Maula, Markku; Stam, Wouter

Enhancing rigor in quantitative entrepreneurship research

Published in:
Entrepreneurship Theory and Practice

DOI:
10.1177/1042258719891388

Published: 01/11/2020

Document Version
Publisher's PDF, also known as Version of record

Entrepreneurship Theory and Practice
2020, Vol. 44(6) 1059–1090

DOI: 10.1177/1042258719891388
journals.sagepub.com/home/etp
SAGE

Enhancing Rigor in Quantitative Entrepreneurship Research

Markku Maula[1] and **Wouter Stam**[2]

Abstract

Reflecting on common empirical concerns in quantitative entrepreneurship research, recent calls for improved rigor and reproducibility in social science research, and recent methodological developments, we discuss new opportunities for further enhancing rigor in quantitative entrepreneurship research. In addition to highlighting common key concerns of editors and reviewers, we review recent methodological guidelines in the social sciences that offer more in-depth discussions of particular empirical issues and approaches. We conclude by offering a set of best practice recommendations for further enhancing rigor in quantitative entrepreneurship research.

Keywords

editorial, entrepreneurship, research methods, quantitative research

The field of entrepreneurship has come a long way over the past decades as a result of entrepreneurship scholars' strong dedication to developing its theoretical basis using increasingly sophisticated research methods (Busenitz et al., 2003; Crook et al., 2010; Davidsson, 2016; Wiklund et al., 2011). Nevertheless, there is currently a strong push for more rigor in quantitative entrepreneurship research, which is driven by several recent developments: a growing recognition that the popular empirical approaches in entrepreneurship research have important limitations, a broad concern about the reproducibility of prior findings in social science research, and rapid advances in available research methods that allow for conducting more rigorous studies. In this editorial, we review these important developments and discuss how they create new opportunities for authors, editors, and reviewers to further enhance rigor in quantitative entrepreneurship research.

The need to improve rigor and transparency is not limited to entrepreneurship research. Research practices in many social science fields are increasingly criticized for lacking the rigor and transparency that would allow others to understand and replicate research findings (e.g., Baker, 2016; Gelman, 2018; Hubbard, 2015; McShane et al., 2019; Miguel et al., 2014; Nosek

[1]Aalto University, Espoo, Finland
[2]Vrije Universiteit Amsterdam, The Netherlands

Corresponding Author:
Markku Maula, Department of Industrial Engineering and Management, Aalto University, PO Box 15500, FI-00076, AALTO, Finland.
Email: markku.maula@aalto.fi

et al., 2015). This "replication crisis" has now been broadly acknowledged in economics (Christensen & Miguel, 2018), sociology (Freese, 2007), political science (Laitin & Reich, 2017), and psychology (Lindsay, 2015). The field of business and management is not immune to these concerns (Bergh et al., 2017; Starbuck, 2016). Indeed, many leading journals have begun to revise their editorial policies to ban poor research practices and promote more rigor and transparency (e.g., Antonakis, 2017; Bettis et al., 2016; Chen, 2018; Hahn & Ang, 2017; Lewin et al., 2016; Meyer et al., 2017).

In addition to the push for more rigor and the need to address the identified problems of the prevailing null hypothesis significance testing (NHST) paradigm (e.g., Gigerenzer, 2004; Matthews, 2019; McShane et al., 2019; Nosek et al., 2015; Schwab et al., 2011; Wasserstein et al., 2019),[1] rapid developments in empirical methods and software tools also create new opportunities for quantitative entrepreneurship research. These include novel approaches to causal inference (e.g., rapidly evolving quasi-experimental techniques that use the potential outcomes [PO] framework; (Abadie & Cattaneo, 2018; Imbens & Wooldridge, 2009; Imbens & Rubin, 2015) and causal graphs/directed acyclic graphs [DAGs]; (Pearl, 1995; 2009; 2016; Rohrer, 2018), an increasing application of Bayesian statistics (e.g., Andraszewicz et al., 2015; Gelman et al., 2013; Lohrke et al., 2018; Vandekerckhove et al., 2018), big data and data science methods (e.g., George et al., 2016; Schwab & Zhang, 2019; Tonidandel et al., 2018), machine learning (e.g., Kolkman & van Witteloostuijn, 2019; Mullainathan & Spiess, 2017), and powerful and innovative visualizations (e.g., Ertug et al., 2018; Greve, 2018; Healy, 2018; Levine, 2018). Given these important recent developments, it is clear that many new opportunities have emerged for enhancing rigor in quantitative entrepreneurship research. Doing so is particularly important given the unique challenges of entrepreneurship as a research context. Indeed, designing and implementing rigorous entrepreneurship studies is far from trivial because of the uncertainty, heterogeneity, and disequilibrium in entrepreneurial phenomena and the common focus on the emergence of new ventures with limited reliable data available on them (Davidsson, 2016).

Therefore, we view it as important for ETP as a leading entrepreneurship journal to offer a timely review of these recent developments and support authors, reviewers, and editors in further enhancing rigor in quantitative entrepreneurship research. In this editorial, we seek to offer a comprehensive review of the key issues involved in designing, conducting, and reporting high-quality quantitative entrepreneurship research. By covering this whole spectrum of empirical considerations, we seek to complement prior reviews of quantitative research methods published in ETP (e.g., Chandler & Lyon, 2001; Connelly et al., 2010; Dean et al., 2007) and more focused recent discussions in other leading entrepreneurship journals (e.g., Anderson et al., 2019; Wennberg & Anderson, 2019). In addition to reviewing key concerns, this editorial also points to recent methodological advances and guidelines in the social sciences, thereby serving as a key entry point for quantitative entrepreneurship scholars who wish to learn more about new ways of conducting rigorous research.

Table 1 summarizes our seven key best practice recommendations for further enhancing rigor in quantitative entrepreneurship research; these concern the importance of (1) matching the research design with the research problem; (2) understanding the advantages and limitations of particular sources of data; (3) ensuring that the measures measure what they are supposed to measure; (4) selecting appropriate analytical tool(s) depending on the key considerations of your particular research question and empirical setting; (5) reporting the methods and results in a transparent and reproducible manner and interpreting the results carefully and thoughtfully; (6) developing a robust workflow to facilitate reproducibility and to minimize errors; and (7) continuing to learn about the evolving methods applicable to quantitative entrepreneurship research.

Table 1. Suggestions for Enhancing Rigor in Quantitative Entrepreneurship Research.

Key considerations	More detailed suggestions
1. Match the research design with the research problem	If the key research design choices, such as the sample and the measures, are not suitable for answering the research problem, it is hard to address the mismatch in a revision. Therefore, the research problem should drive the research design and the choice of the empirical methods. Be clear whether the research is exploratory or hypothesis testing. Consider causal identification already in the research design
2. Understand the advantages and limitations of particular sources of data	Whether or not the data are collected by you or by a data provider, the data collection process should be understood and explained. The implications of the limitations for answering the research problem should be carefully considered. Consider novel data sources such as web-scraped data and video data
3. Ensure that the measures measure what they are supposed to measure	Careful consideration of the validity of the measures is critical. Be transparent about how measures were constructed, adapted, and validated to ensure the reproducibility of research findings. Consider novel measurement techniques such as text mining
4. Select appropriate analytical tool(s) depending on the key considerations of your particular research question and empirical setting	With the increasing availability of specialized statistical tools and the complex nature of entrepreneurial phenomena, it is critical for scholars to carefully consider the key concerns in their particular settings (e.g., the type of measures, potential endogeneity concerns, potential longitudinal data structure, etc.,) and choose the methods accordingly. Consider new approaches, such as Bayesian statistics, to overcome the limitations of traditional techniques
5. Report the methods and results in a transparent and reproducible manner and interpret the results carefully and thoughtfully	Explain and justify the method choices in a transparent manner so as to enable replication. Employ best practices to probe interaction, mediation, and nonlinear effects. Use online appendices for robustness checks if needed. Interpret the results carefully and thoughtfully, avoiding the term "statistically significant" and considering also practical significance. Use visualizations if possible
6. Develop a robust workflow to facilitate reproducibility and to minimize errors	With the increasing complexity and sophistication of entrepreneurship research, a robust reproducible workflow and data access are becoming increasingly important to facilitate error detection and correction, as well as replication
7. Continue to learn about the evolving methods applicable to quantitative entrepreneurship research	Invest time and effort in lifelong learning about new methods and their applicability to quantitative entrepreneurship research. Given the rapid pace of methodological development, continuous learning is important for the progress of the field. It is the responsibility of entrepreneurship scholars to be able to conduct rigorous, leading-edge quantitative research on entrepreneurship phenomena

In the following sections, we start by reviewing frequent concerns in quantitative ETP submissions and continue by discussing the novel opportunities and complexities arising from new data sources and methodological advancements. We close by offering broader suggestions concerning enhanced rigor in the workflow and openness of reporting to facilitate replication and faster knowledge accumulation in quantitative entrepreneurship research.

Enhancing Rigor Across the Research Process

Research Design

Social scientists increasingly recognize that for research to be relevant, it must *also* be rigorous (Van de Ven & Johnson, 2006; Vermeulen, 2005). Rigor implies that scholars are committed to developing and applying best practices for designing, conducting, and reporting scientific studies to enable faster knowledge accumulation. In the field of entrepreneurship, however, it is often difficult, if not impossible, to design a perfectly rigorous study because of the uncertainty, heterogeneity, and disequilibrium in entrepreneurial phenomena and the common focus on new venturing activities for which reliable data are often unavailable (Davidsson, 2016).

Still, research design problems are an important reason for rejections of ETP manuscripts because they undermine the confidence of the readers toward the rigor of the study and its findings, and because they often cannot be resolved within a normal revision. Among such problems, a mismatch between the research question and the research design is one of the key concerns leading to rejection, and this mismatch has many manifestations. For instance, papers that explain performance or change can rarely be successful if the empirical analysis relies on cross-sectional data. Similarly, papers with empirical measures that are disconnected from the theoretical concepts they are supposed to measure are often rejected in the review process. Sampling problems are common in entrepreneurship research (Short et al., 2010), and challenging to address in a revision. Threats to causality are also increasingly identified as a major concern.

Thus, the most critical task for researchers is to carefully consider the research question and match the research design with the research question. Most of the other empirical choices follow from this task and should be considered early on. For instance, matching the unit of analysis (e.g., entrepreneur- vs. firm-level analysis) in the empirical data collection with the level of theoretical argumentation is straightforward before the data collection, but a mismatch would be difficult to address afterwards. A careful research design anticipates key threats to validity and considers how to eliminate these threats. For instance, when empirically testing one causal explanation, it is important to consider alternative explanations for the association between independent and dependent variables and whether and how they could be ruled out.

Carefully considering these issues before the data are collected tends to be more effective than using post hoc methods to compensate for a poor research design. A carefully designed study indeed often requires less from the data analyses and robustness tests (Aguinis & Vandenberg, 2014; Bettis et al., 2014). That said, it is equally important that authors provide a compelling justification for the chosen research design, discuss the design's weaknesses in addition to its strengths, and offer sufficient details about the sampling criteria and data collection procedures. Doing so allows reviewers and editors to better understand and appreciate the study's research design, thereby facilitating its replicability.

In general, papers that fall methodologically short compared to other recent papers tend to face more criticism from reviewers. This means that as the field of entrepreneurship becomes more mature, scholars face ever higher expectations and requirements concerning the rigor of their quantitative hypothesis-testing research.[2] Therefore, it is important for entrepreneurship scholars to continuously enhance their understanding of the latest advances in research methods.

In the sections below, we seek to help scholars make more informed methodological choices by discussing important methods issues in quantitative entrepreneurship research, pointing to recent in-depth discussions of these issues in the literature, and providing a set of best practice recommendations.

Data Sources

Obtaining high-quality data on entrepreneurial phenomena is difficult. Reliable secondary data are often lacking for early-stage venturing efforts, while gaining access to primary data is challenging when entrepreneurs are too busy and their ventures are rapidly changing. Recently, however, rapid increases in the availability and breadth of data sources are transforming the field. For instance, while in the review of the preceding decade of entrepreneurship research by Chandler and Lyon (2001), a typical study was based on a cross-sectional survey given the argued unavailability of other sources of data, such arguments are no longer felt justified, and cross-sectional single-respondent survey studies are now rarely accepted without compelling justification. The ongoing big data revolution indeed enables entrepreneurship scholars to assemble large amounts of data from a variety of sources that offer many new possibilities for studying entrepreneurial phenomena (Davidsson, 2016). At the same time, this growth in data sources also creates new challenges for scholars who wish to use them in empirical studies (George et al., 2016; Schwab & Zhang, 2019; Wenzel & Van Quaquebeke, 2018). Consequently, we discuss some key issues related to different data sources for quantitative entrepreneurship research below.

Survey data. A popular data source in the field is survey data. Despite the known criticisms of the survey method (e.g., potential nonresponse bias, retrospective bias, common method bias, measurement error, and perceptual nature of data), one of its primary benefits is that surveys allow for the direct measurement of complex, latent constructs through multi-item scales. For many interesting entrepreneurship phenomena, the only way to obtain data is to ask people. Unfortunately, survey sampling frames are often unknown or inaccurate because many entrepreneurial efforts are difficult to detect or will have failed prior to data collection (Davidsson, 2016). Measurement validity is another frequent concern that can be compromised, for instance, when surveys utilize individual respondents to measure firm- or industry-level constructs (Davidsson & Wiklund, 2001). For these reasons, it is important to provide a clear exposition of the survey data collection procedures and clarify how and why particular respondents were selected; what measures were taken to minimize potential biases; and how the multi-item scales were constructed, adapted, and validated (Aguinis et al., 2018). A lack of transparency often causes confusion and negative reactions among reviewers, which is understandable given that survey data are prone to biases and, thus, require careful treatment (Huang et al., 2015). In addition, we observe that submissions all too often rely on a single survey, single respondent design. While this may be unavoidable or acceptable in some cases, we do believe that multi-wave surveys and surveys using multiple informants (e.g., multiple founders from the same venture team) are quickly becoming the new standard, as they help address the limitations of cross-sectional single respondent surveys. There is also increasing recognition of the need to adapt survey research practices for work examining nontraditional contexts (Kriauciunas et al., 2011)—settings that receive increasing attention from entrepreneurship researchers.

Archival data. Entrepreneurship scholars frequently use archival data to test their hypotheses. Examples of archival data include data recorded in official government registers (e.g., linked employer-employee data or tax filings; e.g., Goetz et al., 2015), industry association records (e.g., data on venture capital investments or entry and exit of firms), commercial databases (e.g., data on companies and their patents, venture capital (VC) investments, initial public offerings (IPOs), and mergers and acquisitions (M&As) (e.g., Dalle et al., 2017), as well as various scholarly data collection initiatives, such as the Panel Study of Entrepreneurial Dynamics (PSED; Reynolds, 2007), the Global Entrepreneurship Monitor (GEM; Levie et al., 2014; Reynolds et al., 2005), the Kauffman Firm Survey (KFS; Farhat et al., 2018; Farhat & Robb, 2018), and others (e.g., Brown et al., 2017). Compared to primary data, archival data have several potential advantages, including the ability to quickly and unobtrusively collect a large number of repeated measures across different levels of analysis, allowing scholars to draw better causal inferences and employ multilevel research methods (Wennberg, 2005). However, archival data have its own problems. One major weakness is that the available variables are rarely ideal measures (Singleton & Straits, 2017). Typically, multiple sources of archival data need to be merged to address focal research questions, which is not a trivial process and therefore requires care (e.g., Christen, 2019; Tarasconi & Menon, 2017). The messiness of archival data thus requires that authors provide a detailed account of the databases that were used, clarify the sampling criteria and procedures that were used to construct the databases and associated measures (Farhat & Robb, 2018; Zhang & Shaw, 2012; Wennberg, 2005), and discuss the magnitude and implications of missing data (e.g., Enders, 2010; Newman, 2014). Many commercial databases are primarily designed to serve commercial clients who require a snapshot of the current situation, not a comprehensive historical record. Therefore, issues such as backfilling data and updating categories afterward can cause problems when historical data points are reclassified based on current information. Many commercial databases are based on voluntary data provision and may only cover relatively successful firms that have survived for some time, thus leading to potential sample selection biases. Therefore, archival data collected by official government registers (e.g., linked employer–employee data in many countries) can sometimes be valuable for addressing research questions in entrepreneurship, but they also often have important limitations because the data were originally collected for other purposes and their quality can differ depending on the country and the government register. Thus, as with all other data sources, it is crucial that authors understand and explain to readers how the data were produced and the limitations of these procedures (Connelly et al., 2016).

Website-scraped data. The big data revolution has created many new ways to collect high-quality data on entrepreneurial phenomena by scraping data from the Internet. Web scraping refers to the automated identification, extraction, and coding of information stored on websites. Compared to traditional data sources, web scraping offers several key advantages, including the ability to quickly and unobtrusively collect large-scale data on entrepreneurial activities that unfold across actors, levels, and time (Braun et al., 2018; Landers et al., 2016). In turn, advances in data science methods and software tools increasingly allow entrepreneurship scholars to leverage web-scraped data to develop and test theories (George et al., 2016; Prüfer & Prüfer, 2019). Recent studies indeed illustrate how applying techniques such as machine learning and text mining to web-scraped data can offer new insights into entrepreneurial phenomena (Kolkman & van Witteloostuijn, 2019; Lee et al., 2017; Obschonka et al., 2017). It is important to note, however, that researchers must be aware of the weaknesses of web-scraped data and be extremely careful in designing and executing theory-driven web-scraping projects, allowing them to better address ethical concerns and draw more meaningful conclusions (for guidelines, see Braun et al. (2018) and Landers et al. (2016)).

Video data. Video data are increasingly used to study the behaviors of entrepreneurs, investors, and other relevant stakeholders. Video recordings offer exciting new possibilities to capture rich data on, for instance, the bodily expressions, emotions, decision-making, and identity development of entrepreneurs—concepts that can be difficult to capture through other means (Zundel et al., 2018). A growing number of entrepreneurship scholars are discovering the power of video-based research by coding preexisting videos to measure independent or dependent variables and by creating artificial videos that can be used in experiments (e.g., Chen et al., 2009; Ciuchta et al., 2018). Thus, we predict that ETP will receive more submissions that both leverage these unique benefits and address the specific challenges associated with video research (Toraldo et al., 2018).

Experimental data. Experiments can help to overcome the endogeneity problems in entrepreneurship research that are prevalent in research designs based on observational data (Hsu et al., 2017). However, only a small number of ETP submissions currently employ experimental data. We foresee this number to increase because the experimental method offers opportunities to generate better causal evidence on various micro and macro theories in entrepreneurship that can be tested by manipulating independent variables in laboratory as well as field settings (e.g., see special issues by Acs et al. (2010) and Williams et al. (2019)). However, authors must be aware of the key pitfalls in designing, running, and reporting experiments (Chatterji et al., 2016; Hsu et al., 2017; Singleton & Straits, 2017) because reviewers often rightly challenge their construct and external validity when manuscripts do not clearly explain and justify the experimental designs (e.g., Grégoire et al., 2019). For instance, simplified experiments conducted using undergraduate students (or outsourced to MTurk or another service) may not be able to adequately capture the high stakes and emotions that entrepreneurs encounter in real-world situations, so their appropriateness for a particular research problem should be justified carefully (Hsu et al., 2017; Stevenson & Josefy, 2019). Assuming the subjects are appropriate, experiments must also be designed to realistically capture the focal phenomenon. Here, a key limitation is that experiments typically focus on only a few effects, and thus, the results may not be generalizable to the real-world complexities that entrepreneurs must navigate. The experimental vignette methodology can sometimes be used to increase the realism of entrepreneurship experiments (Aguinis & Bradley, 2014). Although not always feasible, *field experiments* that use randomization in naturally occurring settings can also be a great way to conduct realistic experiments in entrepreneurship research (Chatterji et al., 2016; Harrison & List, 2004; Levitt & List, 2009). Overall, given the multitude of different experimental designs with important tradeoffs,[3] it is important to consider the focal research problem carefully when designing an experimental study.

Triangulating multiple data sources. A growing trend in the field is to employ research designs in which multiple data sources, often quantitative and qualitative, are combined in a single study. These mixed methods approaches (Molina-Azorin et al., 2017) hold much promise for entrepreneurship research because they allow scholars to leverage the strengths, and possibly minimize the weaknesses, of different data sources and thus arrive at more insightful findings than could be produced by one method alone (but it is good to remember that this could also happen across studies in research programs). The field of entrepreneurship, which has witnessed various calls for studying context effects and causal mechanisms, could thus greatly benefit from studies that combine different forms of data to triangulate research findings and pinpoint the mechanisms and contextual boundaries for the results (Jick, 1979; Molina-Azorín et al., 2012). Nevertheless, it is important to be aware of the challenges and limitations of mixed methods research and provide a compelling justification for the chosen approach in addition to clearly reporting how the different techniques were combined (Creswell & Plano Clark, 2011; Turner et al., 2017).

Data transparency. While it is common and understandable for authors to seek publication of multiple papers from the same dataset, we observe that this data overlap is not always accurately reported in the submitted manuscripts. At ETP, we therefore now require that authors explicitly discuss data overlaps in a separate document that offers a detailed account of how and which variables in the focal study differ from those in prior publications. Including a uniqueness analysis table (Kirkman & Chen, 2011) in a cover letter, which lists the variables of the current and prior studies, along with explanations of their operationalizations, can be an efficient approach to convey this information to the editors. Although using the same or a very similar variable across different studies based on the same dataset can be problematic (e.g., leading to "salami sliced" publications; Antonakis, 2017)—particularly when it concerns the dependent variables—this practice can also be a requirement when prior publications from the same dataset indicate the importance of particular predictors; that is, not including these predictors could severely bias the model estimates so that variable overlap is actually recommended to obtain valid research findings in such situations. It is also better to use a well-defined and validated existing measure rather than develop marginally different variants presenting them as new, as the former approach improves the opportunities for replication, meta-analysis, and overall progress and accumulation of the research.

Measurement

Quantitative empirical entrepreneurship research typically tests hypotheses derived from one or several theories. For the proper testing of theories, it is crucial that the focal concepts are clearly defined (e.g., Podsakoff et al., 2016; Singleton & Straits, 2017) and that the construct operationalizations are appropriate such that the measures actually measure what they are supposed to measure (e.g., Borsboom et al., 2004). All too often, there are wide gaps between theoretical arguments and empirical measures, which is often seen as a critical and difficult to address problem leading to rejections of many ETP manuscripts. Below, we discuss some important measurement concerns in quantitative empirical entrepreneurship research.

Construct validity. When testing hypotheses concerning relationships between theoretical constructs, construct validity is of central importance. To ensure construct validity, authors should provide clear construct definitions, demonstrate that the study's empirical indicators reflect their underlying constructs (Edwards, 2003), and ideally, report sufficient tests so that reviewers and readers can independently assess the construct validity (Aguinis et al., 2018; DeVellis, 2016). This is not an easy task for entrepreneurship scholars who frequently study novel phenomena for which validated measures may not always be readily available. Nevertheless, whenever possible, we recommend that authors use multi-item scales rather than single or categorical indicators to measure complex constructs, provide sufficient evidence of scale reliability and dimensionality when multiple indicators are used, and make efforts to validate the focal measures by using alternative operationalizations of the same construct and analyze the convergent validity (Carlson & Herdman, 2012). Indeed, when the study's findings can be replicated with multiple measures of the focal constructs, this will enhance their perceived validity. In assessing the convergent and discriminant validity of constructs, it is worth noting that many widely used "rules of thumb" and cutoff criteria have been dismissed in the methods literature as "methodological myths" (Lance & Vandenberg, 2009), such as the false notion that a measure is internally consistent when Cronbach's α exceeds .70 (Guide & Ketokivi, 2015; McNeish, 2018). It is therefore useful to carefully consider the validity of measures in a more substantive fashion.

Transforming variables. One area that requires more attention and care relates to the common practice of applying (nonlinear) transformations of variables. The use of transformations can sometimes be appropriate, but submissions often apply transformations without offering a clear explanation for why and how the transformation was performed.[4] This is problematic because using transformed variables can severely undermine the validity and interpretation of findings (Becker et al., 2018). It is therefore important for authors to provide a compelling (theoretical) rationale for applying a particular type of transformation, report exactly how the original variables have been transformed, and properly discuss how the transformations may affect the interpretations of the research findings. In addition, the visualized results should be made consistent with the transformations by properly adjusting the scaling of the Y and X axes and recognizing that the relationship is no longer linear when plotting the results or applying back transformations of the variables into their original units of analysis.[5]

Common method bias. Method bias can be a concern in survey-based research (Podsakoff et al., 2003). While survey data is often needed when studying particular entrepreneurial phenomena, a single-respondent cross-sectional survey design makes it difficult to remedy potential common methods bias. For instance, the commonly used Harman's single factor test is generally not sufficient to dismiss the presence of common method bias (Guide & Ketokivi, 2015; Podsakoff et al., 2003). It is advantageous for authors to use more sophisticated techniques, such as the confirmatory factor analysis (CFA) marker technique, which requires the inclusion of appropriate marker variables in the questionnaire (Podsakoff et al., 2012). Overall, post hoc tests can never fully compensate for a poor ex-ante research design (Richardson et al., 2009).

Control variables. Entrepreneurial phenomena are typically complex and influenced by a large variety of factors, many of which are not included in the study's hypothesized model. To avoid omitted variable bias and to rule out alternative explanations, it is therefore important to include theoretically relevant control variables in statistical analyses (Antonakis et al., 2010). Unfortunately, it is common for submissions to either lack sufficient statistical controls or contain a list of control variables without adequate explanations for their inclusion. This can be problematic because adding additional controls could actually bias the model estimates and result in less accurate research findings if they contain measurement error (Aguinis & Vandenberg, 2014; Spector & Brannick, 2011) or if they are outcome variables of the focal independent (i.e., "bad controls," Angrist & Pischke, 2009, pp. 64–68). We therefore encourage authors to select control variables based on theoretical considerations and offer an explicit justification for why they are included and how their inclusion affected the research findings.

Novel measurement techniques. For the field to move forward, it is important that entrepreneurship scholars remain alert to new opportunities for novel construct measurement. Computer-aided text analysis (CATA), topic modeling, and natural language processing (NLP) using machine learning are relatively novel approaches to derive measures from textual content (Banks et al., 2018; Hannigan et al., 2019). Spurred by the ongoing big data revolution, these approaches enable researchers to mine huge amounts of data (e.g., website scraped data) to develop longitudinal measures at relatively low costs (George et al., 2014; 2016). However, there are important concerns in the use of these methods, such as how to validate CATA-based scales and address measurement error (McKenny et al., 2018; Short et al., 2010). The implication is that the authors applying these novel techniques must explicitly discuss potential sources of measurement error, the extent to which measurement error variance is present, and their implications for interpreting the research findings. In some areas of entrepreneurship research, new opportunities could also emerge from new measurement and data collection methods in other fields of science, such as neuroimaging in the analysis of entrepreneurial decision making (e.g., Lahti et al., 2019; Laureiro-Martínez et al., 2015), or other biological measures related to genetics, hormones, and physiology in other domains of entrepreneurship research (c.f., Nofal et al., 2017).

Analyses

In recent years, scholars' methodological toolkit has expanded tremendously in terms of the types of data analyses that can be performed to test a study's research hypotheses. Below we highlight several important issues with regard to data analyses that we commonly encounter in manuscripts submitted to ETP. While most of these issues are related to testing null hypotheses using the prevailing frequentist paradigm, the maturation of entrepreneurship theories increasingly warrants the analyses to build on prior findings and to focus on the effect sizes and their uncertainty instead of simply testing whether an effect is different from zero. This paves the way for the increased adoption of Bayesian methods, which we discuss last.

Choice of the analytical method. At the most basic level, an important question is which types of data analyses must be performed to test a particular research hypothesis. Authors often seem to assume that the answer to this question is very straightforward and do not sufficiently explain and justify their analytical method. Here, authors must be forthright by discussing the strengths and limitations of the focal method and presenting theoretical arguments and/or empirical evidence regarding the appropriateness of the chosen method for testing the proposed theory. They may also consider combining different analytical techniques to enable replication of the study's core findings and thereby generate more insightful findings.

Longitudinal analysis. Entrepreneurial phenomena are often dynamic in nature, suggesting that longitudinal research is needed to unravel these temporal dynamics. Using longitudinal data obviously offers many benefits, as it enables researchers to better handle unobserved heterogeneity, improve causal inference, and explain change (Bliese et al., 2019). Fortunately, however, the increasing availability of big data is opening up many new opportunities for entrepreneurship scholars to collect longitudinal data (Schwab & Zhang, 2019). At the same time, this means that they must familiarize themselves with the unique decisions and challenges associated with conducting and reporting longitudinal data analyses. For instance, the use of longitudinal data alone is not sufficient to control for unobserved heterogeneity, if fixed effects regression or another appropriate longitudinal method is not used. While the Hausman test was previously considered as the key criterion when choosing between random effects versus fixed effects models, it has now been recognized that the theoretical understanding concerning between-unit versus within-unit variance is central and that random effects models (and generalized estimating equations [GEEs]) that mix these two are often not ideal default models (Certo et al., 2017; Dieleman & Templin, 2014). A hybrid model that explicitly separates the between- and within-unit effects is often a preferable approach (Certo et al., 2017; Dieleman & Templin, 2014; Mundlak, 1978; Schunck & Perales, 2017). Other longitudinal methods include dynamic panel data models (i.e., a lagged dependent variable as a covariate) estimated using techniques such as generalized method of moments (GMM; e.g., Bond, 2002; Bun & Sarafidis, 2015); event history models (also called survival analysis or duration models), which may be needed if time to an event is of interest and right censoring presents a problem (e.g., Allison, 2014; Cleves et al., 2016); latent growth curve models, which explicitly model change (instead of only variance; e.g., Bollen & Curran, 2006; McArdle, 2009); and difference-in-differences (DID) models for causal inference (Lechner, 2011). For longitudinal analysis, it is important to explicitly probe the assumptions underlying these methods, test the form and stability of relationships over time (including appropriate temporal lags), address the issues of attrition and missing data, and evaluate the suitability of the study's research design and dataset for longitudinal analyses (Bergh & Holbein, 1997; Ployhart & Vandenberg, 2010).

Addressing endogeneity. Many manuscripts we receive test hypothesized causal relationships with cross-sectional, nonexperimental data, which is concerning because establishing causality generally not only requires that two variables are correlated but also that the independent variable temporally precedes the dependent variable and that the presumed causal effect cannot be explained by other causes (Antonakis et al., 2010; Kenny, 1979; Shadish et al., 2002).[6] In many submissions, omitted variable bias constitutes an important validity threat as a prime source of endogeneity, which manifests itself when a predictor is correlated with the error term (Antonakis et al., 2010; Hamilton & Nickerson, 2003; Roberts & Whited, 2013). In longitudinal research designs, the use of fixed-effects models controls for time-invariant unobserved heterogeneity and can therefore help address many sources of omitted variable bias. However, in management research aiming for managerial relevance, self-selection-based omitted variable bias is often particularly relevant and requires more attention (Clougherty et al., 2016; Hamilton & Nickerson, 2003). Endogeneity may substantially bias one's research findings, so it is critical that authors explicitly address endogeneity in their manuscripts from both a theoretical and an empirical standpoint (Ketokivi & McIntosh, 2017). Theoretically, it is important to consider the extent to which focal predictors should be treated as sufficiently exogenous and offer strong arguments for why this is true (or, alternatively, they must refrain from making causal claims in their theorizing). It is also important to identify potential endogeneity concerns and determine how serious the threats are that they cause for the validity of the analysis. Empirically, authors are expected to address the identified endogeneity concerns to the greatest realistic extent by testing for different sources of endogeneity and correcting for endogeneity using appropriate methods, including quasi-experimental techniques (Antonakis et al., 2010; Hamilton & Nickerson, 2003; Roberts & Whited, 2013), such as instrumental variable (IV) estimation (Bascle, 2008; Semadeni et al., 2014), regression discontinuity design (RDD) models (e.g., Calonico et al., 2014; 2019), difference-in-differences (DID) models (e.g., Lechner, 2011), and synthetic control methods (e.g., Abadie et al., 2015), or through structural econometric models (e.g., Low & Meghir, 2017).

This said, a few cautionary notes are in order. First, statistical correction for endogeneity must be performed with great care because researchers often misuse these methods due to a lack of understanding of their assumptions and limitations (Bascle, 2008; Ketokivi & McIntosh, 2017). Indeed, Certo et al. (2016) detailed examination of the use of Heckman models demonstrates that scholars frequently misunderstand when and how to use particular procedures for addressing different types of endogeneity. This is problematic because the appropriateness of certain correction techniques depends on the specific source of endogeneity that must be addressed (e.g., omitted variables, measurement error, simultaneous causality; Roberts & Whited, 2013). The implication is that authors must justify why a specific technique is appropriate for addressing a particular source of endogeneity in the focal study and ensure that it is correctly implemented by performing relevant tests and robustness checks (Bascle, 2008; Semadeni et al., 2014).

Second, alternative econometric methods for causal inference are evolving rapidly (e.g., Abadie & Cattaneo, 2018; Imbens & Wooldridge, 2009), largely building on the potential outcomes (PO) framework (Rubin, 1974; 2005), which focuses on the potential outcomes of a unit of interest under alternative states of a cause (i.e., "counterfactuals"; e.g., Imbens, 2019; Imbens & Rubin, 2015; Morgan & Winship, 2007). This rapid development and the multitude of approaches require care in understanding the assumptions of the alternative techniques and in identifying appropriate technique(s) for a particular problem at hand. For instance, although temporal dependence has traditionally been a key assumption for causal inference, in modern approaches to causal inference, the assumption about the temporal dependence may not be always needed (see, e.g., Imbens & Rubin, 2015; Pearl, 2009, p. 57-59). In models involving matching, propensity score matching (PSM) has been by far the most popular choice. However, PSM has been found to suffer from important limitations (King & Nielsen, 2019), with coarsened exact matching (CEM) being recommended as a robust alternative (Iacus et al., 2009; 2012; King & Nielsen, 2019). Furthermore, when advocating for the PO framework (Rubin, 1974) for causal inference, structural equation modeling (SEM) has previously been falsely portrayed as an incompatible (and as an inferior approach for causal inference) approach without recognizing their equivalency (Bollen & Pearl, 2013; Pearl, 2012). And although different disciplines have used different notations for seemingly incompatible approaches to causal inference, causal graphs/directed acyclic graphs (DAGs) have been identified as useful tools for identifying the required assumptions for causal inference and thereby helping to unify various approaches to causal inference (Pearl, 1995; 2009; 2016; Pearl & Mackenzie, 2018; Rohrer, 2018).

Third, we must emphasize that post hoc statistical corrections for endogeneity often cannot compensate for a poor ex-ante research design (Shaver, 2019). While appropriate correction techniques may often be helpful for increasing the validity of research findings if they can be implemented appropriately, we strongly believe that entrepreneurship research will gain even more when researchers start to design their studies from day one with a focus on anticipating and overcoming key endogeneity concerns. This means, for instance, that whenever possible, they try to construct true random samples instead of sampling on the dependent variable (i.e., also survey ventures that did not survive) and conduct more experiments in the lab or field. Doing so will help the field to draw more informative inferences about causal relationships without having to resort to post hoc correction techniques that often mask fundamental problems in our research designs.

Finally, although addressing potential endogeneity is an important concern in rigorous entrepreneurship research, it can never be fully solved (i.e., there will always be untested assumptions; Ketokivi & McIntosh, 2017). Therefore, to not limit future quantitative entrepreneurship research to a limited set of questions that can be studied using particular methods (Shaver, 2019), we echo the recommendation of Ketokivi and McIntosh (2017) to consider what can be realistically expected in a study of a particular question. Given the complexity of the focal phenomena in entrepreneurship research and the inherent limitations of empirical methods, improved causal identification requires a cumulative body of empirical research (Shaver, 2019).

Nonlinear models. In many entrepreneurship studies, nonlinear models are used. For instance, when the outcome of interest is a limited dependent variable, such as a binary variable or a count variable that takes on a limited number of discrete values, a nonlinear model is generally needed. Examples include studies that seek to predict the likelihood of new firm survival, the number of opportunities identified by individuals, or the entry mode used by new international ventures. In all these cases, the choices or outcomes cannot be operationalized as a continuous dependent variable. Although a linear probability model (LPM) approach is sometimes considered as an alternative approach to facilitate easier interpretation of the marginal effects for binary outcomes (e.g., Angrist & Pischke, 2009; Greene, 2012; Wooldridge, 2010, pp. 562–565; Wooldridge, 2015, pp. 224-228), models with limited dependent variable models are generally estimated using alternative modeling techniques, such as logit, probit, tobit, Poisson, negative binomial regression, and survival analysis (e.g., Cameron & Trivedi, 2005). However, correctly implementing these methods and interpreting the results requires great care (Wiersema & Bowen, 2009). Consequently, we encourage authors to familiarize themselves with best practices for analyzing and interpreting these nonlinear models. For instance, common pitfalls relate to the improper interpreting of coefficients, modeling interactions, comparing coefficients across groups, and probing measures of model fit (Hoetker, 2007). The interpretation of effects in nonlinear models has been a subject of debate and has led to many alternative approaches (e.g., Ai & Norton, 2003; Angrist & Pischke, 2009; Greene, 2010; Zelner, 2009); therefore, it is important to consider the interpretation carefully (Hoetker, 2007). Although this debate has not resulted in any unproblematic silver bullet solutions, it is imperative that authors explicitly address the interpretation carefully in their manuscripts, including plotting of the effect sizes and their confidence intervals (see also our related discussion of reporting the effect sizes below).

Moderation and mediation analyses. Entrepreneurship researchers currently often go beyond merely examining the main effects of focal predictors by conducting moderator and mediator analyses. While moderation analyses are central for understanding under what conditions a theory's predictions hold and, thereby, identify the boundaries of the theory (Andersson et al., 2014; Boyd et al., 2012), mediation analyses can be helpful in pinpointing the causal mechanisms that explain why a predictor may influence the criterion (MacKinnon, 2008). With increased attention to causal inference, the methods for causal mediation analysis have also developed rapidly (see, e.g., Hicks & Tingley, 2011; Keele et al., 2015; Tingley et al., 2014). However, there is often confusion about how moderation and mediation effects should be tested and how the results should be interpreted, especially if they are combined in the same model (Aguinis et al., 2017; Edwards & Lambert, 2007). Common issues we observe in submitted manuscripts include inappropriate dichotomization of moderators, not controlling for the main effects in the model, stating that centering helps to alleviate collinearity issues (see Dalal & Zickar, 2012), inaccurate plotting of the interaction effects (see a related discussion below concerning reporting effects), and failure to conduct simple slope tests (e.g., Dawson, 2014). It is also important to recognize that moderation effects require particular attention in fixed-effects and hybrid models (Shaver, 2019), multi-level (Aguinis & Culpepper, 2015; Andersson et al., 2014), and nonlinear models (e.g., Greene, 2010). Care is also needed when hypothesizing and testing U-shaped or inverted U-shaped relationships (Haans et al., 2016). Indeed, we frequently observe that authors do not formally test and plot the precise nature of the proposed relationships or that they inaccurately report or visualize relevant turning points outside the range of their data. We thus encourage authors to be aware of the many intricacies involved in testing moderation and mediation effects and incorporate available best practice recommendations into their manuscripts.

Multilevel analyses. Many theories in entrepreneurship specify relationships between constructs operating at different levels of analysis, including the individual, team, firm, regional, and country levels (Davidsson & Wiklund, 2001). These relationships may involve cross-level direct effects (e.g., firm-level entrepreneurial orientation influencing individual-level entrepreneurial behavior) or cross-level moderation effects (e.g., national institutions influencing the relationship between a new venture's entry strategy and performance) that can be tested using multilevel techniques, including multilevel random coefficient modeling (RCM; e.g., hierarchical linear modeling [HLM]), mixed-effects models (Rabe-Hesketh & Skrondal, 2012), and multilevel structural equation modeling. One of the primary advantages of using these methods is that they explicitly address the nestedness of the data (i.e., observations cannot be assumed to be independent) and enable researchers to disentangle what proportion of the variance in a particular outcome is produced by variables at different levels of analysis. However, despite these potential benefits, multilevel analyses can be challenging because they require that researchers carefully evaluate and report the multilevel properties of their data (Mathieu & Chen, 2011). Additional care is needed when predicting and testing cross-level interactions in multilevel models (Aguinis & Culpepper, 2015; Andersson et al., 2014). Accordingly, authors should familiarize themselves with these issues prior to designing and executing multilevel studies, and they should address them explicitly in manuscripts submitted to ETP. For instance, they are expected to clearly define and empirically identify relevant levels of analyses (e.g., unit membership and its dynamics), probe the suitability of their data for multilevel analyses (e.g., statistical power and intraclass correlation), consider the model assumptions (e.g., random effects assumptions), justify the modeling techniques used (e.g., HLM versus SEM), and report the multilevel properties of the data (e.g., effect sizes within and between levels). For additional recommendations, see Aguinis et al. (2013).

Bayesian approaches. When entrepreneurship theories mature and some empirical studies have established that an effect seems to exist, the natural progress is to build on prior empirical findings to better understand the effect sizes and their uncertainty. Bayesian statistics are particularly suitable for this purpose (e.g., Gelman et al., 2013; Vandekerckhove et al., 2018) and can prove to be increasingly valuable for entrepreneurship research (Lohrke et al., 2018). Bayesian inference allows researchers to quantify empirical evidence for *any* hypothesis, including the null hypothesis (i.e., X is unrelated to Y), and it enables testing more complex models while combining prior findings with new data, all of which make Bayesian estimation an attractive alternative to traditional significance testing using *p* values (Zyphur & Oswald, 2015). However, entrepreneurship scholars must also be aware of the potential limitations of Bayesian approaches and recognize that Bayesian statistics must also be used with great care and may not always fit the purpose of a particular study (Gigerenzer & Marewski, 2015). However, the computational issues and difficulty of implementation that have previously prevented wider adoption of the Bayesian approach are no longer barriers given the recent integration of Bayesian tools in frequently used statistical software, such as Stata (Lohrke et al., 2018; McCann & Schwab, 2019).

To conclude, entrepreneurship researchers frequently encounter many separate threats to validity simultaneously, and each threat requires a sophisticated treatment. While each issue can often be addressed in isolation with an appropriate approach, an additional challenge in entrepreneurship research is that the field's theoretical models have become increasingly complex, encompassing multiple main effects, as well as moderation or mediation effects. This complexity makes it more difficult to address endogeneity, sample selection, and other concerns in one model. For instance, many econometric techniques to address endogeneity are relatively easy to apply when there is one main effect in a linear model, but they are not easy to apply when the hypothesized model includes multiple endogenous variables with interaction effects to be estimated in a nonlinear model. Although many recent methodological advances, such as Stata's extended regression models (ERM), make it easier to simultaneously address multiple threats to validity (Stata, 2018), it is important that authors carefully prioritize these issues in their own studies. Doing so will help them to make informed choices regarding the main analytical techniques that are most appropriate and may point to possible robustness analyses that should be reported after the main analyses (an issue we will discuss below in more detail). Furthermore, adding complexity to the analysis is not valuable in itself—instead, the simplest feasible analysis approach is generally preferable unless there is a specific reason why it is not appropriate and a more complex analysis approach is needed. Finally, the methods should follow the theoretical progress from identifying a potential mechanism and its effect to a more precise understanding of the effect size and its uncertainty, for which Bayesian approaches can be increasingly valuable.

Reporting

Transparency of reporting has become an increasingly recognized aspect of conducting high-quality entrepreneurship and management research that enhances its credibility and reproducibility (Aguinis et al., 2018). As the field matures, it becomes increasingly important to consider effect sizes instead of mere statistical significance (e.g., Edwards & Berry, 2010; Schwab, 2015). Below, we discuss several recurring issues and the associated recommendations related to the reporting practices of manuscripts submitted to ETP.

Statistical significance. Statistical significance and the interpretation of *p* values have received significant attention in recent years (e.g., Wasserstein & Lazar, 2016; 2019). The *p* value is defined as "the probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value" (Wasserstein & Lazar, 2016), but it is often misinterpreted (Benjamin & Berger, 2019; Greenland et al., 2016; McShane et al., 2019; Wasserstein et al., 2019).[7] As defined by Wasserstein et al. (2019), while "p-values can indicate how incompatible the data are with a specified statistical model" they "do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone." When discussing the statistical significance of the hypothesis tests, the reporting of the exact *p* values and interpreting them in the focal context is recommended instead of focusing on some particular cut-offs (Betensky, 2019; Wasserstein et al., 2019). The excessive focus on arbitrary significance level thresholds to categorize findings either as significant or not significant has recently raised serious concerns across disciplines, leading to calls to ban the term "statistical significance" altogether (e.g., Amrhein et al., 2019; McShane et al., 2019; Wasserstein et al., 2019). To remove incentives for placing too much emphasis on *p* values, or even tweaking analyses to obtain a specific required significance level—a major concern known as p-hacking—many journals also in our fields have already started banning asterisks (e.g., Bettis et al., 2016; Meyer et al., 2017). Despite this very serious concern, a small practical benefit for readers of the asterisks as a form of visualization has been the visual guidance that they give readers when assessing the consistency of significance levels across multiple models and variables in many tables. However, echoing the arguments by Bettis et al. (2016), McShane et al. (2019); Wasserstein and Lazar (2016; 2019), and others, we note that *p* values should not be the only metric of interest in reporting (e.g., effect sizes and their confidence intervals, as discussed below, are increasingly central) and definitely should not be used in a threshold manner as a publication criteria because replication and reporting of nonresults are also important for cumulative and reproducible entrepreneurship research. When conducting multiple hypothesis tests, it might also be appropriate to consider an adjustment for the false discovery rate (FDR; Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001). Nevertheless, when reporting significance levels, it is important for authors to focus on exact *p* values as continuous measures (and clarify whether they are based on two-tailed or one-tailed significance tests, which is often unclear in manuscripts submitted to ETP . Most importantly, whatever the *p* values, they should not be used to claim that a finding is "statistically significant" or "statistically nonsignificant" but a more thoughtful approach to the interpretation of the findings should be adopted (Anderson, 2019; Wasserstein et al., 2019). *P* values continue to have value but should not be misused (Greenland, 2019; Krueger & Heck, 2019; Wasserstein & Lazar, 2016).

Effect size. Whereas the *p* value can indicate how incompatible the data are with a specified statistical model, it does not measure the size of an effect or the importance of a result (Wasserstein & Lazar, 2016). That is, a result with a low *p* value may have little substantive significance when the predictor only explains a negligible fraction of the variance of a dependent variable. For instance, Combs (2010) notes that the increasing availability of large samples empowers researchers to find ever smaller effects that may have little practical relevance. Furthermore, as the field matures, the relevant questions are no longer binary questions of whether or not, but of how much (e.g., Calin-Jageman & Cumming, 2019; Edwards & Berry, 2010). It is therefore critical that authors report effect sizes, which essentially capture the expected change in a dependent variable that results from a change of an independent variable. Reporting effect sizes and their confidence intervals in their original units facilitates intuitive interpretations (Calin-Jageman & Cumming, 2019; Cumming, 2014). However, also reporting the standardized effect sizes (e.g., Cohen's *d* or Pearson's *r*) and their confidence intervals in individual studies facilitates the accumulation of the research and paves the way to a better understanding of the true effect sizes in different contexts (Cumming, 2011; 2014; Schwab, 2015). When reporting and interpreting confidence intervals, it is important to remember that a particular threshold is not any more meaningful in the interpretation of confidence intervals than in the interpretation of p values (e.g., Greenland et al., 2016; Greenland, 2019; Matthews, 2019).

Although the interpretation of the effect size is always context specific and there is no single ideal test, there are multiple approaches that can be used depending on the situation. To understand the effect of a change in the independent variable, marginal effects (especially average marginal effects) are often a helpful way to assess and communicate the effect sizes (e.g., Williams, 2012). This can be fruitfully combined with plotting the marginal effects graphically, which is particularly relevant and informative when interpreting interaction effects in nonlinear models (Greene, 2012, p. 733-741; Hoetker, 2007; Mize, 2019; Wulff, 2015). Other approaches for binary outcomes include the linear probability model (LPM) approach, which is occasionally applied as an easily interpretable approach for binary dependent variables (e.g., Angrist & Pischke, 2009; Cameron & Trivedi, 2005, p. 466-471; Greene, 2012, p. 727; Wooldridge, 2015, p. 224-228), as well as simulation-based approaches (King et al., 2000; Zelner, 2009). As there is no single universal best way to estimate effect sizes, authors should carefully and creatively judge which particular estimate is most appropriate for their particular study (Aguinis et al., 2010; Schwab et al., 2011). This means that submissions should clearly state which particular effect size metrics have been used and report confidence intervals around the effect size estimates if possible (Cumming, 2011; 2014; Schwab, 2015).

In interpreting the effect sizes, we recommend that authors not rely on seemingly random cutoff values to judge the practical significance of their findings but also incorporate qualitative judgments that consider what really matters to practitioners in the specific context of the study (Aguinis et al., 2010; Greenland et al., 2016). Furthermore, when the attention shifts from identifying statistically significant relationships to understanding true effect sizes, alternative approaches such as Bayesian statistics (e.g., Gelman et al., 2013; Vandekerckhove et al., 2018) can prove increasingly valuable for entrepreneurship research (Lohrke et al., 2018).

Visualizations. Entrepreneurial activities are often highly heterogeneous, uncertain, and subject to change over time. Accordingly, it can be very informative to include visualizations of data (Ertug et al., 2018; Greve, 2018; Healy, 2018; Levine, 2018). Graphs enable entrepreneurship scholars to more efficiently communicate the distributional properties of their data and provide more fine-grained insights into the uncertainty of observed effects and outcomes (Schwab, 2018). This means that in addition to standard presentations (e.g., interaction plots), creativity is encouraged to develop novel graphical approaches for evidence presentation (e.g., heat maps). This is particularly valuable and relevant when analyzing big data (Tay et al., 2018). Facilitated by new visualization tools in data science, such as ggplot2 for R, Matplotlib, Seaborn, Plotly, and Bokeh for Python, PROC SGPLOT for SAS and many others, there are increasing opportunities for creating novel insights through better visual presentations in quantitative entrepreneurship research.[8]

Statistical power. Statistical power refers to the likelihood that researchers correctly conclude that the null hypothesis (i.e., the correlation between X and Y is zero; $\rho = 0$) should be rejected (Aguinis & Vandenberg, 2014). Studies in entrepreneurship vary widely in their statistical power, which is determined by a combination of the true population effect, the study's sample size, analysis technique, and the chosen significance level (i.e., α). Few studies, however, explicitly report the statistical power, which is an important consideration in the design phase of research

(Hoenig & Heisey, 2001). Reporting the statistical power provides insight because power analysis shows how likely it is that a study's findings can be attributed to either very low or high statistical power. We therefore recommend that authors provide a power analysis and explain if and how the statistical power may have affected the study's conclusions.

Missing data. A pervasive problem faced by any researcher is missing data. Rarely, if ever, are all data fully available for all cases in a dataset. Missing data may occur because of survey nonresponse, changes in data reporting requirements, technical errors, the merging of partially non-overlapping databases, or random events. Missing data may severely affect research findings because it can easily lead to sample attrition and associated sample selection biases (Aguinis et al., 2018; Certo et al., 2016; Enders, 2010; Newman, 2014). For instance, focusing on a subset of cases with complete data (i.e., listwise deletion) can distort the results when these cases differ systematically from the larger population. Therefore, careful consideration of the magnitude of the issue and appropriate missing data treatment (e.g., multiple imputation [MI] or full information maximum likelihood [FIML]) are needed (Newman, 2014). To facilitate reproducibility, it is critical that authors explicitly report the extent of missing data in their manuscripts, along with explanations for how missing data were handled and may have affected the research findings (Aguinis et al., 2018).

Outliers and influential cases. Given the frequently skewed distributions in the data on entrepreneurial phenomena (Davidsson, 2016), it is not uncommon for datasets to have outliers, that is, observations that differ substantially from others. Outliers can have a large impact on the research findings, particularly when moderation or nonlinear relationships are tested, but they can also contain valuable information that contributes to forming new theoretical insights (Crawford et al., 2015). Some of the outliers may result from errors in records, but some might be accurate data points that lie at a distance from other data points. These are often very interesting for entrepreneurship research, which often focuses on such rare events as venture capital backing or IPOs of young firms (Davidsson, 2016).[9] However, not all outliers are equally interesting or influential, indicating that different types of outliers require different types of treatment (Aguinis et al., 2013). It is therefore critical that authors clearly report how they defined, identified, and handled outliers. Providing this information should help readers to better understand the distributional properties of the data and the extent to which research findings might be driven by a few influential observations.

Robustness analyses. A common challenge in quantitative entrepreneurship research is that in comparison to textbook examples in econometrics focusing on one main effect, it is more typical that there are multiple (potentially endogenous) independent variables with contingent effects, which requires the interaction effects to be analyzed as well. Therefore, many solutions developed in econometrics are not easily applicable or at least do not solve all the modeling problems simultaneously. Addressing this modeling uncertainty frequently creates a need to run additional alternative models as robustness tests when all the modeling assumptions cannot be satisfied with full certainty in one model (Anderson, 2019). Running multiple alternative models requires both careful consideration of the priority of the concerns (e.g., what is the theoretically best justified main model) as well as consideration of how to interpret the findings from multiple incomplete models. As in other research design choices, the choice of the main model and the robustness models should be based on careful consideration and justification based on what is most appropriate for testing the hypotheses and addressing the focal research problem. We also recommend that authors report their robustness analyses transparently (in an online appendix if more space is needed), as it facilitates replication and allows editors and reviewers to better grasp the patterns in the data.

Transparency and attention to detail. Overall, there is a need for more transparent descriptions of the choices made during the empirical analyses so that readers can better assess the quality of the research (Aguinis et al., 2018). When entrepreneurship scholars embrace greater transparency in their reporting practices, this facilitates better replication and faster knowledge accumulation in the field (Bettis et al., 2016; Shrout & Rodgers, 2018). Accordingly, there is a premium for conscientiousness when authors carefully craft the methods and the results sections of their papers by clearly documenting their empirical efforts, starting from the major choices and continuing down to the smallest details. For instance, if there are obvious typos or other small errors, it is difficult for a reader to believe that the complex analyses have been executed without flaws. Complete and careful reporting of methodological choices is consequently critical for establishing the credibility of the research (Zhang & Shaw, 2012). Sometimes the page limits of the articles do not allow sufficiently detailed descriptions of the methods and robustness analyses. In such situations, online appendices can be a good approach to offer enough detail.[10]

Workflow and Data Management

Addressing the above-discussed issues in quantitative empirical research would be easier if the data were simple and the analyses proceeded linearly without errors, thus leading to final results after one run. In practice, that is probably never the case. For instance, reviewers often come up with one or more additional required control variables, requiring all the steps in the research process to be repeated, which, among other reasons, makes it important for the research to be reproducible.

Rigorous empirical analyses in entrepreneurship research tend to involve combinations of many data sources, derivation of variables from raw data, and the creation of multiple tables and plots of empirical tests (and various diagnostics related to them). When implemented in sophisticated statistical software, such as Stata or R, without careful planning and documentation, the resulting code easily becomes complex and difficult for outsiders to understand, as with almost any software project. Additionally, similar to software programming, errors can occur easily (Wilson et al., 2014). For instance, in empirical projects, this is sometimes manifested in descriptive statistics that do not make sense for an observant reviewer who knows the focal phenomenon, data, and prior research. Therefore, a careful data management workflow and documentation are important to eliminate potential bugs and other problems (Long, 2009; Mitchell, 2010). Two pairs of eyes usually spot more errors than one, so it is advisable to share the code at least among the author team and cross-check the code and analyses to prevent errors from being missed in the final analyses.[11]

Although data management has generally not been a major topic in current PhD programs, its importance is increasing as the volume and complexity of data grows (Braun et al., 2018; Christen, 2019; Schwab & Zhang, 2019). Data management plans and study preregistrations are also increasingly required by research funding agencies to ensure proper treatment of data, as well as to facilitate open science (Mellor & Nosek, 2018; Nosek et al., 2015). For instance, in economics, many journals are requiring or encouraging the data and analysis codes to be submitted to facilitate replication (Christensen & Miguel, 2018).[12] Although it is currently not a requirement at ETP that the data and analysis codes are published, publishing them can be valuable and certainly facilitates replication and the progress in the field (see also, Ethiraj et al., 2017 for argumentation of the benefits for voluntary disclosure of data for replication). At present, analysis codes and data can be made available for replication as online appendices.

Conclusion

As in other areas of research, expectations for rigor in quantitative empirical entrepreneurship research are growing rapidly, not the least because of the across-the-fields recognition of the fallibility of research findings (e.g., Baker, 2016; Camerer et al., 2016; Gelman, 2018; Honig et al., 2018; Miguel et al., 2014; Open Science Collaboration, 2015), insufficient replication and replicability (Bettis et al., 2016), and the growing maturity of the field. Entrepreneurship research, as a multidisciplinary field, has come a long way in establishing its legitimacy, but could still learn from other disciplines that have recently sought to improve their research practices. In this editorial, we have thus sought to push the rigor in our field forward by reviewing the methodological best practices for designing, conducting, and reporting rigorous quantitative entrepreneurship research.

Table 1 offers a synthesis of our seven key recommendations. While setting very strict requirements would risk creating a methodological straitjacket that would not work for some (e.g., novel) areas of entrepreneurship research, the reviewed approaches do offer many opportunities for quantitative entrepreneurship scholars to enhance rigor and transparency. We realize that submitting authors might fear that following our recommendations may lead to excessively long manuscripts and actually increase their chances of rejection, because it will alert editors and reviewers to the methodological limitations of their studies. Here, we would argue that the opposite is true. ETP is indeed strongly committed to enhancing rigor and transparency in the field, as reflected by the recommendations in this editorial and related initiatives, such as the opportunity for authors to publish data files and supplementary materials in online appendices.

Still, an important role of this editorial is to educate reviewers and editors on what is realistic to expect and how critical empirical concerns can be solved. This should facilitate constructive and insightful methodological comments instead of repeating methodological myths or rejecting papers on the basis of unrealistic requirements concerning, for instance, endogeneity. It needs to be remembered that there is no such thing as a perfect empirical study that could "confirm" a hypothesis. Instead, empirical evidence could at best be used to reject a hypothesis (Kuhn, 1962; Popper, 1959; Shadish et al., 2002), and even that is not very realistic with the many assumptions related to statistical analyses (e.g., Greenland et al., 2016). Altogether this makes a thoughtful interpretation, transparency, and replication very important (e.g., Amrhein et al., 2019; Nosek et al., 2015; Shaver, 2019). When tackling new questions of interest in quantitative entrepreneurship research, we need to be realistic, aim for the theoretically best solutions given the limitations and the current understanding of the key concerns, and be transparent to enable reproducibility, learning, and the progress of the field. Causal identification is best achieved by a cumulative body of empirical research with replications focusing not only on the generalizability of the existing findings but also on advancing better-identified tests (Shaver, 2019). As authors, reviewers, and editors of quantitative entrepreneurship research, we also need to continue investing in our methodological learning given the current rapid progress of quantitative methods, which is likely to lead to new knowledge and significant further developments in many areas discussed in this editorial.

For instance, while the American Statistical Association (ASA) had already previously considered but decided not to recommend banning the use of the term "statistically significant," that step has now been taken (Wasserstein et al., 2019). The use of phrases such as "significant" or "nonsignificant" based on dichotomized p values are now strongly discouraged (Hurlbert et al., 2019; Wasserstein et al., 2019). While there is no agreement on a simple replacement for the expired "$p < .05$" paradigm, and any one-size-fits-all approach to statistical inference is not an appropriate expectation, there are many clear recommendations offered in the 43 articles of the recent major special issue of the *American Statistician* and summarized by Wasserstein et al.

(2019) in two sentences, as follows: "Accept uncertainty. Be thoughtful, open, and modest." Remember "ATOM." In essence, statistics cannot turn uncertainty into certainty, so we need to accept uncertainty (e.g., visualize the uncertainty in the effect sizes), be thoughtful in our research design, statistical analysis, and interpretation of the results (e.g., be clear whether we are conducting exploratory versus confirmatory research), be open (e.g., embracing open science practices), and be modest (e.g., understand the limits of our work).

In closing, we hope that our broad review of the key empirical concerns and approaches in quantitative entrepreneurship research and of the recent advances in methodological approaches stimulate entrepreneurship scholars to continuously learn and update their empirical toolboxes, enabling enhanced rigor and faster knowledge accumulation in our field.

Acknowledgments

The authors gratefully acknowledge insightful comments from Rene Bakker, Ari Hyytinen, Mikko Ketokivi, Juhana Peltonen, Mikko Rönkkö, Andreas Schwab, Diemo Urbig, Arjen van Witteloostuijn, Johan Wiklund, an anonymous reviewer, and four senior editors of ETP.

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 iDs

Markku Maula https://orcid.org/0000-0002-7030-5059
Wouter Stam https://orcid.org/0000-0003-2821-2163

Notes

1. These issues are discussed in more detail in the section on statistical significance on page 26.
2. In addition to quantitative deductive or hypothesis-testing research, which forms the majority of the current quantitative entrepreneurship research, inductive or exploratory quantitative research is also on the rise, largely driven by the growth of big data and related methods (Schwab & Zhang, 2019; Wennberg & Anderson, 2019).
3. The experimental research designs used in entrepreneurship research can be classified as within-subject vs. between-subject designs, the latter of which can either be controlled experiments (random assignment of subjects) or quasi-experiments (non-random assignment of subjects; Grant & Wall, 2009; Hsu et al., 2017).
4. For instance, a logarithmic transformation to address the skewness of a dependent variable is rarely an ideal solution (there are usually other better approaches, and the logarithmic transformation should be chosen based on theoretical reasons related to the functional form, not empirical reasons related to the distribution of the variables; (e.g., Wooldridge, 2015).
5. For instance, in Stata, the marginscontplot module (Royston, 2013) and the subsequent marginscontplot2 module facilitate the plotting of marginal effects on the original scale of the transformed independent variables.
6. However, in modern approaches to causal inference, the assumption about the temporal dependence may not always be needed (see, e.g., Imbens & Rubin, 2015; Pearl, 2009, p. 57-59; Wunsch et al., 2010). Regarding quasi-experimental techniques, if regression discontinuity design (RDD) or

instrumental variable (IV) regression or a structural model can be applied appropriately, cross-sectional data can also offer informative results.

7. We thank an autonomous reviewer for highlighting an important further clarification that the *p* value is a conditional probability calculated assuming that the null hypothesis is true. If the statistical significance test leads to the rejection of the null hypothesis, this also implies that the *p* value can serve no additional purpose because the assumption for its calculation was ruled out. The resulting dilemma is that we do not have a probability statement for the hypothesis of interest (i.e., the effect of interest). This means that without additional analyses, we cannot make probability statements or quantified uncertainty evaluations for the effect of interest. For this reason, it is important to consider other options to obtain this information and to evaluate the uncertainty associated with the hypothesized effects (e.g., graphs of effect distributions and Bayesian statistics), as discussed elsewhere in this paper.
8. For instance, *Administrative Science Quarterly* has recently collected and made available resources for visualization in organizational research: https://asqblog.com/2017/11/29/asq-improving-evidence-presentation-resources-and-tools/ Also, DIME, the impact evaluation unit of the World Bank Research Group, has made available a helpful visual library with the associated Stata code: https://worldbank.github.io/Stata-IE-Visual-Library/
9. For studies seeking to explain extreme performance or underperformance instead of mean performance, quantile regression (e.g., Koenker & Hallock, 2001) may be a useful method instead of or in addition to OLS regression (e.g., Golubov et al., 2015; Hamilton, 2000).
10. The ETP author pages offer instructions on online appendices, including data appendices.
11. Wilson et al. (2014) review best practices for scientific computing, which are valuable especially in research projects involving more complex data wrangling and analyses.
12. For instance, the *American Economic Association* has an explicit policy in their journals (e.g., *American Economic Review*) "…to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication," unless an exception is requested and granted, for instance, because of proprietary or legally restricted data (American Economic Association, 2018).

References

Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, *10*(1), 465–503.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495–510.

Acs, Z., Audretsch, D., Desai, S., & Welpe, I. (2010). On experiments in entrepreneurship research. *Journal of Economic Behavior & Organization*, *76*(1), 1–2.

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351–371.

Aguinis, H., & Culpepper, S. A. (2015). An expanded decision-making procedure for examining cross-level interaction effects with multilevel modeling. *Organizational Research Methods*, *18*(2), 155–176.

Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods*, *20*(4), 665–685.

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, *39*(6), 1490–1528.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301.

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*(1), 83–110.

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*(1), 569–595.

Aguinis, H., Werner, S., Lanza Abbott, J., Angert, C., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, *13*(3), 515–539.

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economic Letters*, *80*(1), 123–129.

Allison, P. D. (2014). *Event history and survival analysis: Regression for longitudinal event data*. SAGE Publications.

American Economic Association. (2018). Data availability policy. Retrieved October 11, 2018, from https://www.aeaweb.org/journals/policies/data-availability-policy.

Amrhein, V., Greenland, S., & McShane, B. (2019). Retire statistical significance. *Nature*, *567*, 305–307.

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, *73*(supp 1), 262–270.

Anderson, A. A. (2019). Assessing statistical results: Magnitude, precision, and model uncertainty. *The American Statistician*, *73*(supp 1), 118–121.

Anderson, B. S., Wennberg, K., & McMullen, J. S. (2019). Editorial: Enhancing quantitative theory-testing entrepreneurship research. *Journal of Business Venturing*, *34*(5), 105928.

Andersson, U., Cuervo-Cazurra, A., & Nielsen, B. B. (2014). From the editors: Explaining interaction effects within and across levels of analysis. *Journal of International Business Studies*, *45*(9), 1063–1071.

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, *41*(2), 521–543.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. https://press.princeton.edu/books/paperback/9780691120355/mostly-harmless-econometrics

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *Leadership Quarterly*, *28*(1), 5–21.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120.

Baker, M. (2016). Is there a reproducibility crisis? A nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help. *Nature*, *533*(7604), 452–455.

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, *33*(4), 445–459.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, *6*(3), 285–327.

Becker, T. E., Robertson, M. M., & Vandenberg, R. J. (2018). Nonlinear transformations in organizational research: Possible problems and potential solutions. *Organizational Research Methods*, In press.

Benjamin, D. J., & Berger, J. O. (2019). Three recommendations for improving the use of p -values. *The American Statistician*, *73*(sup1), 186–191.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*(1), 289–300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188.

Bergh, D. D., & Holbein, G. F. (1997). Assessment and redirection of longitudinal analysis demonstration with a study of the diversification and divestiture relationship. *Strategic Management Journal*, *18*(7), 557–571.

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*(3), 423–436.

Betensky, R. A. (2019). The *p* -value requires context, not a threshold. *The American Statistician*, *73*(supp 1), 115–117.

Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. (2016). Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*, *37*(2), 257–261.

Bettis, R., Gambardella, A., Helfat, C., & Mitchell, W. (2014). Quantitative empirical analysis in strategic management. *Strategic Management Journal*, *35*(7), 949–953.

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. *Strategic Management Journal*, *37*(11), 2193–2203.

Bliese, P. D., Schepker, D. J., Essman, S. M., & Ployhart, R. E. (2019). Bridging methodological divides between macro- and microresearch: Endogeneity and methods for panel data. *Journal of Management*, *38*, 014920631986801.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. John Wiley & Sons.

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301–328). Springer Netherlands.

Bond, S. R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Economic Journal*, *1*(2), 141–162.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071.

Boyd, B. K., Takacs Haynes, K., Hitt, M. A., Bergh, D. D., & Ketchen, D. J. (2012). Contingency hypotheses in strategic management research: Use, disuse, or misuse? *Journal of Management*, *38*(1), 278–313.

Braun, M. T., Kuljanin, G., & DeShon, R. P. (2018). Special considerations for the acquisition and wrangling of big data. *Organizational Research Methods*, *21*(3), 633–659.

Brown, G. W., Ouimet, P. P., Robinson, D. T., & Zoller, T. (2017). Understanding entrepreneurship: Facilitating academic research with a shared data repository. UNC & Duke University Working Paper.

Bun, M. J. G., & Sarafidis, V. (2015). Dynamic panel data models. In B. H. Baltagi (Ed.), *Oxford handbook on panel data* (pp. 76–110). Oxford University Press.

Busenitz, L. W., West III, G. P., Shepherd, D., Nelson, T., Chandler, G. N., & Zacharakis, A. (2003). Entrepreneurship research in emergence: Past trends and future directions. *Journal of Management*, *29*(3), 285–308.

Calin-Jageman, R. J., & Cumming, G. (2019). The new statistics for better science: Ask how much, how uncertain, and what else is known. *The American Statistician*, *73*(sup1), 271–280.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. *The Review of Economics and Statistics*, *101*(3), 442–451.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, *14*(4), 909–946.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. -H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer,

T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436.
Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.
Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*(1), 17–32.
Certo, S. T., Busenbark, J. R., Woo, H. -S., & Semadeni, M. (2016). Sample selection bias and heckman models in strategic management research. *Strategic Management Journal*, *37*(13), 2639–2657.
Certo, S. T., Withers, M. C., & Semadeni, M. (2017). A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal*, *38*(7), 1536–1556.
Chandler, G. N., & Lyon, D. W. (2001). Issues of research design and construct measurement in entrepreneurship research: The past decade. *Entrepreneurship Theory and Practice*, *25*(4), 101–113.
Chatterji, A. K., Findley, M., Jensen, N. M., Meier, S., & Nielson, D. (2016). Field experiments in strategy research. *Strategic Management Journal*, *37*(1), 116–132.
Chen, G. (2018). Supporting and enhancing scientific rigor. *Journal of Applied Psychology*, *103*(4), 359–361.
Chen, X. P., Yao, X., & Kotha, S. (2009). Entrepreneur passion and preparedness in business plan presentations: A persuasion analysis of venture capitalists' funding decisions. *Academy of Management Journal*, *52*(1), 199–214.
Christen, P. (2019). Data linkage: The big picture. *Harvard Data Science Review*, *1*(2).
Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, *56*(3), 920–980.
Ciuchta, M. P., Letwin, C., Stevenson, R., McMahon, S., & Huvaj, M. N. (2018). Betting on the coachable entrepreneur: Signaling and social exchange in entrepreneurial pitches. *Entrepreneurship Theory and Practice*, *42*(6), 860–885.
Cleves, M. A., Gould, W. W., & Marchenko, Y. V. (2016). *An introduction to survival analysis using Stata* (3rd ed.). Stata Press.
Clougherty, J. A., Duso, T., & Muck, J. (2016). Correcting for self-selection based endogeneity in management research: Review, recommendations and simulations. *Organizational Research Methods*, *19*(2), 286–347.
Combs, J. G. (2010). Big samples and small effects: Let's not trade relevance and rigor for power. *Academy of Management Journal*, *53*(1), 9–13.
Connelly, B. L., Ireland, R. D., Reutzel, C. R., & Coombs, J. E. (2010). The power and effects of entrepreneurship research. *Entrepreneurship Theory and Practice*, *34*(1), 131–149.
Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*, 1–12.
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*(5), 696–713.
Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage.
Crook, T. R., Shook, C. L., Morris, M. L., & Madden, T. M. (2010). Are we there yet?: An assessment of research design and construct measurement practices in entrepreneurship research. *Organizational Research Methods*, *13*(1), 192–206.
Cumming, G. (2011). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* Routledge.
Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29.

Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, *15*(3), 339–362.

Dalle, J. -M., Den Besten, M., & Menon, C. (2017). *Using Crunchbase for economic and managerial research*. OECD Publishing.

Davidsson, P. (2016). *Researching entrepreneurship: Conceptualization and design* (2nd ed.). Springer.

Davidsson, P., & Wiklund, J. (2001). Levels of analysis in entrepreneurship research: Current research practice and suggestions for the future. *Entrepreneurship Theory and Practice*, *25*(4), 81–100.

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*(1), 1–19.

Dean, M. A., Shook, C. L., & Payne, G. T. (2007). The past, present, and future of entrepreneurship research: Data analytic trends and training. *Entrepreneurship Theory and Practice*, *31*(4), 601–618.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). Sage publications.

Dieleman, J. L., & Templin, T. (2014). Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: A simulation study. *Plos One*, *9*(10), e110257.

Edwards, J. R. (2003). Construct validation in organizational behavior research. In J. Greenberg (Ed.), *Organizational behavior: A management challenge* (pp. 327–371). Erlbaum.

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, *13*(4), 668–689.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*(1), 1–22.

Enders, C. K. (2010). *Applied missing data analysis*. The Guilford Press.

Ertug, G., Gruber, M., Nyberg, A., & Steensma, H. K. (2018). From the editors—A brief primer on data visualization opportunities in management research. *Academy of Management Journal*, *61*(5), 1613–1625.

Ethiraj, S. K., Gambardella, A., & Helfat, C. E. (2017). Improving data availability: A new SMJ initiative. *Strategic Management Journal*, *38*(11), 2145–2146.

Farhat, J., Matusik, S., Robb, A., & Robinson, D. T. (2018). New directions in entrepreneurship research with the Kauffman firm survey. *Small Business Economics*, *50*(3), 521–532.

Farhat, J., & Robb, A. (2018). Analyzing complex survey data: The Kauffman firm survey. *Small Business Economics*, *50*(3), 657–670.

Freese, J. (2007). Replication standards for quantitative social science: Why not sociology? *Sociological Methods and Research*, *36*(2), 153–172.

Gelman, A. (2018). The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. *Personality and Social Psychology Bulletin*, *44*(1), 16–23.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321–326.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, *59*(5), 1493–1507.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606.

Gigerenzer, G., & Marewski, J. N. (2015). Surrogate science: The idol of a universal method for scientific inference. *Journal of Management*, *41*(2), 421–440.

Goetz, C., Hyatt, H., McEntarfer, E., & Sandusky, K. (2015). The promise and potential of linked employer-employee data for entrepreneurship research.. Working Paper No. 21639.

Golubov, A., Yawson, A., & Zhang, H. (2015). Extraordinary acquirers. *Journal of Financial Economics*, *116*(2), 314–330.

Grant, A. M., & Wall, T. D. (2009). The neglected science and art of quasi-experimentation. *Organizational Research Methods*, *12*(4), 653–686.

Greene, W. H. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters*, *107*(2), 291–296.

Greene, W. H. (2012). *Econometric analysis* (International ed.). Pearson Education Limited.

Greenland, S. (2019). Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolution with *s-values*. *The American Statistician*, *73*(sup1), 106–114.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350.

Grégoire, D. A., Binder, J. K., & Rauch, A. (2019). Navigating the validity tradeoffs of entrepreneurship research experiments: A systematic review and best-practice suggestions. *Journal of Business Venturing*, *34*(2), 284–310.

Greve, H. R. (2018). Show me the data! improving evidence presentation for publication. *Management and Organization Review*, *14*(2), 423–432.

Guide, V. D. R., & Ketokivi, M. (2015). Notes from the editors: Redefining some methodological criteria for the journal. *Journal of Operations Management*, *37*(1), v–0.

Haans, R. F. J., Pieters, C., & He, Z. -L. (2016). Thinking about u: Theorizing and testing u- and inverted u-shaped relationships in strategy research. *Strategic Management Journal*, *37*(7), 1177–1195.

Hahn, E. D., & Ang, S. H. (2017). From the editors: New directions in the reporting of statistical results in the Journal of world business. *Journal of World Business*, *52*(2), 125–126.

Hamilton, B. H. (2000). Does entrepreneurship pay? an empirical analysis of the returns to self-employment. *Journal of Political Economy*, *108*(3), 604–631.

Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic Organization*, *1*(1), 51–78.

Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, *13*(2), 586–632.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Healy, K. (2018). *Data visualization: A practical introduction*. Princeton University Press.

Hicks, R., & Tingley, D. (2011). Causal mediation analysis. *The Stata Journal*, *11*(4), 605–619.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician*, *55*(1), 19–24.

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, *28*(4), 331–343.

Honig, B., Lampel, J., Baum, J. A. C., Glynn, M. A., Jing, R., Lounsbury, M., Schüßler, E., Sirmon, D. G., Tsui, A. S., Walsh, J. P., & van Witteloostuijn, A. (2018). Reflections on scientific misconduct in management: Unfortunate incidents or a normative crisis? *Academy of Management Perspectives*, *32*(4), 412–442.

Hsu, D. K., Simmons, S. A., & Wieland, A. M. (2017). Designing entrepreneurship experiments: A review, typology, and research agenda. *Organizational Research Methods*, *20*(3), 379–412.

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845.

Hubbard, R. (2015). *Corrupt research: The case for reconceptualizing empirical management and social science*. Sage Publications.

Hurlbert, S. H., Levine, R. A., & Utts, J. (2019). Coup de grâce for a tough old bull: "Statistically significant" expires. *The American Statistician*, *73*(sup1), 352–357.

Iacus, S. M., King, G., & Porro, G. (2009). CEM: Software for coarsened exact matching. *Journal of Statistical Software*, *30*(9), 1–27.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24.

Imbens, G. W. (2019). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. NBER Working Paper No. 26104.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, *24*(4), 602–611.

Keele, L., Tingley, D., & Yamamoto, T. (2015). Identifying mechanisms behind policy interventions via causal mediation analysis. *Journal of Policy Analysis and Management*, *34*(4), 937–963.

Kenny, D. A. (1979). *Correlation and causality*. Wiley.

Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, *52*(1), 1–14.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*(4), 435–454.

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, *44*(2), 347–361.

Kirkman, B. L., & Chen, G. (2011). Maximizing your data or data slicing? recommendations for managing multiple submissions from the same dataset. *Management and Organization Review*, *7*(3), 433–446.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*(4), 143–156.

Kolkman, D., & van Witteloostuijn, A. (2019). Data science in strategy: Machine learning and text analysis in the study of firm growth. Tinbergen Institute Discussion Paper TI 2019-066/VI.

Kriauciunas, A., Parmigiani, A., & Rivera-Santos, M. (2011). Leaving our comfort zone: Integrating established practices with unique adaptations to conduct survey-based strategy research in nontraditional contexts. *Strategic Management Journal*, *32*(9), 994–1010.

Krueger, J. I., & Heck, P. R. (2019). Putting the p-value in its place. *The American Statistician*, *73*(sup1), 122–128.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago.

Lahti, T., Halko, M. -L., Karagozoglu, N., & Wincent, J. (2019). Why and how do founding entrepreneurs bond with their ventures? neural correlates of entrepreneurial and parental bonding. *Journal of Business Venturing*, *34*(2), 368–388.

Laitin, D. D., & Reich, R. (2017). Trust, transparency, and replication in political science. *Political Science and Politics*, *50*(1), 172–175.

Lance, C. E., & Vandenberg, R. J. (Eds.). (2009). *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*: Routledge.

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods*, *21*(4), 475–492.

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration–exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal*, *36*(3), 319–338.

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, *4*(3), 165–224.

Lee, J., Hwang, B., & Che, H. (2017). Are founder CEOs more overconfident than professional CEOs? Evidence from S&P 1500 companies. *Strategic Management Journal*, *38*, 751–769.

Levie, J., Autio, E., Acs, Z., & Hart, M. (2014). Global entrepreneurship and institutions: An introduction. *Small Business Economics*, *42*(3), 437–444.

Levine, S. S. (2018). Show us your data: Connect the dots, improve science. *Management and Organization Review*, *14*(2), 433–437.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1–18.

Lewin, A. Y., Chiu, C. -Y., Fey, C. F., Levine, S. S., McDermott, G., Murmann, J. P., & Tsang, E. (2016). The critique of empirical social science: New policies at management *and organization review*. *Management and Organization Review*, *12*(4), 649–658.

Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, *26*(12), 1827–1832.

Lohrke, F. T., Carson, C. M., & Lockamy, A. (2018). Bayesian analysis in entrepreneurship decision-making research: Review and future directions. *Management Decision*, *56*(5), 972–986.

Long, J. S. (2009). *The workflow of data analysis using Stata*. Stata Press.

Low, H., & Meghir, C. (2017). The use of structural models in econometrics. *Journal of Economic Perspectives*, *31*(2), 33–58.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Erlbaum.

Mathieu, J. E., & Chen, G. (2011). The etiology of the multilevel paradigm in management research. *Journal of Management*, *37*(2), 610–641.

Matthews, R. A. J. (2019). Moving towards the post $p<0.05$ era via the analysis of credibility. *The American Statistician*, *73*(sup1), 202–212.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*(1), 577–605.

McCann, B. T., & Schwab, A. (2019). Bayesian analysis in strategic management research: Time to update your priors. *Strategic Management Review*, Forthcoming.

McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. (2018). What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management*, *44*(7), 2909–2933.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, *73*(sup1), 235–245.

Mellor, D. T., & Nosek, B. A. (2018). Easy preregistration will benefit any research. *Nature Human Behaviour*, *2*(2), 98–98.

Meyer, K. E., van Witteloostuijn, A., & Beugelsdijk, S. (2017). What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies*, *48*(5), 535–551.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B. A., Petersen, M., Sedlmayr, R., Simmons, J. P., Simonsohn, U., & Van der Laan, M. (2014). Social science. promoting transparency in social science research. *Science*, *343*(6166), 30–31.

Mitchell, M. N. (2010). *Data management using stata: A practical handbook*. Stata Press.

Mize, T. D. (2019). Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science*, *6*, 81–117.

Molina-Azorin, J. F., Bergh, D. D., Corley, K. G., & Ketchen, D. J. (2017). Mixed methods in the organizational sciences. *Organizational Research Methods*, *20*(2), 179–192.

Molina-Azorín, J. F., López-Gamero, M. D., Pereira-Moliner, J., & Pertusa-Ortega, E. M. (2012). Mixed methods studies in entrepreneurship research: Applications and contributions. *Entrepreneurship & Regional Development*, *24*(5-6), 425–456.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge University Press.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69–85.

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, *17*(4), 372–411.

Nofal, A., Nicolaou, N., & Symeonidou, N. (2017). Biology and entrepreneurship. In G. Ahmetoglu, T. Chamorro-Premuzic, B. Klinger, & T. Karcisky (Eds.), *The Wiley handbook of entrepreneurship* (pp. 259–272). Chichester, UK: John Wiley & Sons Ltd.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., & Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.

Obschonka, M., Fisch, C., & Boyd, R. (2017). Using digital footprints in entrepreneurship research: A Twitter-based personality analysis of superstar entrepreneurs and managers. *Journal of Business Venturing Insights*, *8*, 13–23.

Open Science Collaboration (2015). Psychology. estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *4*(1), 669–688.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Pearl, J. (2012). The causal foundations of structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 68–91). Guilford Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*(1), 94–120.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*(1), 539–569.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. *Organizational Research Methods*, *19*(2), 159–203.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. -Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Popper, K. R. (1959). *The logic of scientific discovery*. Basic Books.

Prüfer, J., & Prüfer, P. (2019). Data science for entrepreneurship research: Studying demand dynamics for entrepreneurial skills in the Netherlands. *Small Business Economics*, *54*(2).

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using stata*. Stata Press.

Reynolds, P. D. (2007). New firm creation in the United States a PSED I overview. *Foundations and Trends® in Entrepreneurship*, *3*(1), 1–150.

Reynolds, P. D., Bosma, N., Autio, E., Hunt, S., De Bono, N., Servais, I., Lopez-Garcia, P., & Chin, N. (2005). Global entrepreneurship monitor: Data collection design and implementation 1998–2003. *Small Business Economics*, *24*(3), 205–231.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, *12*(4), 762–800.

Roberts, M. R., & Whited, T. M. (2013). Endogeneity in empirical corporate finance. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the economics of finance* (Vol. 2, pp. 493–572). Elsevier.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42.

Royston, P. (2013). Marginscontplot: Plotting the marginal effects of continuous predictors. *Stata Journal*, *13*(3), 510–527.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322–331.

Schunck, R., & Perales, F. (2017). Within- and between-cluster effects in generalized linear mixed models: A discussion of approaches and the xthybrid command. *Stata Journal*, *17*(1), 89–115.

Schwab, A. (2015). Why all researchers should report effect sizes and their confidence intervals: Paving the way for meta–analysis and evidence–based management practices. *Entrepreneurship Theory and Practice*, *39*(4), 719–725.

Schwab, A. (2018). Investigating and communicating the uncertainty of effects: The power of graphs. *Entrepreneurship Theory and Practice*, *42*(6), 823–834.

Schwab, A., & Zhang, Z. (2019). A new methodological frontier in entrepreneurship research: Big data studies. *Entrepreneurship Theory and Practice*, *43*(5), 843–854.

Schwab, A., Abrahamson, E., Starbuck, W. H., & Fidler, F. (2011). Researchers should make thoughtful assessments instead of null-hypothesis significance tests. *Organization Science*, *22*(4), 1105–1120.

Semadeni, M., Withers, M. C., & Certo, T. S. (2014). The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal*, *35*(7), 1070–1079.

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

Shaver, J. M. (2019). Interpreting interactions in linear fixed-effect regression models: When fixed-effect estimates are no longer within-effects. *Strategy Science*, *4*(1), 25–40.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods*, *13*(2), 320–347.

Short, J. C., Ketchen, D. J., Combs, J. G., & Ireland, R. D. (2010). Research methods in entrepreneurship: Opportunities and challenges. *Organizational Research Methods*, *13*(1), 6–15.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*(1), 487–510.

Singleton, R., & Straits, B. C. (2017). *Approaches to social research* (6th ed.). Oxford University Press.

Spector, P. E., & Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, *14*(2), 287–305.

Starbuck, W. H. (2016). 60th Anniversary essay: How journals could improve research practices in social science. *Administrative Science Quarterly*, *61*(2), 165–183.

Stata. (2018). *Stata extended regression models reference manual release 15*. Stata Press.

Stevenson, R. M., & Josefy, M. (2019). Knocking at the gate: The path to publication for entrepreneurship experiments through the lens of gatekeeping theory. *Journal of Business Venturing*, *34*(2), 242–260.

Tarasconi, G., & Menon, C. (2017). *Matching crunchbase with patent data. OECD Science, Technology and Industry* (*Working Papers, No. 2017/07*). OECD Publishing.

Tay, L., Ng, V., Malik, A., Zhang, J., Chae, J., Ebert, D. S., Ding, Y., Zhao, J., & Kern, M. (2018). Big data visualizations in organizational science. *Organizational Research Methods*, *21*(3), 660–688.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38.

Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational Research Methods*, *21*(3), 525–547.

Toraldo, M. L., Islam, G., & Mangia, G. (2018). Modes of knowing: Video research and the problem of elusive knowledges. *Organizational Research Methods*, *21*(2), 438–465.

Turner, S. F., Cardinal, L. B., & Burton, R. M. (2017). Research design for mixed methods. *Organizational Research Methods*, *20*(2), 243–267.

Van de Ven, A. H., & Johnson, P. E. (2006). Knowledge for theory and practice. *Academy of Management Review*, *31*(4), 802–821.

Vandekerckhove, J., Rouder, J. N., & Kruschke, J. K. (2018). Editorial: Bayesian methods for advancing psychological science. *Psychonomic Bulletin & Review*, *25*(1), 1–4.

Vermeulen, F. (2005). On rigor and relevance: Fostering dialectic progress in management research. *Academy of Management Journal*, *48*(6), 978–982.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p < 0.05". *The American Statistician*, *73*(sup1), 1–19.

Wennberg, K. (2005). Entrepreneurship research through databases: Measurement and design issues. *New England Journal of Entrepreneurship*, *8*(2), 9–19.

Wennberg, K., & Anderson, B. S. (2019). Editorial: Enhancing the exploration and communication of quantitative entrepreneurship research. *Journal of Business Venturing*, .

Wenzel, R., & Van Quaquebeke, N. (2018). The double-edged sword of big data in organizational and management research: A review of opportunities and risks. *Organizational Research Methods*, *21*(3), 548–591.

Wiersema, M. F., & Bowen, H. P. (2009). The use of limited dependent variable techniques in strategy research: Issues and methods. *Strategic Management Journal*, *30*(6), 679–692.

Wiklund, J., Davidsson, P., Audretsch, D. B., & Karlsson, C. (2011). The future of entrepreneurship research. *Entrepreneurship Theory and Practice*, *35*(1), 1–9.

Williams, D. W., Wood, M. S., Mitchell, J. R., & Urbig, D. (2019). Applying experimental methods to advance entrepreneurship research: On the need for and publication of experiments. *Journal of Business Venturing*, *34*(2), 215–223.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal*, *12*(2), 308–331.

Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., & Wilson, P. (2014). Best practices for scientific computing. *PLoS Biology*, *12*(1), e1001745.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT press.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage Education.

Wulff, J. N. (2015). Interpreting results from the multinomial logit model: Demonstrated by foreign market entry. *Organizational Research Methods*, *18*(2), 300–325.

Wunsch, G., Russo, F., & Mouchart, M. (2010). Do we necessarily need longitudinal data to infer causal relations? *Bulletin of Sociological Methodology*, *106*(1), 5–18.

Zelner, B. A. (2009). Using simulation to interpret results from logit, probit, and other nonlinear models. *Strategic Management Journal*, *30*(12), 1335–1348.

Zhang, Y. A. N., & Shaw, J. D. (2012). Publishing in AMJ—Part 5: Crafting the methods and results. *Academy of Management Journal*, *55*(1), 8–12.

Zundel, M., MacIntosh, R., & Mackay, D. (2018). The utility of video diaries for organizational research. *Organizational Research Methods*, *21*(2), 386–411.

Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management*, *41*(2), 390–420.

Author Biographies

Markku Maula is professor of Venture Capital, Head of the Institute of Strategy and Venturing, and Deputy Head (Research) of the Department of Industrial Engineering and Management, Aalto University, Finland. His research interests include venture capital, corporate venturing, mergers and acquisitions, and, more generally, the intersections of strategy, entrepreneurship, international business, innovation, and corporate finance. His research has been published in journals including *Academy of Management Journal, Entrepreneurship Theory & Practice, Journal of Business Venturing, MIS Quarterly, Organization Science, Research Policy, Strategic Entrepreneurship Journal, and Strategic Management Journal*. He serves as Editor of *Entrepreneurship Theory & Practice*.

Wouter Stam is professor of Entrepreneurship at Vrije Universiteit Amsterdam. He was previously a Faculty Member at Hong Kong University of Science and Technology and a Visiting Scholar at the Wharton School, University of Pennsylvania. His research interests include new venture creation, corporate entrepreneurship, and entrepreneurial action in nascent markets.