

Meta-analytic Decisions and Reliability: A Serendipitous Case of Three Independent Telecommuting Meta-analyses

Levi R. G. Nieminen · Jessica M. Nicklin ·
Tara K. McClure · Madhura Chakrabarti

Published online: 22 June 2010
© Springer Science+Business Media, LLC 2010

Abstract

Purpose Despite the potential for researcher decisions to negatively impact the reliability of meta-analysis, very few methodological studies have examined this possibility. The present study compared three independent and concurrent telecommuting meta-analyses in order to determine how researcher decisions affected the process and findings of these studies.

Methodology A case study methodology was used, in which three recent telecommuting meta-analyses were re-examined and compared using the process model developed by Wanous et al. (*J Appl Psychol* 74:259–264, 1989).

Findings Results demonstrated important ways in which researcher decisions converged and diverged at stages of the meta-analytic process. The influence of researcher divergence on meta-analytic findings was neither evident in all cases, nor straightforward. Most notably, the overall effects of telecommuting across a range of employee outcomes were generally consistent across the meta-analyses, despite substantial differences in meta-analytic samples.

Implications Results suggest that the effect of researcher decisions on meta-analytic findings may be largely *indirect*, such as when early decisions guide the specific moderation tests that can be undertaken at later stages. However, directly comparable “main effect” findings appeared to be more robust to divergence in researcher decisions. These results provide tentative positive evidence

regarding the reliability of meta-analytic methods and suggest targeted areas for future methodological studies.

Originality This study presents unique insight into a methodological issue that has not received adequate research attention, yet has potential implications for the reliability and validity of meta-analysis as a method.

Keywords Meta-analysis · Methodological · Replication · Reliability · Validity · Telecommuting

Introduction

Meta-analysis is a quantitative method for integrating the findings of existing primary studies. Widely viewed as a more objective and rigorous alternative to the process of narrative literature review, meta-analysis has become increasingly popular across numerous research disciplines (Schulze 2007). As with any research method, the growing popularity of meta-analysis has been paralleled by a great deal of concern focused on various methodological and statistical issues (Schulze 2004).¹ One such issue, which is the focus of the current study, involves the potential for various researcher decisions to bias meta-analytic outcomes (Shadish et al. 2002). However, with the exception of data simulation studies that compare different statistical approaches to meta-analysis, very little research has focused on this issue (Aguinis et al. 2009; Bobko and Roth 2008).

As with narrative literature reviews, researchers have the opportunity to shape meta-analytic findings via

L. R. G. Nieminen (✉) · T. K. McClure · M. Chakrabarti
Department of Psychology, Wayne State University,
5057 Woodward Ave, Detroit, MI 48111, USA
e-mail: levi.nieminen@gmail.com

J. M. Nicklin
Department of Psychology, University of Hartford,
West Hartford, CT, USA

¹ A comprehensive summary of these methodological issues is beyond the scope of the current paper; interested readers can refer to recent overviews by Burke and Landis (2003) and Schulze (2007).

unintentional biases and errors or via intentional distortion (Allen and Preiss 1993; Egger and Smith 1998; Hale and Dillard 1991). Researcher decisions at nearly every stage of the meta-analytic process provide potential influence points, ranging from the initial selection of a research topic to the interpretation of moderation analyses (Aguinis et al. 2009; Geyskens et al. 2009; Wanous et al. 1989). However, unlike narrative literature reviews, the subjective component underlying meta-analysis is often less recognizable or not as well understood (Briggs 2005). In part, this may be due to: (1) page limitations in journals, which greatly restricts the amount of detail authors can provide, (2) disincentives for conducting meta-analytic replication studies, or (3) a general lack of methodological work in this area. Regarding the latter point, Bobko and Roth (2008) suggested that the vast majority of methodological research has failed to address the most basic and pertinent issues involved in meta-analysis (e.g., primary study selection, operationalization of theory), instead focusing on statistical refinements. Similarly, recent studies by Aguinis et al. (2009) and Geyskens et al. (2009) focused primarily on the impact of analytical decisions within meta-analysis (e.g., statistical corrections) and did not, for example, address how literature search methods affected study outcomes. Nonetheless, understanding the “magnitude of experimental effects” as a cumulative function of all meta-analytic decisions is critical for gauging the reliability and validity of meta-analysis as a method (Beaman 1991, p. 252).

The current study sought to address recent calls for methodological studies investigating the full range of researcher decisions in meta-analysis, including those that occur during early stages (e.g., Bobko and Roth 2008; Schulze 2007). Using a case study methodology, we undertook a systematic comparison of the process and results of three meta-analyses on the employee-level outcomes associated with telecommuting. The meta-analyses were conducted concurrently by independent research teams, providing a unique opportunity to study researcher decisions. The primary aims of our case study included identifying key decision points in the meta-analytic process and providing insight into the degree to which researcher choices ultimately affected meta-analytic conclusions. Past research has been only partially informative, and several studies appear conflicting regarding the magnitude of researcher effects on meta-analytic outcomes (e.g., Aguinis et al. 2009 versus Geyskens et al. 2009).

Meta-analytic Decisions, Reliability, and Validity

Researcher decisions are evident at nearly every stage of the meta-analytic process. Wanous et al. (1989) identified eight out of 11 total stages that involved some form of

researcher decision. These eight decision stages included: defining the research question of interest, determining inclusion criteria, setting literature search parameters and conducting the search, selecting the final sample of studies to include, extracting relevant statistical information, coding study characteristics, determining rules for aggregating variables, and selecting moderators. The remaining *non-decision* stages involved computing meta-analytic effect sizes, determining whether a search for moderators is warranted, and conducting moderation tests. However, even these computational stages require judgment calls (see Aguinis et al. 2009; Geyskens et al. 2009). Researchers decide what computational approach to use (Schmidt and Hunter 1999), whether a fixed- or random-effects model is more appropriate (Field 2001; Kisamore and Brannick 2008; Overton 1998), which statistical corrections should be applied, how to detect and deal with outliers (Beal et al. 2002), and the criterion that should be used to determine whether or not a search for moderators is justified (Cortina 2003).

A recent study by Geyskens et al. (2009) demonstrated the potential effects of these decisions on meta-analytic estimates. These authors conducted four meta-analyses using data on management topics. Using an iterative process, they assessed the change in meta-analytic estimates as a consequence of decisions related to five statistical corrections, outlier deletion, methods of detecting the presence of moderators, and analysis strategies for moderation tests. Overall, the conclusions drawn by Geyskens et al. are consistent with the findings of previous data simulation studies, which have shown that some researcher decisions *can* have substantial effects on meta-analytic estimates. On the other hand, studies that have examined researcher decisions across completed meta-analyses have failed to demonstrate a consistent linkage between specific researcher decisions and substantive meta-analytic outcomes, raising some doubt as to the extent of this problem *in practice* (Aguinis et al. 2009; Barrick et al. 2001; Bullock and Svyantek 1985; Stewart and Roth 2004; Wanous et al. 1989).

Still, given the ubiquity of researcher decisions in meta-analysis and the fact that some decisions *can* affect meta-analytic outcomes, several researchers have suggested that the subjective component in meta-analysis is more problematic than is recognized (e.g., Briggs 2005; Geyskens et al. 2009). One implication is that different meta-analysts *could* reach different conclusions regarding the same literature. Stated alternatively, researcher decisions present a challenge to the *reliability* and *validity* of meta-analysis as a method, where *reliability* refers to the degree to which independent meta-analytic investigations produce similar findings, and *validity* refers to the extent to which meta-analytic findings represent the true nature of the

phenomenon being studied. Although the validity of meta-analytic conclusions is the ultimate concern, evidence to support validity typically cannot be established directly because the “true” relationships are not known. On the other hand, reliability can be tested by examining the degree of correspondence between independently conducted meta-analyses on the same topic. A high degree of correspondence provides one form of evidence, albeit indirect, for the validity of meta-analytic methods (Allen and Preiss 1993). Conversely, independent meta-analyses that fail to converge in their findings naturally raises specific questions regarding the validity of the meta-analyses involved, and more generally, regarding meta-analytic methods as a whole.

In this way, independently conducted meta-analyses with the same research focus provide a key lens for examining the effect of researcher decisions on meta-analytic outcomes.² Of particular interest is whether independent meta-analysts make different choices, and if so, the degree to which conclusions diverge as a consequence. With these goals in mind, the following section summarizes the findings of key studies to have addressed these issues. The literature reviewed underscores four main shortcomings of prior methodological research in this area, thereby providing the impetus for the current study: (1) there are very few systematic methodological studies of actual meta-analytic decisions (excluding simulation studies), (2) researchers have focused exclusively on meta-analyses that appeared to present contradictory findings, (3) researchers have generally lacked the depth of information needed to understand what caused contradictory findings, and (4) there are mixed contentions regarding the prevalence of contradictory meta-analyses in the literature and the seriousness of researcher subjectivity.

Key Studies of Meta-analytic Decisions

The most widely cited study of meta-analytic decisions is the aforementioned study by Wanous et al. (1989). Their re-examination of four pairs of divergent meta-analyses suggested that the studies were more consistent than originally thought once researcher judgment was taken into account. They concluded that early researcher decisions involving inclusion criteria, search methods, and the selection of the final sample of studies are key influence points in the meta-analytic process, perhaps more so than researchers’ decisions thereafter. For example, in two

instances the discrepancy between meta-analyses could be traced to the inclusion of one or two influential studies. A final case demonstrated that decisions occurring beyond the study accumulation phase could be similarly influential. In this instance, researchers extracted different correlations from eight out of 17 overlapping studies and made different choices regarding whether to interpret meta-analytic estimates separately for facets of job satisfaction versus combine them and interpret the overall effect.

An important caveat of the Wanous et al. (1989) study, as it bears on the preceding discussion regarding the reliability and validity of meta-analytic methods, was that the differences between pairs were “probably not large by past standards of narrative review” (p. 263). For example, although the difference between meta-analyses reported by Scott and Taylor (1985) and Hackett and Guion (1985) regarding the job satisfaction-absenteeism relationship ($r = -.15$ vs. $r = -.07$, respectively) might be described by some as moderate, others may question whether these estimates call for substantively different conclusions. If one concludes that the divergence between these meta-analyses was relatively minor, then the effect of researcher decisions on meta-analytic findings should likewise be qualified as mild. However, as Wanous et al. remind us, interpreting the magnitude of effects is also a judgment call.

A less ambiguous example of contradictory meta-analyses was provided by Bullock and Svyantek (1985). Their meta-analysis failed to replicate an earlier meta-analysis by Terpstra (as cited by Bullock and Svyantek), which had reported a positivity bias among organizational development studies, such that studies of lower methodological rigor tended to yield more positive findings. Bullock and Svyantek noted two key difficulties in attempting to replicate the earlier meta-analysis. First, their search yielded a sample of 90 studies as compared to Terpstra’s 52, despite setting similar search parameters (e.g., time period) and inclusion criteria. Second, they noted that the coding rules used by Terpstra for the categorical independent (methodological rigor) and dependent variables (positivity of findings) were unclear, and thus, difficult to replicate. Unfortunately, lack of necessary information about the prior meta-analysis prevented Bullock and Svyantek from fully understanding which of these factors ultimately contributed to the contradictory findings.

The majority of subsequent attempts to reconcile contradictory meta-analyses have similarly struggled to pinpoint the key sources of divergence. One example involves two meta-analyses on the relationships between personality traits from the Five Factor Model (FFM) and job performance. Barrick and Mount (1991) found that conscientiousness alone was a significant predictor across job and criteria types, whereas Tett et al. (1991) found that only emotional stability was significant and conscientiousness

² The primary focus of our literature review was on studies of concurrent meta-analyses because they provide the most commensurate basis for studying meta-analytic decisions. Alternatively, *meta-analytic updates* are less useful for this purpose, having been conducted after a substantial body of new literature on a topic has accumulated.

had the weakest effect among the five factors. Tett et al. suggested that the discrepancy might have been due to Barrick and Mount's failure to use absolute values when averaging effect sizes within primary studies or their inclusion of several additional criteria types and fewer dissertations. They also suggested that the two meta-analyses served different purposes, with their own meta-analysis focused primarily on a methodological rather than substantive issue. Alternatively, a later critique by Ones et al. (1994) traced the discrepant findings to several decisions made by Tett et al., such as their exclusion of exploratory studies in some analyses, unclear procedures regarding the assignment of personality scales to the FFM traits, and their failure to consider the moderating effect of job type.³ Ultimately, it was not well understood which factors contributed to the divergence of these meta-analyses (Barrick et al. 2001). Without full access to necessary information from both research groups, insights were based in part on speculation.

A more positive conclusion resulted from Barrick et al.'s (2001) comprehensive review of 11 meta-analyses on the relationships between FFM traits and performance criteria. Their quantitative review demonstrated that, with the exception of Tett et al.'s findings for agreeableness, emotional stability, and openness to experience, meta-analytic estimates of trait relationships with performance criteria differed only trivially from one meta-analysis to another. Therefore, they concluded that, "independently conducted meta-analytic studies are relatively consistent and perhaps more so than has been previously recognized" (p. 21).

Given the consistent meta-analytic findings described by Barrick et al. (2001), it might be possible that methodological improvements since the publication of Wanous et al. (1989) and Bullock and Svyantek (1985) have been effective toward reducing the subjective component in meta-analysis. These early methodological case studies served as an important stimulus for the development of guidelines for conducting, reporting, and evaluating meta-analyses (e.g., Campion 1993; Rothstein and McDaniel 1989). Recent reviews provide mixed evidence, suggesting that some aspects of meta-analytic practice have become more standardized than others (Aguinis et al. 2009; Dieckmann et al. 2009; Geyskens et al. 2009; Oyer 1997). More generally, it is recognized that meta-analytic guidelines can only go so far toward limiting the role of researcher judgment.

The findings of a recent study by Aguinis et al. (2009) suggest that further revision to existing guidelines may be

unnecessary. These authors examined the relationship between 21 researcher decisions and the magnitude of 5,581 effect size estimates from 196 meta-analyses published in several top organizational sciences journals between January 1982 and August 2009. Non-significant effects were found for the majority of decisions, and significant but not practically meaningful effects were found for the remainder. Aguinis et al. concluded that the majority of judgment calls do not affect conclusions in a substantial manner.

Although we do not wish to criticize the approach taken by Aguinis et al. (2009), we do feel that our case study complements and extends these authors' study in a few important ways. First, given that their study focused on the effect of researcher decisions in the aggregate (i.e., across meta-analyses from a wide range of topics), it is possible that important researcher effects at the level of matched pairs of meta-analyses were obscured within the larger pattern of results. In this way, our case study makes a unique contribution by focusing specifically on meta-analyses of the same topic. A second limitation was their lack of attention to the non-statistical decisions in meta-analysis. Because this is also true of the aforementioned studies, a unique objective of our study was to examine the consequences of early meta-analytic decisions.

Along these lines, Bobko and Roth (2008) called for additional methodological studies to address how conceptualization of the research question and constructs involved affects meta-analyses. Ironically, these influential aspects may be least amenable to standardization by methodological guidelines. Instead, they suggest the need for clearer thinking by meta-analysts on a case-by-case basis. An earlier paper by Stewart and Roth (2004) provides an illustration. Stewart and Roth re-examined Miner and Raju's (2004) meta-analytic update of their prior meta-analysis on entrepreneurial risk propensity (i.e., Stewart and Roth 2001), challenging these authors' inclusion of several primary studies that measured variables they considered to be irrelevant to the study's focus. By removing these irrelevant studies, Stewart and Roth yielded estimates much closer to those reported in their original meta-analysis.

Another issue with conceptual and statistical consequences for meta-analysis is the choice between fixed-, random-, or mixed-effects models (Bobko and Roth 2008). Whereas fixed-effects models (FE) assume that all study effects are from a single underlying population, random-effects (RE) models assume any number of underlying populations whose parameters may differ due to both random and true (e.g., variability due to un-measured moderators) sources of effect variability (Hunter and Schmidt 2000). A number of studies demonstrate that the mean estimates and confidence intervals yielded by FE and

³ In addition, Ones et al. identified several computational errors in the moderator analyses conducted by Tett et al., although those particular analyses were not relevant to the discrepant findings noted above.

RE models diverge as a function of increasing effect heterogeneity, with the latter being the more conservative method (Hunter and Schmidt 2000; Kisamore and Brannick 2008; Overton 1998). Beyond these statistical repercussions, model choice is guided by several factors including the extent to which important contextual variables are known (Overton 1998), the type of inference the meta-analyst wishes to make, and the number of studies that are available (Konstantopoulos and Hedges 2009).

In summary, although prior research clearly demonstrates that researcher decisions *can* affect meta-analytic conclusions (e.g., Geyskens et al. 2009), no general consensus appears to exist regarding whether meta-analytic reliability is routinely decreased as a result. Furthermore, attempts to understand the specific causes of divergence between meta-analyses have presented mixed evidence. In many cases, a lack of necessary information has precluded stronger conclusions, but more generally, the literature contains few explicit studies of meta-analytic decisions. Existing studies have focused on statistical decisions and have failed to address some of the more basic issues that might shape substantive outcomes. As a result, it is unclear whether additional methodological guidelines should be advanced, and if so, the specific stages of meta-analysis that should be targeted. Toward this end, the present case study makes a unique contribution by presenting a detailed account of the full range of meta-analytic decisions and their effect on the outcomes of three independent and concurrent meta-analyses of the same research topic.

The Present Study: A Serendipitous Case

Meta-analyses by Gajendran and Harrison (2007), Nicklin et al. (2009), and Nieminen et al. (2008) were each concerned with the employee-level outcomes (e.g., job satisfaction) associated with telecommuting, an alternative work arrangement in which employees complete regular work duties away from their traditional office location. For the remainder of the paper they are referred to as GH, NI, and NM, respectively. The meta-analyses were conducted independently, such that the three groups were comprised of non-overlapping researchers. Also, there was no correspondence between groups prior to the completion of the meta-analyses, nor were any of the groups aware that the other meta-analyses were being conducted. The meta-analyses were undertaken concurrently and completed in March of 2007 (GH) and August of 2007 (NI and NM).

Our case study approach afforded several advantages over prior methodological research. First, unlike the majority of previous studies, the meta-analyses studied here were not selected on the basis that they reported contradictory findings, but instead on serendipity alone.

This allowed for a fairer test of the degree to which independent meta-analyses with the same research question draw conclusions. To our knowledge, no prior study has reported a similar test using independent and concurrent meta-analyses on the same topic. Previous studies were potentially confounded by the accumulation of new primary studies (e.g., Barrick et al. 2001), differences in the research topics addressed (e.g., Aguinis et al. 2009), or because they focused specifically on contradictory meta-analyses (e.g., Wanous et al. 1989). Second, given the present authors' involvement in two of the three meta-analyses, this study was afforded greater access to detailed information than has been true of previous studies. Consequently, we were able to study a broader range of meta-analytic decisions, as opposed to being constrained to only those decisions that were described by authors in their published manuscripts (e.g., Aguinis et al. 2009). Third, the approach taken here was not focused on identifying differences alone. Rather, the broader intent was to examine the full range of researcher decisions occurring within each step of the meta-analytic process, including those where general consensus was apparent.

The 11-stage process model developed by Wanous et al. (1989) was used as a framework for comparing the three meta-analyses. The 11 stages can be thought of as representing early, middle, and late aspects of the meta-analytic process, where early stages (1–4) involve compiling a meta-analytic sample, middle stages (5–7) involve extracting and coding information from the collected studies, and late stages (8–11) involve executing the analysis and interpreting findings. Accordingly, the comparative analysis is discussed in “[Early Stages](#)”, “[Middle Stages](#)”, and “[Late Stages](#)” sections. Although attempts were made to use all three meta-analyses in all comparisons, there were some instances in which re-examination had to be limited to the current authors' meta-analyses (i.e., NI and NM) because of the availability of necessary information.⁴

Early Stages

Table 1 presents detailed findings of the comparative analysis for the early stages. Overall, it is apparent that at least some evidence of divergence was observed at all four stages, although the degree to which meta-analyses differed varied by stage. The most noteworthy difference between meta-analyses was observed at stage 4. Approximately one-third to one-half of the studies included were unique to a particular meta-analysis, and moreover, among the 62

⁴ Requests for additional information were sent to Gajendran and Harrison. However, in some cases necessary details were not available from these authors.

Table 1 Summary of comparative analysis for the early stages of meta-analysis

Description of key points of convergence and divergence across meta-analyses by stage
<p>Stage 1. Defining the research domain</p> <p>Definition of telecommuting</p> <p>Telecommuting was defined similarly across meta-analyses; <i>telecommuting</i> was not limited to a particular alternative work location (e.g., home versus satellite office) or frequency (i.e., part-time vs. full-time); however, contractors and self-employed persons with home offices were excluded</p> <p>Dependent variables studied</p> <p>Out of 14 total dependent variables studied, 4 were common to all three meta-analyses, 8 were common to at least 2 meta-analyses, and 4 were unique to a particular meta-analysis</p>
<p>Stage 2. Determining inclusion criteria</p> <p>Inclusion criteria</p> <p>All three meta-analyses required that primary studies included the following: (1) telecommuting as an independent variable, (2) at least one dependent variable of interest, and (3) necessary statistical information to compute an effect size. However, NM additionally required that studies include a non-telecommuter control group (i.e., traditional office workers)</p>
<p>Stage 3. Conducting the literature search</p> <p>Electronic databases^a</p> <p>The electronic databases searched included: PsycINFO (all), ERIC (all), Dissertation Abstracts International (all), EBSCO (GH, NI), JSTOR (GH, NI), ABI/Inform (GH, NM), Web of Science (GH, NM), Elsevier Science Direct (GH), Academic Ideal (GH), Sociological Abstracts (GH), Business Source Complete (NI), EconLit (NI), and PAIS International (NI)</p> <p>Search terms</p> <p>Search terms included: <i>telecommute</i> and iterations thereof (all), <i>telework</i> and iterations thereof (all), <i>flex-place</i> (all), <i>work at home</i> (all), <i>remote work</i> (GH, NM), <i>distributed work</i> (GH), <i>virtual work</i> (NI), <i>mobile work</i> (GH, NI)</p> <p>Manual searches</p> <p>All three meta-analyses searched reference lists of book chapters and reviews; GH and NI contacted known telecommuting researchers and searched conference programs for the Society of Industrial and Organizational Psychology (1995–2006) and the Academy of Management (1954–2006)</p> <p>Studies retrieved</p> <p>NI retrieved 325 studies, GH retrieved 212, and NM retrieved 115</p>
<p>Stage 4. Selecting the final sample of studies</p> <p>Final samples</p> <p>In total, 62 studies were included. GH included 46 studies ($k = 127$); NI included 32 studies ($k = 90$); NM included 22 studies ($k = 82$). Among GH's sample, 10 studies (22%) were common to all meta-analyses, 16 (35%) overlapped one other meta-analysis, and 20 (43%) were unique. Among NI's sample, 10 studies (32%) were common to all meta-analyses, 12 (39%) overlapped one other meta-analysis, and 9 (29%) were unique. Among NM's sample, 10 studies (45%) were common to all meta-analyses, 6 (27%) overlapped one other meta-analysis, and 6 (27%) were unique</p>

GH Gajendran and Harrison (2008), NI Nicklin et al. (2009), NM Nieminen et al. (2008)

^a Cases in which similarity was observed across all three meta-analyses are denoted "all"

total studies, only 10 were included in all three meta-analyses.⁵ Thus, the meta-analyses yielded samples that differed considerably, despite sharing the same general purpose. The remainder of this section focuses on why such differences occurred.

Bobko and Roth (2008) described how conceptualization of research questions and constructs guide subsequent decisions. Consistent with their comments, our re-examination underscored two "conceptual" differences prior to the literature search, both of which were identified as plausible reasons for the divergence in meta-analytic samples. First, different strategies were adopted for specifying

the dependent variables of interest. GH advanced a model in which variables associated with autonomy, work–family interface, and relationship quality were posited to mediate the relationship between telecommuting and more distal work-related outcomes. NI used the job characteristics model (e.g., Hackman and Oldham 1975) and self-determination theory (e.g., Ryan and Deci 2000) as their basis for selecting variables and specifying hypotheses. Alternatively, NM adopted a more empirical strategy, focusing primarily on the outcomes that had been most commonly studied and those demonstrating mixed findings. As a result, moderate divergence was noted in terms of the dependent variables studied in each meta-analysis, with NM including the most dependent variables, followed by NI and GH. This did not, however, lead to a larger sample for NM.

⁵ A full listing of the studies included in each meta-analysis are available upon request.

A second key difference involved the inclusion criteria that were used. Two types of studies were most commonly observed in the telecommuting literature: (1) studies that compare telecommuters to non-telecommuters, and (2) studies that compare telecommuters of different frequencies along a continuum. NM argued that combining both types of studies was inappropriate because they represent slightly different operationalizations of the telecommuting variable and included only those studies with a non-telecommuting control group. Alternatively, GH and NI combined both types of studies in their meta-analyses and later demonstrated via categorical moderation tests that effects did not differ consistently as a function of study type. In order to determine the contribution of these two factors to the divergence observed in meta-analytic samples, we re-examined the 31 studies that were non-overlapping between NI and NM. Unfortunately, lack of necessary information prohibited a similar analysis involving GH. Among the twenty studies included by NI but not NM, ten were not located by NM's literature search; three were located by NM's literature search, but attempts to retrieve the studies were unsuccessful; six were retrieved but were discarded due to the absence of a non-telecommuter group; and finally, one was retrieved but judged not to meet NM's operational definition of telecommuting. Among the 11 studies included by NM but not NI, nine were not located by NI's literature search; two were located and retrieved, but later discarded due to NI's judgment that the studies were redundant with two other studies. Re-analysis failed to locate an instance in which a non-overlapping study was attributed to differences in the dependent variables of interest to these meta-analyses.

In considering the reason for these differences, there was some evidence to indicate that NI's search was more comprehensive than NM's. For example, NI's search was more exhaustive of unpublished sources and retrieved many more studies initially for closer inspection. However, that NM also located nine unique studies suggests that rigor alone cannot account for the observed differences and underscores the idiosyncratic nature of literature searching (Greenhalgh and Peacock 2005).

Middle Stages

Table 2 presents detailed findings of the comparative analysis for the middle stages. Although similar decision rules were noted for handling various study designs (e.g., longitudinal studies), several cases were noted in which different effect sizes were extracted. These differences were attributed to researchers' operationalization of the constructs involved. For example, in one instance NI extracted separate effects for *personal* strain and *work*

strain (and later aggregated these), whereas NM extracted the effect for *work* strain only. Likewise, NI focused on affective and overall commitment, whereas NM derived an average of affective, normative, and continuance commitment whenever facet-level data were available. Different perspectives on whether or not to focus at the global versus facet level similarly influenced how effects were extracted and handled thereafter for job performance and work–family conflict (see stage 7). GH used moderation tests to aid their decisions herein (e.g., non-significant moderation led to the combination of work-to-family and family-to-work conflict), whereas NI and NM based their decisions on sample size and conceptual understanding of the constructs involved.

In terms of stage 6, differences were observed for the potential moderator variables that were selected and the coding rules that were applied. Specifically, four of the seven total moderators tested were common to at least two meta-analyses, and among these, slightly different coding strategies were noted for three. Our subsequent interest was in understanding the extent to which these differences and those noted for previous stages affected meta-analytic findings.

Late Stages

Table 3 summarizes the comparative analysis of researcher decisions during the late stages, and Tables 4 and 5 present the meta-analyses' overall effects and moderation findings, respectively.

Overall Effects

The overall effects shown in Table 4 reaffirm that telecommuting generally has a small, beneficial effect on a range of employee outcomes, although the potential for moderation was evident in the majority of cases. Prior decisions led the meta-analyses to examine different criteria or to operationalize criteria differently, rendering direct comparisons difficult in some cases. For example, GH's decision to separate self-rated performance from other types of performance criteria resulted in a unique finding in comparison to the other meta-analyses. Whereas the effects for performance criteria were generally small and positive, GH found a somewhat larger effect for their non-self-rating category. Consequently, GH's conclusions regarding the effect of telecommuting on performance were somewhat different than NI and NM, not as a function of contradictory findings but because GH addressed an additional issue not considered by the others.

On the other hand, whenever direct comparisons were possible, a high degree of convergence was observed, with one exception that is described in more detail

Table 2 Summary of comparative analysis for the middle stages of meta-analysis

Description of key points of convergence and divergence across meta-analyses by stage
<p>Stage 5. Extracting statistical information</p> <p>Handling variety in study designs^a</p> <p>For cross-sectional designs, means, standard deviations, sample sizes for telecommuter and non-telecommuter groups were extracted (all). When data were reported separately for multiple telecommuter groups, a single aggregate telecommuter group was formed (all). For studies that did not include a non-telecommuter control group, GH and NI extracted correlations representing the relationship between frequency of telecommuting and dependent variables. For studies that used longitudinal designs, all three meta-analyses extracted cross-sectional effect size information for the latest available time point (i.e., telecommuters vs. non-telecommuters)^b</p> <p>Handling multiple dependent variables^c</p> <p>In general, when multiple measures of a dependent variable were reported, GH derived the median value, whereas NI and NM used the mean value. In comparing the effect sizes that were extracted from studies common to NI and NM, differences were noted in 2 of 9 instances for job satisfaction, 2 of 6 for job performance/productivity, 1 of 4 for work-to-family conflict (WTF) and family-to-work conflict (FTW), and 1 of 7 for stress</p> <p>Stage 6. Coding</p> <p>NI and NM used multiple independent coders and solved any discrepancies through consensus meetings. GH used multiple coders and solved any discrepancies through a third party referee. Inter-coder agreement was 88% for NI, 100% for NM, and not reported by GH</p> <p>Variables coded</p> <p>The following seven potential moderators were coded: publication status (all), gender (all), frequency of telecommuting (GH, NI), dichotomous versus continuous operationalization of the telecommuting variable (GH, NI), amount of experience with telecommuting (GH), age (NI), professional vs. technical job type (NI)^d</p> <p>Coding rules</p> <p>Different coding strategies were noted for <i>publication status</i> (NI and NM coded published vs. unpublished, whereas GH coded into three categories based on the methodological rigor of peer review), <i>gender</i> (NM and NI coded predominantly male vs. predominantly female, whereas GH coded the percentage of females as a continuous variable) and <i>frequency of telecommuting</i> (GH and NI coded frequency as either high vs. low, where GH defined high frequency as greater than 2.5 days per week, and NI defined high frequency as greater than 30 h per week)</p> <p>Stage 7. Decisions regarding grouping of variables^e</p> <p>For <i>job performance</i>, GH formed separate categories for self-rated performance vs. objective or supervisor-rated performance; NI formed separate categories for overall, productivity, and effectiveness criteria; NM averaged across all available indicators of performance to form an overall job performance criterion</p> <p>For <i>work-family conflict</i>, all three meta-analyses coded WTF and FTW separately. However, GH ultimately combined WTF and FTW conflict (after a non-significant moderation test) with work-life balance (reverse coded); NI separated WTF and FTW constructs; NM separated WTF, FTW and work-life balance variables^f</p>

GH Gajendran and Harrison (2008), NI Nicklin et al. (2009), NM Nieminen et al. (2008)

^a Cases in which similarity was observed across all three meta-analyses are denoted “all.” ^b NM included three longitudinal studies, two of which were included by GH and one of which was included by NI. ^c Only differences in extraction strategies were noted here. For all other variables, NI and NM handled the occurrence of multiple dependent variables similarly. ^d All three meta-analyses reported coding additional variables, which could not be tested as moderators due to insufficient sample, including: whether entry into the telecommuting arrangement was voluntary, task interdependence, industry, organization and work-group characteristics (e.g., number of employees), employee level, technology-related variables, characteristics of the home office (e.g., whether or not the telecommuter has a separate home-office workspace). ^e Only differences in grouping of variables were noted here. All other variables were grouped using similar strategies. ^f Work-life balance was defined by Cree (1999) as “being able to meet family (personal) responsibilities while at the same time meeting the demands of the job” (p. 166). Thus, there is conceptual overlap between work-family balance and work-family conflict. However, NM maintained a distinction between them because work-family balance is not readily combined with work-to-family conflict (WTF) or family-to-work conflict (FTW), both of which imply a direction of the role conflict experienced (i.e., work conflicting with family or family conflicting with work), whereas work-life balance is non-directional

below—NM’s finding for role stress. Specifically, the maximum difference observed in terms of uncorrected sample-weighted mean r ’s was .05 (excluding role stress), and more commonly estimates differed by only .01 or .02. Moreover, it is apparent that GH and NI’s decision to correct estimates for criterion unreliability had little effect on meta-analytic findings. Thus, the majority of overall effects differed only trivially across meta-analyses for those variables that were directly comparable.

As noted above, an important exception was observed for the outcome variable role stress. Specifically, GH and NI found a small, beneficial effect of telecommuting on role stress, whereas NM found a small effect in the opposite direction. In order to determine the source of this discrepancy, we re-examined the studies that contributed to the role stress findings reported by NI and NM. Results suggested two key points of divergence: (1) three non-overlapping primary studies, and (2) the aforementioned

Table 3 Summary of comparative analysis for the late stages of meta-analysis

Description of key points of convergence and divergence across meta-analyses by stage
<p>Stage 8. Estimating overall effects</p> <p>Meta-analytic model and statistical corrections</p> <p>GH and NM followed Hedges and Olkin's (1985) method with a fixed-effects model, whereas NI followed Hunter and Schmidt's (2004) method with a random-effects model. Consistent with a "bare bones" approach to meta-analysis, NM reported sample-weighted mean effects. Alternatively, GH and NI corrected estimates for criterion unreliability. For studies that did not contain reliability information, GH and NI imputed the mean reliability of other studies that measured the same variable</p> <p>Findings</p> <p>See Table 4</p>
<p>Stage 9. Determining if moderation tests are warranted</p> <p>Heterogeneity of effects</p> <p>Using Hunter and Schmidt's 75% rule, NI concluded that moderator tests were justified in all cases. Alternatively, GH and NM applied Hedge's and Olkin's <i>Q</i>-statistic method, but with slightly different strategies. GH applied <i>Q</i>-tests at the variable level, and found sufficient heterogeneity to search for moderators in all cases, except for autonomy, quality of relationship with supervisor, and job satisfaction. In contrast, NM applied <i>Q</i>-tests at the grouped variable level and found sufficient heterogeneity to search for moderators within work, non-work, and job characteristics domains^a</p>
<p>Stage 10. Selecting moderators</p> <p>Rationale—theory driven vs. exploratory</p> <p>Theoretical rationales were provided for moderation analyses involving gender (all), telecommuting frequency (GH, NI), mean age (NI), amount of experience with telecommuting (GH), and job type (NI). Publication status (all) and dichotomous vs. continuous operationalizations of telecommuting (GH, NI) were exploratory with stated methodological purposes^b</p> <p>Sample size cut-off rules</p> <p>GH and NM required at least four effects at each level of categorical moderators to justify moderation analyses, whereas NI required three effects at each level</p>
<p>Stage 11. Conducting moderation tests</p> <p>Analysis strategy</p> <p>GH conducted moderation tests within variables and in some cases collapsing across all variables (i.e., at the study level); NI conducted moderation tests within variables only; NM conducted tests within grouped variables or domains^c</p> <p>Moderation findings</p> <p>See Table 5</p>

GH Gajendran and Harrison (2008), NI Nicklin et al. (2009), NM Nieminen et al. (2008)

^a In NM's meta-analysis, the *work* domain was a composite of job performance, job satisfaction, organizational commitment, work stress, turnover intentions, and perceived career advancement criteria. The *non-work* domain was a composite of work-life balance, work-to-family conflict and family-to-work conflict criteria. The *job characteristics* domain was a composite of flexibility, autonomy, and total work hours criteria. Due to small sample sizes, moderation tests proceeded at the domain level rather than within individual variables. ^b Cases in which similarity was observed across all three meta-analyses were denoted "all." ^c For NM, moderation analyses proceeded within grouped variables or domains, corresponding to *work* (job performance, job satisfaction, organizational commitment, work stress, turnover intentions, and perceived career advancement opportunities), *non-work* (work-to-family conflict, family-to-work conflict, and work-life balance) and *job characteristics* (flexibility, autonomy, and total work hours) criteria types

case in which different effects were extracted from the same study. Specifically, two studies that were included by NI (but not NM) reported reduced stress associated with telecommuting ($r = -.21$ from an unpublished government data set and $r = -.13$ from Raghuram and Weisenfeld 2004). Furthermore, the unpublished study was particularly influential in NI's meta-analysis due to its much larger sample size ($N = 1,635$) than the other studies included. Conversely, the one study that was unique to NM reported increased stress associated with telecommuting ($r = .08$ from Staples 2001). In this way, the pattern of non-overlapping studies happened to maximize the difference between NI and NM's finding for role stress.

Finally, although it is not advisable to over interpret the meaningfulness of significance tests for effects that are both small and based on very large samples (Cortina and Dunlap 1997; Eden 2002), a few comments regarding the similarity of statistical inferences across meta-analyses are warranted. Among the nine dependent variables for which a comparison across meta-analyses was possible, there were two cases in which different statistical inferences were made. Specifically, NM found non-zero effects for job satisfaction and job performance (as evidence by 95% confidence intervals that did not include 0), whereas NI's estimates for these variables were non-significant, despite that the magnitude of effects were very similar across meta-analyses (r 's = .04 for NM versus r 's = .06 for NI).

Table 4 Effects of telecommuting on employee outcomes by meta-analysis

Employee outcomes	Gajendran and Harrison (2007)				Nicklin et al. (2009)				Nieminen et al. (2008)			
	<i>k</i>	<i>n</i>	Mean <i>r</i> (Est. ρ) ^a	Mean <i>r</i> , 95% CI	<i>k</i>	<i>n</i>	Mean <i>r</i> (Est. ρ) ^a	Mean <i>r</i> , 95% CI	<i>k</i>	<i>n</i>	Mean <i>r</i>	Mean <i>r</i> , 95% CI
Autonomy	11	3,040	.19 (.22) ^b	.16, .22					7	3,911	.14 ^b	.11, .17
Flexibility									6	1,963	.31 ^b	.27, .35
Job satisfaction	28	7,764	.09 (.10) ^b	.07, .11	21	12,131	.06 (.07)	-.02, .15	15	5,412	.04 ^b	.01, .07
Organizational commitment					10	10,982	.03 (.04)	-.03, .09	6	2,669	.04	.00, .08
Perceived career advancement	8	1,038	.00 (.00)	-.06, .07					4	5,879	.05 ^b	.02, .08
Performance (composite)					18	11,439	.06 (.06)	-.01, .14	11	7,461	.04 ^b	.02, .06
Self-rated	9	7,419	.01 (.01)	-.01, .03								
Supervisor-rated or objective	4	484	.18 (.19) ^b	.09, .26								
Productivity					7	1,191	-.01 (-.01)	-.16, .14				
Effectiveness					11	10,248	.07 (.07) ^b	.01, .13				
Quality of relationship- supervisor	14	2,888	.12 (.12) ^b	.08, .15								
Quality of relationship- co-worker	14	3,269	.00 (.00)	-.03, .03								
Role stress	11	2,406	-.11 (-.13) ^b	-.15, -.07	7	3,240	-.14 (-.15) ^b	-.07, -.25	6	1,493	.07 ^b	.02, .12
Turnover intent	9	7,580	-.08 (-.10) ^b	-.11, -.06	8	4,567	-.09 (-.10) ^b	-.18, -.01	5	6,179	-.10 ^b	-.12, -.07
Work–family conflict (bi-directional)	19	9,852	-.11 (-.13) ^b	-.13, -.10								
WTF conflict ^c					18	10,342	-.11 (-.12) ^b	-.20, -.04	5	794	-.09*	-.16, -.02
FTW conflict ^d					7	1,704	-.08 (-.09)	-.22, .03	4	633	-.07	-.15, .01
Work-life balance ^e									5	6,438	-.03 ^b	-.05, -.01
Work hours									8	7,274	.16 ^b	.14, .18

Gajendran and Harrison (2007) and Nieminen et al. (2008) followed Hedges and Olkin's (1985) method and used a fixed-effects model, whereas Nicklin et al. (2009) followed Hunter and Schmidt's (2004) method and used a random-effects model

^a Estimated population correlation (ρ) is shown in parentheses. Estimates were corrected for attenuation due to unreliability in the criteria measures. ^b indicates that the 95% confidence interval around the sample-weighted mean *r* does not include zero. ^c WTF conflict refers to conflict due to work interfering with family. ^d FTW conflict refers to conflict due to family interfering with work. ^e Work-life balance was defined by Cree (1999) as "being able to meet family (personal) responsibilities while at the same time meeting the demands of the job" (p. 166)

Although this difference did not lead to substantively different conclusions in the original papers regarding the importance of these effects, it is nonetheless interesting to note the role of computational differences. When NM's overall effects were re-computed using Hunter and Schmidt's (2004) formulas and a random-effects model (as was done by NI), four previously significant effects were no longer significant: job satisfaction ($r = .03$, 95% CI = $-.04$ to $.11$), job performance ($r = .03$, 95% CI = $-.02$ to $.07$), turnover intent ($r = -.09$, 95% CI = $-.05$ to $.24$), and work-life balance ($r = -.02$, 95% CI = $-.05$ to $.00$). Thus, NM's use of Hedges and Olkin's (1985) formulas with a fixed-effects model was a generally less conservative approach, although the resulting differences were negligible.

Moderation findings

All three meta-analyses found consistent evidence of effect heterogeneity (see stage 10, Table 3), which suggests the presence of moderators. The most apparent form of divergence at this stage involved the *focus* of moderation tests. Table 5 summarizes moderation findings for the four moderators that were common to at least two of the meta-analyses, including gender, operationalization of telecommuting, telecommuting frequency, and publication status. Even after limiting re-examination to these overlapping moderators, direct comparison across meta-analyses was complicated by differences in the dependent variables studied, the way that moderators were coded (see stage 6, Table 2), and the analysis strategies that were used

Table 5 Summary of moderation findings by meta-analysis

Moderator	Summary of findings
Gender of sample ^a	<p>GH: Percentage of females was positively associated with the effect of telecommuting on <i>supervisor-rated performance</i> and <i>perceived career prospects</i>. Gender did not moderate any other study variables</p> <p>NI: Gender moderated the effect of telecommuting on <i>role stress</i> (female: $r = -.18$, 95% CI = $-.26$ to $-.13$ vs. male: $r = -.02$, 95% CI = $-.02$ to $.02$). Weak support^c was found for the moderating effect of gender on <i>WTF conflict</i>^f (female: $r = -.07$, 95% CI = $-.19$ to $.04$ vs. male: $r = -.13$, 95% CI = $-.08$ to $-.21$). Gender did not moderate <i>job satisfaction</i> or <i>job performance</i></p> <p>NM: Gender moderated the effect of telecommuting on aggregated <i>job characteristics</i>^g variables (female: $r = .16$, 95% CI = $.14$ to $.18$ vs. male: $r = .10$, 95% CI = $.06$ to $.14$). Gender did not moderate the effect of telecommuting on aggregated <i>work</i>^h or <i>non-work</i>ⁱ variables</p>
Operationalization of telecommuting ^b	<p>GH: Operationalization did not moderate effects at the study level^j</p> <p>NI: Weak support was found for the moderating effect of operationalization on <i>turnover intentions</i> (amount: $r = -.11$, 95% CI = $-.19$ to $-.04$ vs. dichotomous: $r = -.05$, 95% CI = $-.14$ to $.04$). Operationalization did not moderate <i>job satisfaction</i>, <i>commitment</i>, <i>job performance</i>, or <i>WTF conflict</i></p>
Telecommuting frequency ^c	<p>GH: Frequency moderated the effect of telecommuting on <i>work–family conflict</i> (high: $r = -.16$, 95% CI = $-.18$ to $-.13$ vs. low: $r = -.05$, 95% CI = $-.10$ to $.00$) and <i>quality of relationship with coworker</i> (high: $r = -.19$, 95% CI = $-.30$ to $-.08$ vs. low: $r = .03$, 95% CI = $-.01$ to $.07$). Frequency did not moderate <i>autonomy</i> or <i>quality of relationship with supervisor</i></p> <p>NI: Weak support^c was found for the moderating effect of frequency on <i>WTF conflict</i> (high: $r = -.01$, 95% CI = $-.12$ to $.10$ vs. low: $r = -.13$, 95% CI = $-.25$ to $-.05$)</p>
Publication status ^d	<p>GH: Publication status did not moderate effects at the study level^j</p> <p>NI: Weak support was found for the moderating effect of publication status on <i>turnover intentions</i> (published: $r = -.04$, 95% CI = $-.14$ to $.06$ vs. unpublished: $r = -.11$, 95% CI = $-.19$ to $-.05$) and <i>job performance</i> (published: $r = .14$, 95% CI = $.01$ to $.28$ vs. unpublished: $r = .04$, 95% CI = $-.01$ to $.11$). Publication status did not moderate <i>job satisfaction</i>, <i>FTW conflict</i>^k, or <i>WTF conflict</i></p> <p>NM: Publication status moderated the effect of telecommuting on aggregated <i>non-work</i> variables (published: $r = .02$, 95% CI = $.00$ to $.04$ vs. unpublished: $r = -.11$, 95% CI = $-.18$ to $-.04$). Publication status did not moderate the effect of telecommuting on aggregated <i>work</i> or <i>job characteristics</i> variables</p>

GH Gajendran and Harrison (2008), NI Nicklin et al. (2009), NM Nieminen et al. (2008)

^a GH coded the percentage of females in a study as a continuous moderator. NI and NM coded studies as either majority *male* or majority *female*.

^b GH and NI coded studies as to whether telecommuting was operationalized as a continuous variable reflecting *amount* of telecommuting or as a *dichotomous* variable reflecting the comparison of telecommuters to traditional office workers. ^c GH coded studies as *high* frequency when the sample telecommuted 2.5 days per week or more on average. NI coded studies as *high* frequency when the sample telecommuted 30 h per week or more on average. ^d GH coded publication status in three categories: journal articles; book chapters and conference papers; unpublished dissertations. NI and NM coded publication status as *published* or *unpublished*. ^e Here, the term “weak support” is used to describe a case in which the confidence intervals associated with effects at the categorical level overlap, but the confidence interval for one of the effects includes 0, whereas the other does not include 0. ^f WTF conflict is conflict resulting from work interfering with family life. ^g The *job characteristics* domain was a combination of flexibility, autonomy, and total work hours criteria. ^h *Work* domain was a composite of job performance, job satisfaction, organizational commitment, work stress, turnover intentions, and perceived career advancement criteria. ⁱ The *non-work* domain was a composite of work-life balance, work-to-family conflict and family-to-work conflict criteria. ^j GH tested for moderation of studies’ operationalization of telecommuting and publication status at the study level (i.e., across variables) rather than within study variables. ^k FTW conflict is conflict resulting from family life interfering with work

(see stage 11, Table 3). Thus, an important point to be taken away from Table 5 is that earlier decisions led to eventual differences in terms of the questions that were addressed by moderation analyses.

Differences notwithstanding, we identified a subset of analyses described in Table 5 as providing the most commensurate basis for comparison among meta-analyses. These cases involved gender and telecommuting frequency as moderators, both of which were tested by GH and NI at the *variable-level* (e.g., does frequency moderate the effect of telecommuting on *work–family conflict*). First, with respect to the moderating effect of gender, NI found weak support for moderation of work–family conflict, whereas

GH found no support.⁶ Neither GH nor NI found support for moderation of job satisfaction. Lastly, GH found support for moderation of supervisor-rated performance (but not for self-rated performance), whereas NI found no support for moderation of overall performance. Second, with respect to the moderating effect of telecommuting

⁶ The term “weak support” is used to describe a situation when the confidence intervals for two estimated mean effects at different levels of a categorical moderator variable overlap one another, but only one set of confidence intervals includes 0. This situation implies that the effect of telecommuting only differs significantly from zero for one level of the moderator variable, but that in a stricter sense, the estimates for each level do not differ significantly from one another.

frequency, both GH and NI found some support for moderation of work–family conflict, although the nature of the moderating effect differed. Specifically, GH's analysis indicated that high frequency telecommuters experienced a greater reduction in work–family conflict than low frequency telecommuters, whereas NI's analysis found weak support for the opposite. However, the marginal nature of NI's finding, suggests that the importance of this difference should not be overstated. Nonetheless, it is apparent that moderation findings demonstrated less convergence than the overall effects.

Beyond this subset of tests, the meta-analyses were generally consistent in demonstrating weak or non-supportive moderation findings, despite clear evidence of effect heterogeneity. Excluding the five findings involving weak support, moderation was supported in only seven out of a total of 35 moderation tests. Furthermore, we question the practical meaningfulness of some of these findings, and by extension, the degree to which they should be weighted in a cross-meta-analysis comparison. For example, NM found that predominantly male studies differed from predominantly female studies by .06 for the effect on aggregated job characteristics. Because the estimates for both were small, this represents a very modest difference in terms of variance accounted for. Given this rather modest difference, it is not particularly alarming that GH did not find a similar effect for autonomy, although operational differences were again difficult to disentangle (e.g., NM's use of domain-level analysis versus GH's use of variable-level analysis). Thus, our re-examination generally fails to find examples of stable and/or consistent patterns of moderation that differ across the meta-analyses, although the operational differences described prevented stronger conclusions.

Discussion

This study took advantage of a serendipitous and unprecedented scenario in which three telecommuting meta-analyses were conducted concurrently by independent research teams. Overall, our re-examination demonstrated that researcher decisions guided the meta-analytic process at nearly every turn, including some decision points that were not previously recognized by Wanous et al. (1989). Beyond simply documenting the ubiquity of the judgment calls in these meta-analyses, we were more concerned with how often and to what extent these decisions were handled differently (or similarly) by independent meta-analysts. Critics of meta-analysis have focused on the number of researcher decisions that enter into the meta-analytic process (e.g., Briggs 2005), without considering the possibility that researchers handle these decisions similarly.

Furthermore, even in those instances where different decisions are made, it is not necessarily the case that meta-analytic outcomes will diverge as a consequence. As a test of this possibility, our re-examination studied the sensitivity of meta-analytic findings to differences in researcher decisions.

First, some evidence of divergence between meta-analyses was noted at nearly every stage of the meta-analytic process, from the parameters used in literature searches to the specific strategies used for conducting moderation tests. However, within each of these stages, a fair degree of convergence was also apparent. For example, although a handful of the dependent variables and moderators examined were unique to a particular meta-analysis, the majority were common to two or more of the meta-analyses. Likewise, the decision rules noted for handling different primary study designs, for coding moderators, and when setting sample size cut-offs for categorical moderation tests reflected more similarity than dissimilarity. Therefore, it is apparent that the meta-analyses diverged in some ways and converged in others; however, in many cases, we would classify the divergence observed as being relatively minor. This is an important finding given that previous research has focused exclusively on identifying decision points that were handled differently.

The question of whether or not these deviations in process manifested in different meta-analytic outcomes needs to be addressed at two levels. On one level, our case study was primarily concerned with making *apples-to-apples* comparisons by focusing on instances in which the meta-analyses could be directly compared due to having assessed the same dependent variables, the same moderators, and so on. Instances of contradictory findings at this level have been the dominant focus of prior writing on this topic because of the direct implications for reliability and validity of meta-analysis as a method. For example, meta-analysts are compelled to speculate on the reasons behind any findings that appear directly at odds with prior meta-analyses. Interestingly, our re-examination generally failed to find meaningful differences for these *apples-to-apples* comparisons, with just a few possible exceptions (e.g., NM's finding for work stress and the moderating effect of telecommuting frequency on work–family conflict). Most notably, we found a high degree of convergence for the effects of telecommuting across a range of dependent variables. On the basis of this overall similarity, the impact of earlier researcher decisions appears to have been minimal.

At a second level, it was interesting to observe how the meta-analyses evolved to adopt different focuses, also as a function of earlier researcher decisions. For example, judgment calls made early on (e.g., the theoretical frameworks adopted, the inclusion criteria that were used,

the focus of variable operationalizations, etc.) led the meta-analyses to adopt different eventual directions in terms of the dependent variables that were studied and the moderation tests that were possible. Although this limited our ability to make across-meta-analysis comparisons in several cases, this was in and of itself important evidence of divergence among the meta-analyses. Moreover, these differences in *direction* generally overshadowed the magnitude of differences in directly comparable findings. Certainly a single case study does not overturn previous examples that have demonstrated more direct consequences of meta-analytic decisions (e.g., Bullock and Svyantek 1985; Geyskens et al. 2009; Stewart and Roth 2004). However, it does suggest that an important consequence of subjectivity has less to do with the reliability of meta-analytic methods and more to do with how researchers guide their meta-analyses toward unique questions.

Our findings deserve some further clarification given recent studies by Aguinis et al. (2009) and Geyskens et al. (2009) that seem contradictory in their conclusions about the seriousness of researcher decisions for meta-analytic practice. Upon consideration of all three studies, a few general conclusions seem warranted. First, like previous data simulation studies (but using actual meta-analytic data), Geyskens et al. demonstrated the *potential* effects of different researcher choices and concluded that certain researcher decisions *if handled differently could* affect meta-analytic outcomes. However, our findings complement and extend those of Aguinis et al. by demonstrating that these potential differences are often not realized among actual meta-analyses, or at least not in the direct manner that is frequently assumed. One reason, which had not been fully demonstrated until our case study, is that many decisions are not handled differently by independent meta-analysts. Another possibility is that decisions can have a cancelling effect on one another across meta-analyses (e.g., when some decisions cause an upward bias and others cause a downward bias). Taken together, it is perhaps unsurprising that studies of hypothetical decisions have generally yielded more severe conclusions than studies of actual decisions across completed meta-analyses.

By viewing the findings of our case study in tandem with those of previous research, we identified three emergent themes with important implications for the reliability of meta-analytic methods. By doing so, we hope to draw attention to several generally overlooked decisions within the meta-analytic process (i.e., in comparison to statistical decisions) that might be targeted as important avenues for future methodological research. Although some specific guidelines for meta-analytic practice can be inferred (e.g., meta-analysts should incorporate sensitivity analysis

more broadly), readers are referred elsewhere for additional guidelines with respect to other decision points (e.g., Dieckmann et al. 2009; Geyskens et al. 2009).

Meta-analytic Sampling

There is accumulating evidence to suggest that despite researchers' best intentions, meta-analytic samples are rarely exhaustive of the available primary studies on a particular topic (Bullock and Svyantek 1985; Dieckmann et al. 2009; Wanous et al. 1989). The present study provides a striking example of this. By re-examining the non-overlapping studies between NI and NM, it appears that failure to locate studies was primarily responsible. Although differences were observed in terms of the databases and search terms that were used, some research suggests that protocol-driven searches (i.e., using pre-specified combinations of search terms and databases) account for a relatively small percentage of the studies that meta-analysts identify in comparison to searching reference lists, personal knowledge, or serendipitous discovery (Greenhalgh and Peacock 2005). Thus, although it might be possible to standardize some literature search procedures, such as protocol-driven searching, it is unlikely that this will result in substantially better coverage of the primary literature.

Alternatively, it may be useful to conceptualize meta-analytic sampling using the same principles that guide sampling theory in single-sample situations (e.g., random sampling and power analysis). Following from this perspective, Cortina (2002) described the diminishing returns of increasing k beyond a number that is sufficient to represent the population of studies accurately. In other words, beyond the point at which k is sufficient for drawing a reliable estimate, the accumulation of additional data points is largely inconsequential. Alternatively, when the number of studies in a domain is relatively small, estimates can be very sensitive to the addition of one or two data points. Therefore, it is the small- k situations in which an exhaustive sampling strategy is most important. Effect heterogeneity within a domain should also be considered because the potential influence of additional data points is increased when drawn from either a single population with wide variability or multiple populations, as when moderators are present.

Thus, although working with a larger subset of available studies is usually preferable from a reliability and validity standpoint (Dieckmann et al. 2009), including all relevant studies may be both unnecessary and not practically feasible under certain conditions. Particularly when the number of studies in a domain has grown to an otherwise prohibitive quantity, it may prove beneficial to focus on improving methods for drawing representative, rather than

exhaustive, samples.⁷ For example, Cortina (2002) suggested focusing on adequate inclusion of studies at each level of suspected moderators. The quality of primary studies is another important factor to consider. For example, some have called for more restrictive sampling techniques that result in fewer, higher quality studies (e.g., Bobko and Roth 2008). However, determining study quality can be complex as underscored by recent alternatives to using publication status as a proxy for quality (e.g., Wells and Littell 2009). Future methodological studies should seek to clarify how best to draw a representative sample of high quality studies in balance with maintaining the rigor of the methods used.

Sensitivity Analysis

Similarly, Allen and Preiss (1993) suggested that meta-analytic replications need not include the exact same studies in order to yield the same conclusions. In many ways, our case study, in which only 10 studies or less than 20% of the total sample, were common to all three meta-analyses, provides further support for this position. At the same time, the discrepant findings observed for role stress provide an important counter-example, whereby meta-analysts' handling of a small number of influential studies was responsible for the divergence that was observed. Similar cases in which a contradictory finding was traced to the inclusion of one or two unique studies have been reported elsewhere (e.g., Wanous et al. 1989). Although such instances likely result from one meta-analyst's failure to locate particular studies, the discrepancy might alternatively be traced to researchers' handling of outlier analysis. However, outlier analysis is not likely to flag a small set of outlying studies working in concert to bias a mean estimate, nor is it likely to flag a moderately heterogeneous effect with a disproportionately large sample size. Therefore, using outlier diagnostics alone may provide an incomplete method for examining the degree to which individual effects are influencing results.

From a broader perspective, a sensitivity analysis approach to meta-analysis is preferred by some (e.g., Dieckmann et al. 2009; Greenhouse and Iyengar 1994). The logic of sensitivity analysis is reflected in Cortina's (2003) recommendation that, "there is little to be lost by presenting results with and without outliers" (p. 434). Although outlier analysis is one important aspect of sensitivity analysis, the full strategy calls for thinking

about the influence of a variety of decisions made by the meta-analyst, and when appropriate, qualifying the description of results to reflect the influence that such decisions may have had (see Greenhouse and Iyengar 1994). For example, sensitivity analysis might entail reporting estimates for alternative statistical models or inclusion criteria if results would have differed. Recent work has also championed the use of improved methods (e.g., trim and fill, funnel plots) for detecting and accounting for a possible publication bias or file-drawer bias as part of a sensitivity analysis (Geyskens et al. 2009; Kromrey and Rendina-Gobioff 2006; Rothstein et al. 2005). In light of concerns over publication bias in meta-analysis, it is perhaps reassuring to note the general pattern of small effects (i.e., $<.20$) that was observed across all three meta-analyses. Although inflated estimates were not a primary concern in these meta-analyses, categorical moderation tests were used to determine (and ultimately defend) the meta-analysts' decision to include unpublished studies. A similar contingency perspective would be beneficial for other potentially important factors (e.g., study quality). In summary, more fully integrating sensitivity analysis across a range of these decision types, should help meta-analysts to avoid or at least acknowledge controversial decisions, in addition to providing research consumers with specific information to judge the reliability of meta-analytic findings.

Moderation Tests

Although our re-examination of the moderation analyses mainly reinforces our earlier point that researcher decisions led the meta-analyses to examine different issues, it also underscores the potential for increased sensitivity to researcher decisions. The reason for this conclusion is twofold. First, several moderation tests involved subgroup analyses with small samples comprising one or more groups. In this situation the influence of one or two unique studies can be substantial, such that the decisions leading up to the analysis are similarly magnified in importance. Second, there are several additional opportunities for researcher influence preceding moderation analysis in comparison to estimation of the overall effects. Still, given that our case study did not find instances of clear discrepancy, with one possible exception being GH and NI's finding regarding the moderating effect of telecommuting frequency on work–family conflict, we are reminded that increased decision-making does not imply that the additional decisions will be handled differently. Nonetheless, as was also suggested by Aguinis et al. (2009), future studies should investigate the reliability of moderation findings across additional samples of meta-analyses.

⁷ We would like to thank an anonymous reviewer for pointing out that many meta-analysts are promoting the idea of using specialized personnel to retrieve studies from a domain (e.g., a reference librarian) and that this could serve as a possible alternative to representative sampling or other approaches to handling prohibitively large research literatures.

Limitations

Adopting a case study approach allowed for an in-depth exploration of researcher decisions and associated consequences at each stage of the meta-analytic process. At the same time, having limited the focus to a detailed account of three specific meta-analyses suggests certain caveats, most notably, with respect to the issue of generalizability. While the meta-analyses reviewed here are representative of the meta-analytic process as it typically unfolds, we recognize that each new meta-analysis can present unique challenges (Dieckmann et al. 2009). Moreover, it is possible that meta-analyses from other domains could involve additional decisions or more difficult decisions, such that the degree of similarity observed here might have been due to the straightforward nature of these meta-analyses in particular. However, we feel that the decisions in these meta-analyses are both representative of common issues involved in a typical meta-analysis and underscore a fair degree of complexity. Nonetheless, the reader is cautioned against over generalizing from the results of this case study.

Further related to the generalizability of these meta-analyses, it may be that the small effects typically found in the telecommuting literature limited the impact of some researcher decisions. For example, whereas the decision to correct for criterion unreliability had little effect on estimates here, elsewhere this decision could yield substantial *gains*, particularly when effects are relatively large to begin with and reliability coefficients are low (Hunter and Schmidt 2004). Similarly, other scenarios may call for additional corrections due to range restriction on the independent variable, criterion variables, or both (e.g., selection-validation research). Therefore, we caution against interpreting our results as indicative of the general unimportance of statistical corrections in meta-analysis (see Geyskens et al. 2009).

Another potential limitation has to do with the possibility of researcher bias, particularly with respect to decisions that were made when comparing and reporting the present authors' own prior meta-analyses. However, several precautions were taken to limit the influence of any conscious or unconscious bias on the findings of our re-examination. First, the meta-analyses were completed independently by non-overlapping sets of researchers, and no contact occurred between the authors of the present manuscript prior to the completion of all three meta-analyses. Thus, our comparison was focused at the level of completed independent meta-analyses, and no results were changed as a function of subsequent correspondence between authors. Instead, correspondence was intended only to fill in any gaps in available information or to clarify

the nature of decisions that were made so that they could be reported in sufficient detail. Second, a methodological strength of our case study was the systematic framework used to guide comparisons. By adopting Wanous et al.'s (1989) 11-stage process model and making earnest efforts to report thorough and accurate information within each stage, we attempted to eliminate the possibility of selective reporting (e.g., the exclusion of unfavorable information about a particular meta-analysis).

Finally, it should be noted that NM's meta-analysis (unpublished) did not undergo the same level of rigorous peer review as GH and NI (published), and GH's is the only study published in a top outlet. Although some research suggests that the rigorousness of peer review is not related to the quality of meta-analytic practices (Geyskens et al. 2009), it is possible to question how potential revisions to NM's meta-analysis as a function of the journal review process may have affected the findings of our re-examination. In response to this question, it is interesting to speculate if including NM's meta-analysis may have skewed our test toward the possibility of observing dissimilarity between meta-analyses, based on the idea that rigorous peer review may act to standardize researcher decisions (e.g., by ensuring that meta-analytic guidelines are followed; Dieckmann et al. 2009; Schulze 2007).

Conclusions

In summary, the present case study sheds light on an important methodological issue that has received relatively little systematic investigation. Our re-examination of three telecommuting meta-analyses highlighted the manner in which researcher decisions affected the *direction and focus of meta-analyses*, while having relatively lesser impact on those meta-analytic findings that were directly comparable between meta-analyses. This was particularly true with respect to the overall effects reported by meta-analyses, which demonstrated a high degree of convergence. In line with Barrick et al.'s (2001) conclusions, we interpret this as additional positive evidence for meta-analytic methods. At the same time, our re-examination suggests that moderation tests may be more vulnerable to researcher idiosyncrasy, given that several additional decisions precede these analyses, and the potential influence of non-overlapping studies is magnified when categorical moderation tests involve a small subset of the total sample of studies. Finally, although further support was gained for the overall small and positive effects associated with telecommuting, consistent evidence of effect heterogeneity suggests the presence of additional moderators.

Acknowledgment We would like to thank Boris Baltes and Christopher Berry for their constructive comments on a previous draft of this manuscript.

References

- Aguinis, H., Dalton, D. A., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2009, August). *Meta-analytic choices and judgment calls: Implications for theory and scholarly impact*. Paper presented at the meeting of the Academy of Management, Chicago, IL.
- Allen, M., & Preiss, R. (1993). Replication and meta-analysis: A necessary connection. *Journal of Social Behavior and Personality*, 8, 9–20.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9–30.
- Beal, D. J., Corey, D. M., & Dunlap, W. P. (2002). On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in the meta-analysis of correlations. *Journal of Applied Psychology*, 87, 583–589.
- Beaman, A. L. (1991). An empirical comparison of meta-analytic and traditional reviews. *Personality and Social Psychology Bulletin*. *Special Issue: Meta-Analysis in Personality and Social Psychology*, 17, 252–257.
- Bobko, P., & Roth, P. L. (2008). Psychometric accuracy and (the continuing need for) quality thinking in meta-analysis. *Organizational Research Methods*, 11, 114–126.
- Briggs, D. C. (2005). Meta-analysis: A case study. *Evaluation Review*, 29, 87–127.
- Bullock, R. J., & Svyantek, D. J. (1985). Analyzing meta-analysis: Potential problems, an unsuccessful replication, and evaluation criteria. *Journal of Applied Psychology*, 70, 108–115.
- Burke, M. J., & Landis, R. S. (2003). Methodological and conceptual challenges in conducting and interpreting meta-analyses. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 287–310). Mahwah, NJ: Lawrence Erlbaum Associates.
- Campion, M. A. (1993). Article review checklist: A criterion checklist for reviewing research articles in applied psychology. *Personnel Psychology*, 46, 705–718.
- Cortina, J. M. (2002). Big things have small beginnings: An assortment of "minor" methodological misunderstandings. *Journal of Management*, 28, 339–362.
- Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods*, 6, 415–439.
- Cortina, J. M., & Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Journal of Applied Psychology*, 2, 161–172.
- Cree, L. H. (1999). Work/family balance of telecommuters. *Dissertation Abstracts International Section B: The Sciences and Engineering*, 59(11-B), 6100.
- Dieckmann, N. F., Malle, B. F., & Bodner, T. E. (2009). An empirical assessment of meta-analytic practice. *Review of General Psychology*, 13, 101–115.
- Eden, D. (2002). Replication, meta-analysis, scientific progress, and AMJ's publication policy. *Academy of Management Journal*, 45, 841–846.
- Egger, M., & Smith, G. D. (1998). Meta-analysis bias in location and selection of studies. *British Medical Journal*, 316, 61–66.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161–180.
- Gajendran, R. S., & Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: Meta-analysis of psychological mediators and individual consequences. *Journal of Applied Psychology*, 92, 1524–1541.
- Geyskens, I., Krishnan, R., Steenkamp, J. B. E. M., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management*, 35, 393–419.
- Greenhouse, J. B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 503–520). New York: Russell Sage Foundation.
- Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *British Medical Journal*, 331, 1064–1065.
- Hackett, R. D., & Guion, R. M. (1985). A re-evaluation of the absenteeism-job satisfaction relationship. *Organizational Behavior and Human Decision Processes*, 35, 340–381.
- Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology*, 60, 159–170.
- Hale, J., & Dillard, J. (1991). The uses of meta-analysis: Making knowledge claims and setting research agendas. *Communication Monographs*, 58, 463–471.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8, 275–292.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- Kisamore, J., & Brannick, M. (2008). An illustration of the consequences of meta-analysis model choice. *Organizational Research Methods*, 11, 35–53.
- Konstantopoulos, S., & Hedges, L. V. (2009). Analyzing effect sizes: Fixed-effects models. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 279–293). New York: Russell Sage Foundation.
- Kromrey, J. D., & Rendina-Gobioff, G. (2006). On knowing what we do not know: An empirical comparison of methods to detect publication bias in meta-analysis. *Educational and Psychological Measurement*, 66, 357–373.
- Miner, J. B., & Raju, N. S. (2004). Risk propensity differences between managers and entrepreneurs and between low- and high-growth entrepreneurs: A reply in a more conservative vein. *Journal of Applied Psychology*, 89, 3–13.
- Nicklin, J. M., Mayfield, C. O., Caputo, P. M., Arboleda, M. A., Cosentino, R. E., Lee, M., et al. (2009). Does telecommuting increase organizational attitudes and outcomes: A meta-analysis. *Pravara Management Review*, 8, 2–16.
- Nieminen, L. R. G., Chakrabarti, M., McClure, T. K., & Baltes, B. B. (2008). *A meta-analysis of the effects of telecommuting on employee outcomes*. Paper presented at the 23rd Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Ones, D. S., Mount, M. K., Barrick, M. R., & Hunter, J. E. (1994). Personality and job performance: A critique of the Tett, Jackson, and Rothstein (1991) meta-analysis. *Personnel Psychology*, 47, 147–156.
- Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3, 354–379.
- Oyer, E. J. (1997). Validity and impact of meta-analyses in early intervention research. *Dissertation Abstracts International Section A: Humanities and Social Sciences*. 57(7-A), 2859.

- Raghuram, S., & Weisenfeld, B. (2004). Work-nonwork conflict and job stress among virtual workers. *Human Resource Management, 43*, 259–277.
- Rothstein, H. R., & McDaniel, M. A. (1989). Guidelines for conducting and reporting meta-analyses. *Psychological Reports, 65*, 759–770.
- Rothstein, H., Sutton, A. J., & Bornstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. Chichester, UK: Wiley.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.
- Schmidt, F. L., & Hunter, J. E. (1999). Comparison of three meta-analysis methods revisited: An analysis of Johnson, Mullen, and Salas (1995). *Journal of Applied Psychology, 84*, 144–148.
- Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge: Hogrefe & Huber.
- Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Journal of Psychology. Special Issue: The State and the Art of Meta-Analysis, 215*(2), 90–103.
- Scott, K. D., & Taylor, D. S. (1985). An examination of conflicting findings on the relationship between job satisfaction and absenteeism: A meta-analysis. *Academy of Management Journal, 28*, 599–612.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- Staples, D. S. (2001). A study of remote workers and their differences from non-remote workers. *Journal of End User Computing, 13*, 3–14.
- Stewart, W. H., Jr., & Roth, P. L. (2001). Risk propensity differences between entrepreneurs and managers: A meta-analytic review. *Journal of Applied Psychology, 86*, 145–153.
- Stewart, W. H., Jr., & Roth, P. L. (2004). Data quality affects meta-analytic conclusions: A response to Miner and Raju (2004) concerning entrepreneurial risk propensity. *Journal of Applied Psychology, 89*, 14–21.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703–742.
- Wanous, J. P., Sullivan, S. E., & Malinak, J. (1989). The role of judgment calls in meta-analysis. *Journal of Applied Psychology, 74*, 259–264.
- Wells, K., & Littell, J. H. (2009). Study quality assessment in systematic reviews of research on intervention effects. *Research on Social Work Practice, 19*, 52–62.