Article

Organizational Research Methods
1-16

DOI: 10.1177/1094428119836485
journals.sagepub.com/home/orm
SAGE

Best Practices in Data Collection and Preparation: Recommendations for Reviewers, Editors, and Authors

Herman Aguinis[1], **N. Sharon Hill**[1], **and James R. Bailey**[1]

Abstract

We offer best-practice recommendations for journal reviewers, editors, and authors regarding data collection and preparation. Our recommendations are applicable to research adopting different epistemological and ontological perspectives—including both quantitative and qualitative approaches—as well as research addressing micro (i.e., individuals, teams) and macro (i.e., organizations, industries) levels of analysis. Our recommendations regarding data collection address (a) type of research design, (b) control variables, (c) sampling procedures, and (d) missing data management. Our recommendations regarding data preparation address (e) outlier management, (f) use of corrections for statistical and methodological artifacts, and (g) data transformations. Our recommendations address best practices as well as transparency issues. The formal implementation of our recommendations in the manuscript review process will likely motivate authors to increase transparency because failure to disclose necessary information may lead to a manuscript rejection decision. Also, reviewers can use our recommendations for developmental purposes to highlight which particular issues should be improved in a revised version of a manuscript and in future research. Taken together, the implementation of our recommendations in the form of checklists can help address current challenges regarding results and inferential reproducibility as well as enhance the credibility, trustworthiness, and usefulness of the scholarly knowledge that is produced.

Keywords

quantitative research, qualitative research, research design

[1]Department of Management, School of Business, The George Washington University, Washington, DC, USA

Corresponding Author:
Herman Aguinis, Department of Management, School of Business, The George Washington University, 2201 G Street, NW, Washington, DC 20052, USA.
Email: haguinis@gwu.edu

We offer best-practice recommendations for journal reviewers, editors, and authors regarding data collection and preparation. Our article has the dual purpose of offering prescriptive information about (a) methodological best practices and (b) how to enhance transparency. We focus on data collection and preparation because these are foundational steps in all empirical research that precede data analysis, production of results, and drawing conclusions and implications for theory and practice.

Specifically regarding transparency, many published articles in management and related fields do not include sufficient information on precise steps, decisions, and judgment calls made during a scientific study (Aguinis, Ramani, & Alabduljader, 2018; Aguinis & Solarino, 2019; Appelbaum et al., 2018; Levitt et al., 2018). One of the most detrimental consequences of insufficient methodological transparency is that readers are unable to reproduce research (Bergh, Sharp, Aguinis, & Li, 2017). That is, insufficient transparency leads to lack of results reproducibility and lack of inferential reproducibility. Results reproducibility is the ability of others to obtain the same results using the same data as in the original study, and it is an important and evident requirement for science (Bettis, Ethiraj, Gambardella, Helfat, & Mitchell, 2016). In addition, inferential reproducibility is the ability of others to draw similar conclusions to those reached by the original authors. Absent sufficient inferential reproducibility, it is impossible for a healthily skeptical scientific readership to evaluate conclusions regarding the presence and strength of relations between variables (Banks et al., 2016; Grand, Rogelberg, Banks, Landis, & Tonidandel, 2018; Tsui, 2013). Also, without sufficient methodological transparency, reviewers are unable to fully assess the extent to which the study adheres to relevant methodological best practices. Moreover, insufficient methodological transparency is a detriment to practice as well. Namely, untrustworthy methodology is an insuperable barrier to using the findings and conclusions to drive policy changes or inform good managerial practices.

The Present Article

We offer recommendations, which we summarize in the form of checklists, that reviewers and editors can use as a guide to critical issues when evaluating data collection and preparation practices in submitted manuscripts.[1] Our recommendations are sufficiently broad to be applicable to research adopting different epistemological and ontological perspectives—including both quantitative and qualitative approaches—and across micro and macro levels of analyses.

Aguinis et al. (2018) proposed a conceptual framework to understand insufficient methodological transparency as a "research performance problem." Specifically, they relied on the performance management literature showing that performance problems result from insufficient (a) knowledge, skills, and abilities (KSAs) and (b) motivation (Aguinis, 2019; Van Iddekinge, Aguinis, Mackey, & DeOrtentiis, 2018). So, if authors are disclosing insufficient details about data collection and preparation procedures, this research performance problem could be explained by researchers' lack of KSAs (i.e., know-how) and lack of motivation (i.e., want) to be transparent. Our article addresses both.

One outcome of using the checklists during the review process could be outright rejection of the submission. But the checklists can further help reviewers express to the authors the serious consequences of not addressing the uncovered issues. This first purpose addresses motivational aspects because authors are more likely to be transparent if they know that failure to disclose necessary information may lead to a manuscript rejection decision. In addition, the checklists have developmental purposes. In other words, the outcome of the review process may be revision and resubmission with some of the reviewers' comments dedicated to making recommendations about, for example, what needs to be improved, what needs to be more transparent, and why these issues are important. Clearly, some of the issues could be addressed in a revision such as performing a different or no data transformation (as we describe later in our article). But others may not be fixable because they involve decisions that need to be made prior to data collection (e.g., research design). Nevertheless, the checklists can still be helpful for reviewers to provide advice to authors regarding their future research. So, this second use of our checklists addresses authors' KSAs.

We address the following four issues regarding data collection: (a) type of research design, (b) control variables, (c) sampling procedures, and (d) missing data management. In addition, we address the following three issues regarding data preparation: (e) outlier management, (f) use of corrections for statistical and methodological artifacts, and (g) data transformations.[2] Next, we offer a description of each of the aforementioned seven issues together with examples of published articles that are exemplary in the steps they took as well as transparent regarding each of the issues we describe. The topics we describe are broad and not specific to any particular field, theoretical orientation, or domain and include exemplars from the micro as well as the macro literature. Also, in describing each, we refer to specific methodological sources on which we relied to offer our best-practice recommendations.

Data Collection

The data collection stage of empirical research involves several choices such as the particular type of research design, what sampling procedures are implemented, whether to use control variables and which ones in particular, and how to manage missing data. As a preview of our discussion and recommendations regarding each of these issues, Table 1 includes a checklist and summary of recommendations together with exemplars of articles that implemented best practices that are also highly transparent regarding each of these issues. Next, we address these four data-collection issues in detail.

Type of Research Design

The ultimate purpose of all scientific endeavors is to develop and test theory, and a critical goal is to address causal relations: Does X cause Y? To establish causal claims, the cause must precede the effect in time (Shadish, Cook, & Campbell, 2002; Stone-Romero, 2011). In other words, the research design must be such that data collection involves a temporal precedence of X relative to Y (Aguinis & Edwards, 2014). Another necessary condition for drawing conclusions about causal relations is the ability to rule out alternative explanations for the presumed causal effect (Shadish et al., 2002).

Because information regarding research design issues is critical for making claims about causal relations between variables, submitted manuscripts need to answer fundamental questions such as: Which data were collected and when? Was a control group used? Were the data collected at different levels of analysis? Was the design more suitable for theory development or theory testing? Was the design experimental or quasi-experimental? Was the design inductive, deductive, or abductive?

For example, in their study on the effects of team reflexivity on psychological well-being, Chen, Bamberger, Song, and Vashdi (2018) provided the following information regarding their research design:

> We implemented a time lagged, quasi-field experiment, with half of the teams trained in and executing an end-of-shift team debriefing, and the other half assigned to a control condition and undergoing periodic postshift team-building exercises . . . . Prior to assigning production teams to experimental conditions (i.e., at T0), we collected data on the three team-level burnout parameters and team-level demands, control, and support. We then assigned 36 teams to the intervention condition and the remaining teams to the control condition on the basis of the shift worked (i.e., day vs. night). (pp. 443-444)

Table 1. Checklist for Data Collection Practices.

Methodological Issue	Questions That Need to Be Addressed in Submitted Manuscripts	Exemplars of Best Practices	Methodological Sources Describing Practices
Type of research design	✓ Which data were collected and when? ✓ Was a control group used? ✓ Were the data collected at different levels of analysis? ✓ Was the design more suitable for theory development or theory testing? ✓ Was the design experimental or quasi-experimental? ✓ Was the design inductive, deductive, or abductive?	• Chen, Bamberger, Song, and Vashdi (2018) • Truelove and Kellogg (2016)	• Shadish, Cook, and Campbell (2002) • Stone-Romero (2011)
Control variables	✓ To what extent do the control variables chosen hold theoretically meaningful relationships with antecedents and outcomes? ✓ How and why were the particular control variables chosen? ✓ Which control variables were initially investigated but then subsequently dropped from the final analysis? ✓ Are the control variables included in the correlation table so that their reliability and correlations with all other variables are reported openly and fully?	• Chiu, Balkundi, and Weinberg (2017) • Ryu, McCann, and Reuer (2018)	• Becker et al. (2016) • Bernerth and Aguinis (2016)
Sampling procedures	✓ Which particular firms and/or individuals were targeted and which eventually included in the study? ✓ What were the procedures used in selecting and recruiting them? ✓ In the case of archival data, what specific databases were used? ✓ What particular type of sampling procedure was used (e.g., snowballing, convenience, purposeful)?	• Kaul, Nary, and Singh (2013) • Follmer, Talbot, Kristof-Brown, Astrove, and Billsberry (2018)	• Aguinis and Lawal (2012) • Teddlie and Yu (2007)
Missing data management	✓ If data imputation approaches were used, what specific techniques were implemented and using which particular software package? ✓ What are the assumptions in the implementation of these procedures (e.g., data missing at random)? ✓ Were some of the initially collected data excluded from analysis, and if yes, why? ✓ What was the approach to eliminating firms/individuals from the study based on missing data concerns (e.g., listwise, pairwise)?	• Kalaignanam, Kushwaha, Steenkamp, and Tuli (2013) • Antonakis, House, and Simonton (2017)	• Newman (2014) • Schafer and Graham (2002)

Note: The Methodological Sources column includes sources reviewers, editors, and authors can consult for more detailed information about each of the issues.

Chen et al. described that the independent variables were collected before the dependent variables and that their design allowed for conclusions about causality (including a control group). The implementation of these procedures allowed them to make claims about causal relations.

As a second example, in a study investigating an early entrant into the U.S. car-sharing industry and factors that facilitate cross-occupational collaboration, Truelove and Kellogg (2016) provided information about their qualitative research design. Specifically, they noted that they used

> an inductive, ethnographic approach that is well-suited for developing new theory .... Data collection at Transco's headquarters began in June 2012 and ended a year later .... We observed 111 meetings: 87 daily "scrum" meetings and 24 cross-departmental meetings in which engineers and marketers discussed progress and challenges on joint projects. This enabled us to track 42 projects attempted during the study's time-frame .... To complement observational data, we conducted 126 formal, semistructured interviews. (p. 667)

In other words, they were clear about their use of an inductive design, which allowed them to accomplish their theory-building goal (as opposed to theory testing).

Control Variables

The topic of control variables is directly related to our previous discussion about drawing causal inferences because it is about ruling out alternative explanations for a hypothesized effect (Aguinis & Vandenberg, 2014; Becker et al., 2016; Bernerth & Aguinis, 2016). When it is not practically possible to implement a research design that allows researchers to prevent or rule out the influence of other variables, the use of statistical controls offers an alternative for addressing potential confounds. Stated differently, the use of statistical controls is a way to improve on limitations of the data collection process because it involves measuring variables suspected of having a relationship with either an antecedent or outcome and including them in the analyses, thereby removing potential antecedent-outcome "contamination" (Bernerth & Aguinis, 2016; Spector & Brannick, 2011).

Important questions related to control variables that reviewers should be able to answer include: To what extent do the control variables chosen in a particular study hold a theoretically meaningful relationships with antecedents and outcomes? How and why were the particular control variables chosen? Which control variables were initially investigated but then subsequently dropped from the final analysis? Are the control variables included in the correlation table so that their reliability and correlations with all other variables are reported openly and fully?

For example, in a study examining the processes through which managers become leaders, Chiu, Balkundi, and Weinberg (2017) implemented best practices in reporting that they

> included several control variables at both the leader and follower levels (L2 and L1). In order to better capture the magnitude of leader network centralities on leadership perception at T2, we controlled for followers' leadership perceptions at the prior data collection wave (L1.T1). Also, we controlled for the sex of both leaders and followers because multiple studies have reported the impact of sex on leadership perception and effectiveness .... Further, we controlled for leaders' personality, because multiple meta-analyses have indicated associations between leadership perception and leader personality. (p. 339)

Their statistical controls were described clearly and precisely, and each had a theory-based justification.

As a second illustration of best practices, Ryu, McCann, and Reuer (2018) examined the extent to which geographic co-location between a focal firm's partner and rivals introduces potential indirect paths of knowledge leakage to rivals. In their description, they explained that they

> controlled for a number of additional factors that the previous literature has argued affect knowledge misappropriation and spillover concerns and therefore could affect alliance governance and design. The alliance literature has long argued that social networks in which alliance partners are embedded provide controls for opportunistic behaviors and thus might also affect the risk of knowledge losses, as well as the alliance design choices firms make . . . we controlled for an alliance dyad's social embeddedness, using variables to capture the partners' prior ties, indirect ties between the two firms in the dyad, and each partner's degree centrality. (p. 953)

Similar to the previous example, Ryu et al. offered clear and theory-based justification for the inclusion of controls.

Sampling Procedures

Organizational research relies on samples rather than populations. Thus, it is important to provide information on the procedures used to select the sample because this is critical for making inferences about the representativeness of the observations and the generalizability of results and conclusions (Aguinis & Lawal, 2012). Moreover, in the case of qualitative research, transparency regarding sampling procedures is particularly pertinent because samples are often nonprobabilistic (Aguinis & Solarino, 2019; Teddlie & Yu, 2007).

Reviewers evaluating manuscripts should be able to answer the following questions about sampling procedures: Which particular firms and/or individuals were targeted, and which were eventually included in the study? What were the procedures used in selecting and recruiting them? In the case of archival data, what specific databases were used? What particular type of sampling procedure was used (e.g., snowballing, convenience, purposeful)?

For example, Kaul, Nary, and Singh (2018) examined the role of nonventure private equity firms in the market for divested businesses. They described their sampling strategy and steps as follows:

> We test our hypotheses in a sample of divestments by publicly listed U.S. manufacturing firms from 1997 to 2010. Data on divestments are drawn from the Securities Data Company (SDC) Platinum database. We begin with all divestments in the U.S. manufacturing sector (SIC 2000-3999) listed in the SDC database for our sample years. These are then matched to parent corporations in Compustat, thus limiting our sample to divestments by publicly listed firms. We then match the remaining transactions to the Execucomp database, since we need data on executive compensation and ownership. Our final sample consists of 1,711 divestments. (p. 1277)

So, sampling procedures were plainly described so that a discerning scientific readership can understand the population to which results can be generalized.

As a second example regarding best practices about sampling procedures, Follmer, Talbot, Kristof-Brown, Astrove, and Billsberry (2018) wrote that they

> used responses from Phase 1 to design the sampling approach for Phase 2 by evaluating when episodes of misfit were most likely to occur and who was most likely to try to address misfit, rather than quickly leave the organization . . . we pursued a theoretically driven sampling

> strategy for Phase 2 . . . . Although leaving is one viable response to misfit, our particular interest was in people who chose to remain in poorly fitting environments. Therefore, we approached people who had recently experienced a change at work and those who had invested in their careers through obtaining a higher-level degree. To gather this sample, we recruited from the LinkedIn and Facebook groups for the alumni associations of a college of business in a large, Midwestern, state university and a small, East Coast, liberal arts college. Respondents qualified for the study if they were currently employed at least part-time and were experiencing misfit due to a recent change at work. (pp. 444-445)

Follmer et al. described their sampling procedures in detail and the rationale that guided these choices.

Missing Data Management

Directly related to the aforementioned discussion regarding sampling procedures, it is likely that the resulting data set includes missing data (Newman, 2014; Schafer & Graham, 2002). This is an issue not only in the context of micro research involving individual-level data collected using surveys (Aguinis, Pierce, Bosco, & Muslin, 2009) but also macro research involving firm-level data available in archival sources and publicly or commercially available databases (Ketchen, Ireland, & Baker, 2013).

There are several questions related to missing data management that submitted manuscripts should address. For example, if data imputation approaches were used, what specific techniques were implemented, and which particular software package was used? What are the assumptions in the implementation of these procedures (e.g., data missing at random)? Were some of the initially collected data excluded from analysis, and if yes, why? What was the approach to eliminating firms/individuals from the study based on missing data concerns (e.g., listwise, pairwise)?

As an example of best practices, Kalaignanam, Kushwaha, Steenkamp, and Tuli (2013) examined for which types of firms and contexts it is more or less beneficial to outsource customer-relationship management. Regarding missing data management, they reported that they

> were able to collect data on IT expenses for 119 of the 158 observations in our sample. For the remaining 39 cases, where IT expenses are missing in a particular year, we used the following heuristic. We searched the 10-K statements and annual reports for the firm in previous and subsequent financial years and replaced the missing value with IT expense from the previous available year. If IT expenses of a firm were not available for any financial year (11 of 39 cases), we replaced the missing value with the average IT expense in the industry (i.e., four-digit or three-digit SIC code) for the given year. (p. 757)

So, Kalaignanam et al. provided precise information on how they handled missing data—and the justification for their procedures.

As a second illustration regarding best practices about missing data management, Antonakis, House, and Simonton (2017) studied whether the relation between intelligence and perceived leadership might be more accurately described by a curvilinear rather than a linear function. Regarding how they managed missing data, they explained the following:

> Because data were missing only on one variable (IQ), we could not directly test the assumption of missing completely at random (MCAR) . . . . Thus, using the full sample we created a variable "missing" coded 0 for when data is complete, or 1 otherwise and regressed the seven fully measured leader individual-difference characteristics (i.e., personality, gender and age)

on the variable "missing" and all the fixed effects . . . the variable "missing" was unrelated both individually and jointly to any of the individual difference measures . . . . As a further test, we examined the MCAR assumption . . . by performing two Monte Carlo simulations . . . . The mean *p* value of the MCAR test was .86 (SE .003; 95% CI [.85 to .87]). Out of the 5,000 simulations, the test was only significant 18 times (0.36%). Thus, overall, the listwise sample with full observations appears to be MCAR. Still, in reporting, we include too results from the full sample using Stata's MLMV estimator—maximum likelihood estimator for missing data in the event that data are MAR or missing at random (which is not testable); the MLMV estimator is still consistent under MAR assumptions. (pp. 1008-1009)

So, Antonakis et al. offered a detailed description of how they assessed the type of missing data pattern and how, based on their assessment, they chose a particular missing data management procedure.

Data Preparation

The data preparation stage takes place after the data are collected and prior to data analysis. Several choices and judgment calls characterize this stage, such as how to handle outliers; whether to use corrections for statistical and methodological artifacts, and which in particular; and whether to transform the data that have been collected. As a preview of our discussion and recommendations regarding each of these issues, Table 2 includes a checklist and summary of recommendations together with exemplars of articles that were highly transparent regarding each of these issues. As in the case of Table 1, Table 2 also includes methodological sources on which we relied to offer our best-practice recommendations. Next, we address these three data preparation issues in detail.

Outlier Management

Outliers are data points that lie far from others (Aguinis, Gottfredson, & Joo, 2013). Because their location is so far from other data points, they often have an outsized influence on parameter estimates as well as standard errors, which are used for hypothesis testing and computing *p* values as well as confidence intervals. Thus, different ways of managing outliers usually alter substantive research conclusions (Aguinis & Joo, 2015).

Submitted manuscripts should address the following questions: What were the specific rules used for defining, identifying, and handling outliers? Were outlying data points error outliers, influential outliers, or interesting outliers? How many outlying data points were deleted from the final sample? What would the results be if the deleted observations had been included? Answering these questions is critical for drawing substantive conclusions and also for future reproducibility attempts.

As an example of best practices regarding outlier management, consider a study by Dineen, Duffy, Henle, and Kiyoung (2017) on how a painful social comparative emotion (i.e., job search envy) transmutes as deviant or normative résumé fraud. In describing how they managed outliers, they reported that they

screened for outliers using Bollen and Jackman's (1990) conservative criteria for the standardized dFits diagnostic statistic. This statistic offers a balance between identifying studentized residuals and influential cases. We also examined the raw data to ensure that the dependent variable for any identified outlying cases was at least four standard deviations from the mean. Using this combined procedure we eliminated one outlying case, for a final sample size of 334. (p. 302)

Table 2. Checklist for Data Preparation Practices.

Methodological Issue	Questions That Need to Be Addressed in Submitted Manuscripts	Exemplar of Best Practices	Methodological Sources Describing Best Practices
Outlier management	✓ What were the specific rules used for defining, identifying, and handling outliers? ✓ Were outlying data points error outliers, influential outliers, or interesting outliers? ✓ How many outlying data points were deleted from the final sample? ✓ What would the results be if the deleted observations had been included?	• Dineen, Duffy, Henle, and Kiyoung (2017) • Neckebrouck, Schulze, and Zellweger (2018)	• Aguinis, Gottfredson, and Joo (2013) • Aguinis and Joo (2015)
Use of corrections for statistical and methodological artifacts	✓ For all corrections, what were the procedures used (e.g., artifact distributions?), and what were the assumptions involved? ✓ Was the measurement error correction applied to the antecedent variable, outcome variable, or both? ✓ If a range restriction correction was used, was it for direct or indirect range restriction, or both?	• Huang, Ryan, Zabel, and Palmer (2014) • Shaffer and Postlethwaite (2012)	• Le, Schmidt, and Putka (2009) • Schmidt and Hunter (2014)
Data transformations	✓ What were the effects of the transformation? ✓ Were hypotheses stated in terms of untransformed scores actually tested with transformed data? ✓ Which particular variables were transformed, what was the specific procedure used, and what is the justification for the particular procedure?	• Courtright, Gardner, Smith, Mccormick, and Colbert (2016) • McDonnell and Brayden (2013)	• Becker, Robertson, and Vandenberg (2019) • Cohen, Cohen, West, and Aiken (2003)

Note: The Methodological Sources column includes sources reviewers, editors, and authors can consult for more detailed information about each of the issues.

So, Dineen et al. were fully transparent about the implementation of their particular outlier management procedure and what guided their choice.

As a second example, Neckebrouck, Schulze, and Zellweger (2018) tested competing propositions about the impact of family on employment practices and reported, “After winsorizing our data at the 1% level to reduce the potential influence of outliers, our final data set consisted of 102,094 firm-years of data from 14,961 private firms” (p. 560). In other words, Neckebrouck et al. explained that they windsorized their data and precisely reported the final size of their sample.

Use of Corrections for Statistical and Methodological Artifacts

The development of valid and reliable measures continues to be an important challenge in organizational research (Aguinis, Henle, & Ostroff, 2001; Bliese, 2018). We are still a long way from the type of effective and precise measures that are in use in other scientific fields, such as the electron microscope in biology, the Hubble telescope in astronomy, and the teleseismometer in geology. A clear sign of the need to improve measurement is that reliability estimates of about .80 or higher are usually considered adequate—meaning that 20% of variance in observed scores is completely random. This level of error is completely unacceptable in older and more established fields such as physics as well as applied fields such as engineering. For example, this level of measurement error would surely have prevented the successful landing of the Rosetta orbiter on a distant comet after chasing it in its orbit around the sun at a speed of about 19,000 miles per hour for 12 years (Gilbert, 2016).

Acknowledging the fallibility of observed scores, organizational researchers often implement a variety of procedures to “correct” measures for what are often labeled methodological and statistical artifacts (Le, Schmidt, & Putka, 2009). The goal of these corrections is to learn what scores and relations between scores would look like if measures were error free (i.e., no sampling error, no measurement error, no range restriction; Schmidt & Hunter, 2014). But, there are several questions that researchers need to answer in their submitted manuscripts. For example, for all corrections, what were the procedures used (e.g., artifact distributions), and what were the assumptions involved? Was the measurement error correction applied to the antecedent variable, outcome variable, or both? If a range restriction correction was used, was it for direct or indirect range restriction, or both?

As an example of best practices regarding this issue, Huang, Ryan, Zabel, and Palmer (2014) implemented several corrections in their meta-analysis of the relation between personality and adaptive performance at work. They reported the following:

> We corrected for measurement error in the criterion using internal consistency estimates for the overall adaptive performance measure at the study level . . . we corrected for direct or indirect range restriction at the study level using information from the HPI manual . . . . For the six samples with unknown validation design, we adopted the direct range restriction formula on these datasets to provide a conservative estimate . . . . As no interrater reliability estimates were available, the corrected validity estimates in the current study would be the lower bound estimates for operational validity. (p. 6)

So, Huang et al. provided precise information on which corrections they used and why.

Another example of best practices regarding corrections is from Shaffer and Postlethwaite (2012) in their meta-analysis to compare the validities of general noncontextualized personality measures and work-specific contextualized measures. They reported the following:

> We analyzed our data using the methods developed by Hunter and Schmidt (2004). This method computes the sample-weighted mean of the observed correlations and observed

> standard deviations from the original studies and then corrects them for statistical artifacts, including predictor unreliability, criterion unreliability, and range restriction.... We computed separate artifact distributions for each of the Big Five traits based on the data available in our data set. Because it is extremely unlikely that the individuals in our sample were selected top-down based on personality scores, corrections for direct range restriction were not appropriate for use in this study. Therefore, in order to correct for range restriction we used the procedures for correcting for indirect range restriction outlined by Hunter and Schmidt (2004). We computed a separate u_x value for each Big Five trait by combining data from each of the studies in our data set with the normative data provided in the test manuals for the various personality scales that were included in our data set. We also computed estimates of predictor reliability.... Although we did not correct for unreliability in the predictor, predictor reliabilities were used to test Hypothesis 3 and in the process of correcting for indirect range restriction. (p. 12)

So, Shaffer and Postlethwaite specified the procedures they used for each of the corrections and their rationale.

Data Transformations

There is increased awareness that in many research domains, variables are not normally distributed; that is, they do not follow a Gaussian pattern in which scores are clustered around the mean, the mean and median are in the same location, and there are fairly small tails to the left and to the right of the mean (Aguinis & O'Boyle, 2014). For example, recent research in the domain of individual performance has provided evidence regarding the prevalence of heavy-tailed distributions due to the presence of star performers (e.g., Aguinis, Ji, & Joo, 2018; Joo, Aguinis, & Bradley, 2017). Similarly, firm performance and firm size (Stanley et al., 1995) as well as firm revenue and revenue growth (Crawford, Aguinis, Lichtenstein, Davidsson, & McKelvey, 2015) also follow non-normal distributions

A common practice is to implement nonlinear transformations to normalize scores (Becker, Robertson, & Vandenberg, 2019). As documented by Becker et al. (2019), researchers also transform their data to reduce heteroscedasticity, linearize relations among variables, change relations between variables from multiplicative to additive, and promote analytical convenience (i.e., to be able to use familiar data-analytic techniques based on the general lineal model; Cohen, Cohen, West, & Aiken, 2003). However, based on their review of 323 articles published between 2012 and 2017, Becker et al. showed that many articles do not include sufficient description of the transformation procedures, their justification, and their effects. So, there is a need to answer the following questions in submitted manuscripts: What were the effects of the transformation? Were hypotheses stated in terms of untransformed scores actually tested with transformed data? Which particular variables were transformed, what was the specific procedure used, and what is the justification for the particular procedure?

As an example of best practices regarding data transformations, consider a study by Courtright, Gardner, Smith, Mccormick, and Colbert (2016) on the relation between ego depletion and abusive supervision. Specifically, they noted:

> Given the low mean level of abusive supervision in our sample, we conducted an analysis with a data transformation appropriate for skewed data by squaring the abusive supervision variable (Tabachnick & Fidell, 1996). In doing so, we found results that mirrored those of our original analysis, with support for tested hypotheses remaining statistically significant. Thus, for the

> sake of interpreting our results in a parsimonious fashion, we report the results of the standard linear regression analyses. (p. 1638)

In other words, Courtright et al. explained the rationale for the transformation and that because results were similar compared to the untransformed data, they focused on the original scores.

A macro-level illustration of the procedure used and rationale for transformation is McDonnell and Brayden's (2013) study examining how firms' strategic reactions differ based on their reputation when they are targeted by consumer boycotts. The authors provided a detailed description of how they transformed the Fortune reputation ranking score as follows:

> Each organization included in the list is given a raw score that ranges from 0 to 10. Scores for roughly 37 percent of the firms in our sample are not reported in the Fortune rankings, however, an indicator that these firms are not central players in the field. The raw scores are also unevenly distributed . . . most scores vary between 6 and 8. Following prior research . . . we adopted an ordinal transformation of the raw Fortune scores to account for the skewed distribution of the scores and to demonstrate a firm's relative reputational position . . . . To create the ordinal transformation of the reputation variable, we used Stata's xtile function to evenly divide the raw scores into three quantiles, or tiers, of ranked firms. We recalculated the tiers for each year, using the scores of every firm in the reputation index. Because we recalculated quantile membership for each year, raw scores vary in their distribution among the quantiles, depending on the shape of the distribution of the raw scores that year . . . . A score of "1" was given to companies in the lowest third of Fortune's annual index in a given year; companies in the middle tier of the rankings in their year received a value of "2," and the highest value, "3," was allotted to all companies in the top third. We assigned unranked firms a value of zero to reflect their peripheral field position . . . . By opting for an ordinal scale, we acknowledge that variation in the rankings is important, but we also operate under the assumption that firms that are not covered in Fortune's index lack the same reputational standing as those that are covered. (p. 398)

McDonnell and Brayden described their transformation procedure in detail, its rationale, as well as assumptions.

Discussion

Organizational research is currently facing a credibility challenge due to the documented inability to reproduce published research (Bergh et al., 2017; Butler, Delaney, & Spoelstra, 2017; Byington & Felps, 2017). But, this is not unique to organizational research, and the same challenge is currently faced by economics (Camerer et al., 2016) and the natural sciences as well (Baker, 2016). These challenges will not simply go away, and as noted by Honig et al. (2018), "today's challenge to the integrity of management scholarship does not come from external demands for ideological conformity, rather from escalating competition for publication space in leading journals that is changing the internal dynamics of our community" (p. 413). Thus, we applaud the initiative by CARMA and *Organizational Research Methods* (*ORM*) to join forces to tackle these issues by publishing a series of articles offering useful resources for reviewers and editors that can also be used by authors. Aguinis, Ramani, and Villamor (2019) predicted that "*ORM* will play an increasingly important role regarding questionable research practices," and we are delighted that this is happening so quickly.

The dual goal of our article is to offer prescriptive information about (a) methodological best practices and (b) how to enhance transparency. In particular, our focus is on the early stages of empirical research—the data collection and preparation phases—because they are foundational and influence the rest of the process ranging from data analysis, to results, and to implications for theory and practice. Also, methodological transparency is key for results and inferential reproducibility. Thus, improving the transparency of published research is a very good step toward addressing the current credibility challenge.

A summary of our recommendations and checklists are included in Tables 1 and 2. But, we offer the following caveats. First, these checklists should not be applied rigidly to all empirical research. The reason is that although they are broad in nature, not all of them apply to absolutely every situation and empirical study. Overall, our view is that the more items that can be checked off the list, the better. But this does not mean that the absence of any particular item has veto power over a manuscript's publication deservingness. This is a judgment call that action editors will have to weigh. Second, as mentioned earlier, these recommendations and checklists can be used for evaluative and also developmental purposes. In terms of their developmental purpose, reviewers and editors can use them to offer advice to authors on what additional information to include in a revised manuscript or in future research efforts (thereby enhancing authors' KSAs), and instructors can use them for teaching doctoral seminars and methods workshops.

In closing, as noted by Aguinis et al. (2019), "*ORM* can be an important source of knowledge regarding state-of-the-science approaches to 'gray areas' in the application of specific methodologies and data analytical approaches and improving the accuracy and transparency of organizational research." We hope our recommendations and checklists will make a useful contribution toward these efforts.

Acknowledgment

We thank Larry J. Williams, Robert J. Vandenberg, George Banks, and two *Organizational Research Methods* reviewers for comments on previous drafts. A previous version of our manuscript was presented as part of the CARMA Advanced Reviewer Development series on September 21, 2018.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Herman Aguinis https://orcid.org/0000-0002-3485-9484

Notes

1. Our recommendations can also be used by reviewers involved in evaluating preregistered reports, which are manuscripts describing a study's method and proposed analyses but not the results.
2. We readily acknowledge that many of our recommendations require that authors provide additional information, which is likely to lengthen manuscripts. Some of this information can be included in the manuscript proper and some in online supplements (as is already the practice in many journals).

References

Aguinis, H. (2019). *Performance management* (4th ed.). Chicago, IL: Chicago Business Press.

Aguinis, H., & Edwards, J. R. (2014). Methodological wishes for the next decade and how to make wishes come true. *Journal of Management Studies*, *51*, 143-174.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*, 270-301.

Aguinis, H., Henle, C. A., & Ostroff, C. (2001). Measurement in work and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 27-50). London: Sage.

Aguinis, H., Ji, Y. H., & Joo, H. (2018). Gender productivity gap among star performers in STEM and other scientific fields. *Journal of Applied Psychology*, *103*, 1283-1306.

Aguinis, H., & Joo, H. (2015). Debunking myths and urban legends about how to identify influential outliers. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 206-223). New York, NY: Routledge.

Aguinis, H., & Lawal, S. O. (2012). Conducting field experiments using eLancing's natural environment. *Journal of Business Venturing*, *27*, 493-505.

Aguinis, H., & O'Boyle, E. (2014). Star performers in twenty-first-century organizations. *Personnel Psychology*, *67*, 313-350.

Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. (2009). First decade of *Organizational Research Methods*: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods*, *12*, 69-112.

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals*, *12*, 83-110.

Aguinis, H., Ramani, R. S., & Villamor, I. (2019). The first 20 years of *Organizational Research Methods*: Trajectory, impact, and predictions for the future. *Organizational Research Methods*. Advance online publication. doi:10.1177/1094428118786564

Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal*. Advance online publication. doi:10.1002/SMJ.3015

Aguinis, H., & Vandenberg, R. J. (2014). An ounce of prevention is worth a pound of cure: Improving research quality before data collection. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 569-595.

Antonakis, J., House, R. J., & Simonton, D. K. (2017). Can super smart leaders suffer from too much of a good thing? The curvilinear effect of intelligence on perceived leadership behavior. *Journal of Applied Psychology*, *102*, 1003-1021.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*, 3-25.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*, 452-454.

Banks, G. C., O'Boyle, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., . . . Adkins, C. L. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management*, *42*, 5-20.

Becker, T. E., Atinc, G., Breaugh, J. A., Carlson, K. D., Edwards, J. R., & Spector, P. E. (2016). Statistical control in correlational studies: 10 essential recommendations for organizational researchers. *Journal of Organizational Behavior*, *37*, 157-167.

Becker, T. E., Robertson, M. M., & Vandenberg, R. J. (2019). Nonlinear transformations in organizational research: Possible problems and potential solutions. *Organizational Research Methods*. Advance online publication. doi:10.1177/1094428118775205

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. (2017). Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization*, *15*, 423-436.

Bernerth, J., & Aguinis, H. (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology*, *69*, 229-283.

Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. (2016). Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*, *37*, 257-261.

Bliese, P. D. (2018). Editorial. *Organizational Research Methods*, *21*, 3-5.

Butler, N., Delaney, H., & Spoelstra, S. (2017). The grey zone: Questionable research practices in the business school. *Academy of Management Learning & Education*, *16*, 94-109.

Byington, E. K., & Felps, W. (2017). Solutions to the credibility crisis in management science. *Academy of Management Learning & Education*, *16*, 142-162.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T-H., Huber, J., Johannesson, M., ... Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*, 1433-1436.

Chen, J., Bamberger, P. A., Song, Y., & Vashdi, D. R. (2018). The effects of team reflexivity on psychological well-being in manufacturing teams. *Journal of Applied Psychology*, *103*, 443-462.

Chiu, C., Balkundi, P., & Weinberg, F. J. (2017). When managers become leaders: The role of manager network centralities, social power, and followers' perception of leadership. *Leadership Quarterly*, *28*, 334-348.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Courtright, S. H., Gardner, R. G., Smith, T. A., Mccormick, B. W., & Colbert, A. E. (2016). My family made me do it: A cross-domain, self-regulatory perspective on antecedents to abusive supervision. *Academy of Management Journal*, *59*, 1630-1652.

Crawford, G. C., Aguinis, H., Lichtenstein, B., Davidsson, P., & McKelvey, B. (2015). Power law distributions in entrepreneurship: Implications for theory and research. *Journal of Business Venturing*, *30*, 696-713.

Dineen, B. R., Duffy, M. K., Henle, C. A., & Kiyoung, L. (2017). Green by comparison: Deviant and normative transmutations of job search envy in a temporal context. *Academy of Management Journal*, *60*, 295-320.

Follmer, E. H., Talbot, D. L., Kristof-Brown, A. L., Astrove, S. L., & Billsberry, J. (2018). Resolution, relief, and resignation: A qualitative study of responses to misfit at work. *Academy of Management Journal*, *61*, 440-465.

Gilbert, D. (October 3, 2016). *Rosetta landing: Comet-chasing probe makes final touchdown*. Retrieved from https://www.cnn.com/2016/09/30/health/rosetta-landing-comet-philae/index.html

Grand, J. A., Rogelberg, S. G., Banks, G. C., Landis, R. S., & Tonidandel, S. (2018). From outcome to process focus: Fostering a more robust psychological science through registered reports and results-blind reviewing. *Perspectives on Psychological Science*, *13*, 448-456.

Honig, B., Lampel, J., Baum, J. A. C., Glynn, M. A., Jing, R., Lounsbury, M., ... van Witteloostuijn, A. (2018). Reflections on scientific misconduct in management: Unfortunate incidents or a normative crisis? *Academy of Management Perspectives*, *32*, 412-442.

Huang, J. L., Ryan, A. M., Zabel, K. L., & Palmer, A. (2014). Personality and adaptive performance at work: A meta-analytic investigation. *Journal of Applied Psychology*, *99*, 162-179.

Joo, H., Aguinis, H., & Bradley, K. J. (2017). Not all non-normal distributions are created equal: Improved theoretical and measurement precision. *Journal of Applied Psychology*, *102*, 1022-1053.

Kalaignanam, K., Kushwaha, T., M., Steenkamp, J. E., & Tuli, K. R. (2013). The effect of CRM outsourcing on shareholder value: A contingency perspective. *Management Science*, *59*, 748-769.

Kaul, A., Nary, P., & Singh, H. (2018). Who does private equity buy? Evidence on the role of private equity from buyouts of divested businesses. *Strategic Management Journal*, *39*, 1268-1298.

Ketchen, D. J., Ireland, R. D., & Baker, L. T. (2013). The use of archival proxies in strategic management studies: Castles made of sand? *Organizational Research Methods*, *16*, 32-42.

Le, H., Schmidt, F. L., & Putka, D. J. (2009). The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organizational Research Methods*, *12*, 165-200.

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*, 26-46.

McDonnell, M-H., & Brayden, K. (2013). Keeping up appearances: Reputational threat and impression management after social movement boycotts. *Administrative Science Quarterly*, *58*, 387-419.

Neckebrouck, J., Schulze, W., & Zellweger, T. (2018). Are family firms good employers? *Academy of Management Journal*, *61*, 553-585.

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, *17*, 372-411.

Ryu, W., McCann, B. T., & Reuer, J. J. (2018). Geographic co-location of partners and rivals: Implications for the design of R&D alliances. *Academy of Management Journal*, *61*, 945-965.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147-177.

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, *65*, 445-494.

Spector, P. E., & Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, *14*, 287-305.

Stanley, M. H. R., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A., & Stanley, H. E. (1995). Zipf plots and the size distribution of firms. *Economics Letters*, *49*, 453-457.

Stone-Romero, E. F. (2011). Research strategies in industrial and organizational psychology: Nonexperimental, quasi-experimental, and randomized experimental research in special purpose and nonspecial purpose settings. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 1, pp. 37-72). Washington, DC: American Psychological Association.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research*, *1*, 77-100.

Truelove, E., & Kellogg, K. C. (2016). The radical flank effect and cross-occupational collaboration for technology development during a power shift. *Administrative Science Quarterly*, *61*, 662-701.

Tsui, A. S. (2013). The spirit of science and socially responsible scholarship. *Management and Organization Review*, *9*, 375-394.

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2018). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management*, *44*, 249-279.

Author Biographies

Herman Aguinis is the Avram Tucker Distinguished Scholar and Professor of Management at The George Washington University School of Business. His research is multidisciplinary and addresses the acquisition and deployment of talent in organizations and organizational research methods. He has published nine books and about 160 articles in refereed journals. He is a fellow of the Academy of Management (AOM), former editor of *Organizational Research Methods*, and received the Academy of Management Research Methods Division Distinguished Career Award. He was elected for the five-year presidency track of AOM and will be serving as AOM program chair elect and vice president elect, program chair and vice president, president elect, president, and past president during 2018 to 2023.

N. Sharon Hill is an associate professor of management at The George Washington University School of Business. Her research focuses on virtual work arrangements and organizational change. Her articles have appeared in several leading journals, including *Organizational Behavior and Human Decision Processes*, *Organization Science*, *Personnel Psychology*, *Leadership Quarterly*, and *Academy of Management Annals*. Dr. Hill is a member of the editorial review board of *Personnel Psychology*.

James R. Bailey is a professor of leadership at The George Washington University School of Business. He has published over 50 scholarly articles, cases, and books and is a frequent contributor to practitioner sources such as the *Harvard Business Review* and the *Washington Post.* He is former editor of the *Academy of Management Learning and Education* and current research coordinator for the Management Education and Development Division of the Academy of Management.