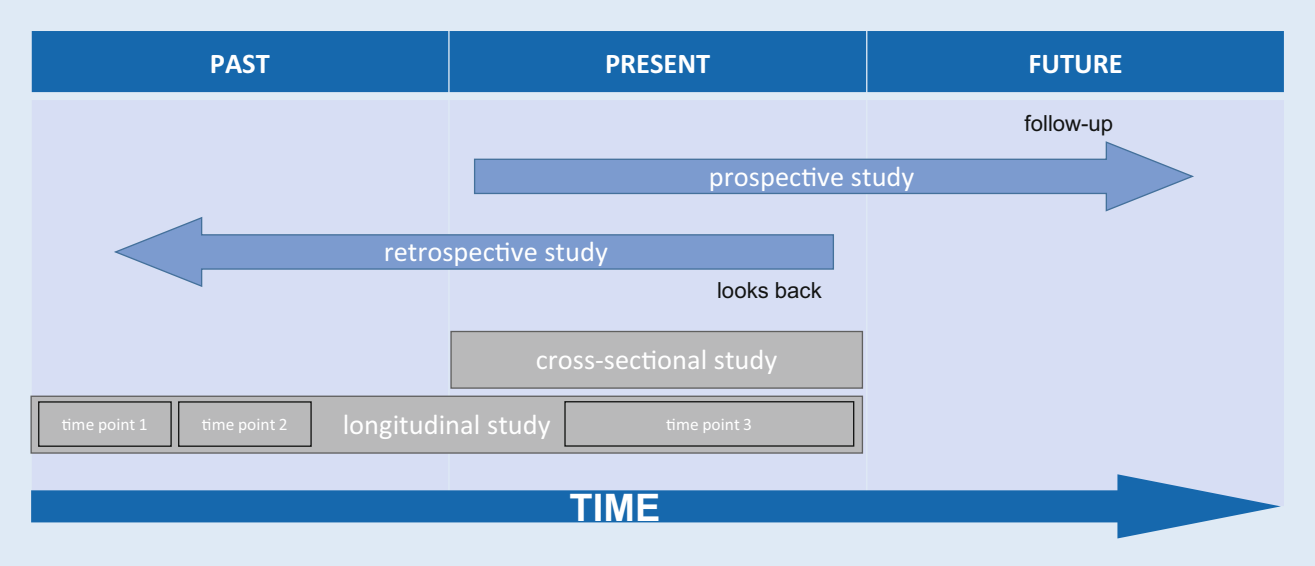

Fig. 2 ▲ Difference between retrospective and prospective observational studies

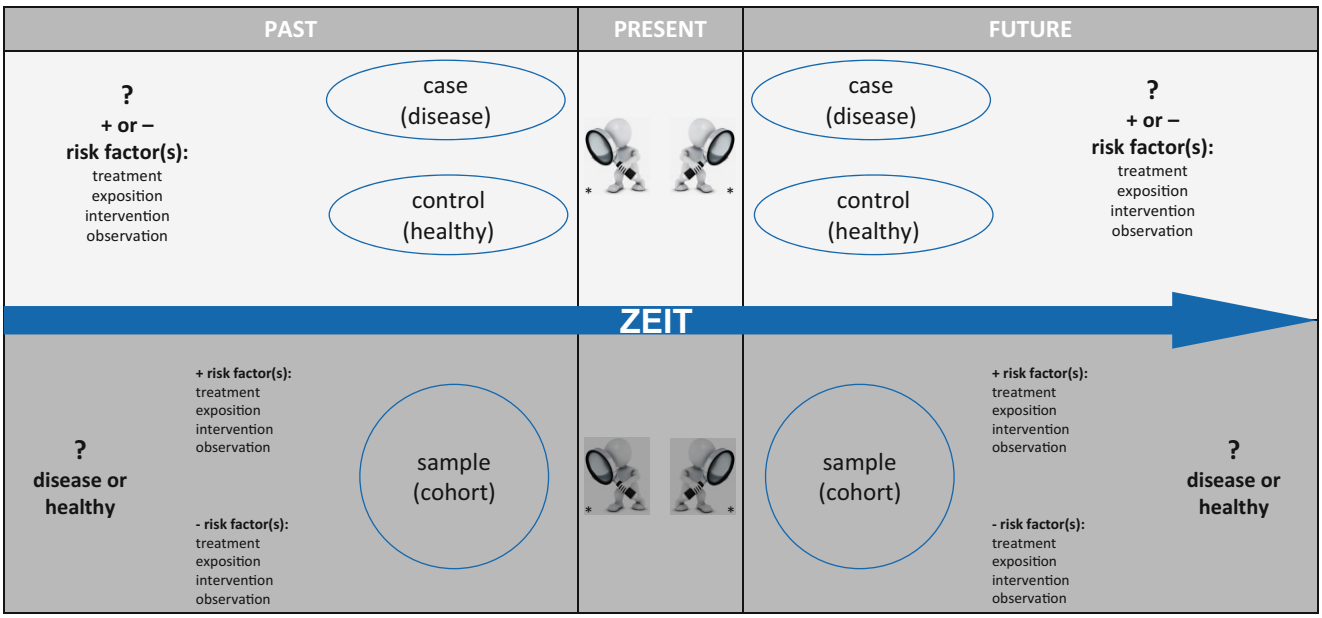

Fig. 3 ▲ Difference between case–control (*top*, *light grey*) and cohort (*bottom*, *dark grey*) studies

PROMs collected by telephone showed significantly higher scores than online-based or face-to-face collection [16]. This is not necessarily a problem as long as the same methods are used consistently throughout the study. However, it regularly occurs that the initial survey was conducted during a face-to-face presentation (at inclusion), but the subsequent final evaluation, especially in the case of long-term follow-up, was conducted by telephone. This could bias postoperative values towards larger effects.

Intervention studies

Prospective RCTs are generally accepted as the gold standard of clinical intervention studies (level 1) [17]. However, even this type of study is subject to certain risks of bias that may call the study results into question, and not every RCT is of equally high quality [5].

Randomization as a solution to selection bias. The main feature of RCTs is randomization. During randomization, patients who meet the inclusion criteria defined before the start of the study are randomly assigned to one of the study groups. This ensures an equal or random distribution of potential confounders between the study groups. Selection bias, as described above, is thus largely excluded or its risk is reduced [17]. Randomization can be performed with various methods. In addition to the classic coin flip, commonly accepted methods include allocation using computer-generated sequences [18, 19]. Non-accepted strategies

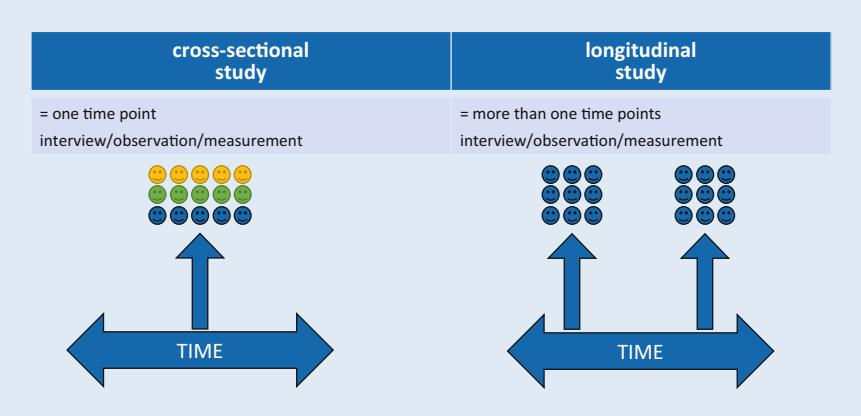

Fig. 4 ▲ Difference between cross-sectional and longitudinal studies

are randomization by sex, date of birth, date of inclusion, or other patient- or clinic-specific factors, as they do not guarantee adequate randomization. It is also important that allocation remains unknown to the including personnel, hence preventing (conscious or unconscious) allocation to a preferred group (blinding of allocation) [20].

One problem that regularly arises in surgical intervention trials is patient (non-)compliance. For example, patients randomized to surgery may decide not to undergo surgery, thereby losing the original randomization. One way to account for this is to perform an intention-to-treat (ITT) analysis, in which patients are still analyzed in the group to which they were originally randomized, regardless of the actual therapy. For a detailed discussion of the specifics of ITT, as well as possible alternatives, we refer the reader to the relevant AGA Research Committee article by Laky et al. [21].

The problem of blinding in orthopedic RCTs. Blinding in orthopedic studies can occur at different levels. Blinding of the surgeon is not possible for obvious reasons. In the case of intervention studies, implementation of studies blinded to the patient is only possible by performing a sham operation, which presents relevant ethical hurdles, so that this can only be justified in very rare cases [22]. Lastly, the person collecting the clinical endpoints (e.g., physical examination at follow-up) can be blinded. If this is not the case, it may lead to observer bias, as knowledge about the patient's treatment may consciously or unconsciously influence the observer's assessments [23]. Both the lack of blinding of the patient and the person following up can lead to significant bias in the reported results. Usually, this leads to overestimation of the effect of a treatment [24, 25]. The challenges that can arise from this problem, as well as possible solutions, have already been reviewed in a previous article from the research committee of Günther et al. [26].

Loss to follow-up and attrition bias. Ideally, all patients included at the beginning of a study are available for follow-up at the end of data collection. However, as patients may move away, die, become unreachable, or no longer be interested in participating in the study, collection of all datasets is unrealistic, especially in longer-term studies. However, subjects lost to follow-up may differ significantly in terms of their characteristics from the remaining patients. Randomization in an RCT, as described above, ensures homogeneity of study participants between trial groups. If the loss-to-follow-up differs between the study groups, this may mean that the homogeneity of the groups obtained by randomization is no longer guaranteed, which may lead to a difference between the groups that can no longer be explained by the intervention. This bias in study results due to the asymmetric omission of patients for the final analysis of a study is referred to as attrition bias [27]. The relevance of attrition bias is highlighted in the work of Akl and colleagues [28]. In their systematic review, up to one third of the included studies lost their significant difference between groups when a significant difference was simulated for patients who were no longer available for follow-up compared to the remaining patients. Generally applicable thresholds for the extent at which the loss of study patients becomes relevant cannot be found in the literature. As a rule of thumb, it is stated that a loss below 5% is indicative of a low risk of attrition bias, whereas a loss above 20% is considered a high risk [27, 29]. However, the ratio of the loss to follow-up between study groups is more important than the absolute number of subjects lost for final analysis [29].

Secondary research (review articles and meta-analyses)

In secondary research, previously published work is summarized as simple (narrative) or systematic reviews with or without meta-analysis. In a systematic review, all previously published primary studies are systematically identified, selected, and critically evaluated for a clearly defined research question. The results are then summarized descriptively or quantitatively using statistical methods in the form of a meta-analysis. Meta-analysis is thus a statistical method for quantitatively combining the results of studies dealing with the same research question into an overall result within the framework of a systematic review. This is intended to increase the informative value compared to the respective individual studies. A simple (narrative) review, on the other hand, is based on a non-systematic selection process, so that a subjective partial selection of the studies can occur.

Conclusion

- **Evidence levels of clinical studies are based on the susceptibility of a study type to bias.**
- **The primary limitation of non-randomized studies (NRS) is that study groups differ in terms of the distribution of known and unknown confounders, which can influence the study results (selection bias).**
- **NRS tend to overestimate the effect of an intervention, which can be due to various reasons.**
- **Randomized controlled trials (RCTs) represent the highest level of evidence, as they are the least susceptible to bias due**

to their methodology. However, randomized controlled trials can also be subject to bias, which can influence the study results.
- The difficulty of blinding in orthopedic RCTs poses a problem.

Corresponding address

Dr. Adrian Deichsel
Department of Trauma, Hand and Reconstructive Surgery, University Hospital Münster
Albert-Schweitzer-Campus 1, 48149 Münster, Germany
adrian.deichsel@ukmuenster.de

Funding. Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest. A. Deichsel, L.N. Münch, B. Laky, and the AGA Research Committee declare that they have no competing interests.

For this article no studies with human participants or animals were performed by any of the authors. All studies mentioned were in accordance with the ethical standards indicated in each case.

References

1. Esene IN, Mbuagbaw L, Dechambenoit G, Reda W, Kalangu KK (2018) Misclassification of case-control studies in neurosurgery and proposed solutions. World Neurosurg 112:233–242. https://doi.org/10.1016/j.wneu.2018.01.171
2. Sackett D (1981) How to read clinical journals: I. why to read them and how to start reading them critically. Can Med Assoc J 124:555–558
3. Sackett DL (1989) Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 95:2s–4s
4. The periodic health examination (1979) Canadian task force on the periodic health examination. Can Med Assoc J 121:1193–1254
5. Smith CS, Mollon B, Vannabouathong C, Fu JM, Sales B, Bhandari M, Whelan DB (2020) An assessment of randomized controlled trial quality in the journal of Bone & joint surgery. JBJS 102:e116. https://doi.org/10.2106/jbjs.18.00653
6. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I et al (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 355:i4919. https://doi.org/10.1136/bmj.i4919
7. Nunan DBC, Aronson JK Selection bias. http://www.catalogofbias.org/biases/selection-bias/
8. Kruse C, Axelrod D, Al-Mohrej OA, Daniel R, Johal H (2023) Effect of local antibiotic prophylaxis on postoperative deep infection in fracture surgery: a systematic review and meta-analysis. J Orthop Trauma 37:e28–e35. https://doi.org/10.1097/bot.0000000000002487
9. Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD (2011) Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2011:Mr12. https://doi.org/10.1002/14651858.MR000012.pub3
10. Musahl V, Karlsson J, Hirschmann MT, Ayeni OR, Marx RG, Koh JL, Nakamura N (2019) Basic methods handbook for clinical orthopaedic research. Springer
11. Morgan CJ (2018) Reducing bias using propensity score matching. J Nucl Cardiol 25:404–406. https://doi.org/10.1007/s12350-017-1012-y
12. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP (2008) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 61:344–349. https://doi.org/10.1016/j.jclinepi.2007.11.008
13. Crutchfield CR, Givens RR, O'Connor M, deMeireles AJ, Lynch TS (2022) Recall bias in the retrospective collection of common patient-reported outcome scores in hip arthroscopy. Am J Sports Med 50:3190–3197. https://doi.org/10.1177/03635465221118375
14. Macchiarola L, Pirone M, Grassi A, Pizza N, Trisolino G, Stilli S, Zaffagnini S (2022) High recall bias in retrospective assessment of the pediatric International Knee Documentation Committee Questionnaire (Pedi-IKDC) in children with knee pathologies. Knee Surg Sports Traumatol Arthrosc 30:3361–3366. https://doi.org/10.1007/s00167-022-06922-7
15. Gotlin MJ, Kingery MT, Baron SL, McCafferty J, Jazrawi LM, Meislin RJ (2020) Recall bias in retrospective assessment of preoperative patient-reported American shoulder and elbow surgeons scores in arthroscopic rotator cuff repair surgery. Am J Sports Med 48:1471–1475. https://doi.org/10.1177/0363546520913491
16. Acosta J, Tang P, Regal S, Akhavan S, Reynolds A, Schorr R, Hammarstedt JE (2020) Investigating the bias in orthopaedic patient-reported outcome measures by mode of administration: a meta-analysis. J Am Acad Orthop Surg Glob Res Rev 4:e20.00194. https://doi.org/10.5435/JAAOSGlobal-D-20-00194
17. Soucacos PN, Johnson EO, Babis G (2008) Randomised controlled trials in orthopaedic surgery and traumatology: overview of parameters and pitfalls. Injury 39:636–642. https://doi.org/10.1016/j.injury.2008.02.011
18. Suresh K (2011) An overview of randomization techniques: an unbiased assessment of outcome in clinical research. J Hum Reprod Sci 4:8–11. https://doi.org/10.4103/0974-1208.82352
19. Kang M, Ragan BG, Park JH (2008) Issues in outcomes research: an overview of randomization techniques for clinical trials. J Athl Train 43:215–221. https://doi.org/10.4085/1062-6050-43.2.215
20. Higgins JPTTJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (2022) Cochrane handbook for systematic reviews of interventions version 6.3. Cochrane, p 2022
21. Laky B, Heuberer PR, Herbst E, Lattermann C, Günther D, Schüttler KF, Wafaisade A, Mathis D, Rössler PP, Kopf S et al (2021) Intention-to-treat-Prinzip. Arthroskopie 34:233–237. https://doi.org/10.1007/s00142-021-00459-6
22. Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP (2002) A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 347:81–88. https://doi.org/10.1056/NEJMoa013259

Zusammenfassung

Studiendesign, Evidenzlevel und wissenschaftlicher Bias

Bei der Begutachtung klinischer Studien sollten Verzerrungen (sog. Bias) berücksichtigt werden. Ein solcher Bias kann in vielfältigen Formen in Studien vorkommen und diese in verschiedene Richtungen beeinflussen. Verzerrungen können durch die Auswahl der Studienpatienten, die Untersucher, die Art sowie die Analyse der Daten auftreten. Hierbei leiden verschiedene Studientypen unter unterschiedlichen potenziellen Verzerrungen. Das Ziel der vorliegenden Arbeit ist es, gängige klinische Studientypen zu beschreiben sowie mögliche Verzerrungen darzustellen.

Schlüsselwörter
Evidenzbasierte Medizin · Klinische Studie · Randomisierte kontrollierte Studie · Statistik · Forschungsvorhaben

23. Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, Brorson S (2013) Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. Cmaj 185:E201–211. https://doi.org/10.1503/cmaj.120744
24. Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, da Costa BR, Flores-Mir C (2018) Influence of blinding on treatment effect size estimate in randomized controlled trials of oral health interventions. BMC Med Res Methodol 18:42. https://doi.org/10.1186/s12874-018-0491-0
25. Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, Brorson S (2012) Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 344:e1119. https://doi.org/10.1136/bmj.e1119
26. Günther D, Herbst E, Laky B, Lattermann C, Mathis DT, Rössler P, Schüttler KF, Wafaisade A, Kopf S (2021) Research-Komitee der Arbeitsgemeinschaft für, A. Herausforderungen von Evidenzlevel-1-Studien in der Unfallchirurgie und Orthopädie. Arthroskopie 34:299–304. https://doi.org/10.1007/s00142-021-00469-4
27. Catalogue of Bias Collaboration, Bankhead C, Aronson JK, Nunan D Attrition bias. https://catalogofbias.org/biases/attrition-bias/
28. Akl EA, Briel M, You JJ, Sun X, Johnston BC, Busse JW, Mulla S, Lamontagne F, Bassler D, Vera C et al (2012) Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. BMJ 344:e2809. https://doi.org/10.1136/bmj.e2809
29. Schulz KF, Grimes DA (2002) Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet 359:781–785. https://doi.org/10.1016/s0140-6736(02)07882-0