Journal of Business Venturing 34 (2019) 105928

Contents lists available at ScienceDirect

Journal of Business Venturing

journal homepage: www.elsevier.com/locate/jbusvent

Editorial: Enhancing quantitative theory-testing entrepreneurship research

Brian S. Anderson[a,b,*], Karl Wennberg[c], Jeffery S. McMullen[d]

[a] Bloch School of Management, University of Missouri-, Kansas City, USA
[b] Faculty of Economics and Business Administration, Ghent University, Belgium
[c] Institute for Analytical Sociology (IAS), Linköping University, Norrköping, Sweden
[d] Kelley School of Business, Indiana University

ARTICLE INFO

Keywords:
Research design
Publishing in *JBV*
Theory-testing research
Causal inference
Researcher degrees of freedom

ABSTRACT

The purpose of this editorial is to discuss methodological advancements to enhance quantitative theory-testing entrepreneurship research. As the impact of entrepreneurship scholarship accelerates and deepens, our methods must keep pace to continue shaping theory, policy, and practice. Like our sister fields in business, entrepreneurship is coming to terms with the replication and credibility crisis in the social sciences, forcing the field to revisit commonly-held assumptions that limit the promise and prospect of our scholarship. Thus, we provide suggestions for reviewers and editors to identify concerns in empirical work, and to guide authors in improving their analyses and research designs. We hope that our editorial provides useful and actionable guidance for entrepreneurship researchers submitting theory-testing papers to *Journal of Business Venturing*.

We are grateful for comments from Herman Aguinis, Anna Dreber Almenberg, Per Davidsson, Frederic Delmar, Timothy Folta, Keith Hmieleski, Alan Johnson, Reddi Kotha, Ali Mohammadi, Ernest O'Boyle, Simon Parker, Miriam Van Praag, Christopher Rider, Nina Rosenbusch, Dean Shepherd, Diemo Urbig, Martin Wallin, Marcus Wolfe, and participants at the IE Business School Tertulias speaker series. Wennberg acknowledges funding from the Royal Swedish Academy of Letters, History and Antiquities.

1. Introduction

As a field, entrepreneurship has much to be proud of in its recent past. Three entrepreneurship journals now rank among the most recognized scholarly publications in business on the *Financial Times 50* list. The Entrepreneurship Division of the Academy of Management grew *four-times* faster than the broader Academy from 2012 to 2016, and more than eighty universities around the world now offer doctoral training in entrepreneurship (Katz, 2018). Though exciting, this rapid growth suggests that our field may no longer be able to assume that its constituents are also members of adjacent fields informed elsewhere about controversial issues or best practices in empirical research. Consequently, *JBV* has a growing moral obligation to make its readers aware of practices impacting entrepreneurship research.

At present, there is growing recognition that many published findings in social science are not likely to replicate—the hallmark for establishing scientific truth—and that the resulting "replication crisis" threatens the legitimacy of academic disciplines relying on quantitative methodologies to test theory (Camerer et al., 2016; Open Science Collaboration, 2015). Although explanations for the

* Corresponding author at: Bloch School of Management, University of Missouri-Kansas City, USA.
E-mail addresses: andersonbri@umkc.edu (B.S. Anderson), karl.wennberg@liu.se (K. Wennberg), mcmullej@indiana.edu (J.S. McMullen).

https://doi.org/10.1016/j.jbusvent.2019.02.001
Received 9 February 2019; Accepted 11 February 2019
Available online 22 February 2019
0883-9026/ © 2019 Elsevier Inc. All rights reserved.

replication crisis range from ample researcher degrees-of-freedom (Gelman and Loken, 2014; Goldfarb and King, 2016; Nuijten et al., 2016; Simmons et al., 2011), to the 'publish-or-perish' promotion and tenure system (Honig et al., 2017),[1] few would question whether a replication crisis exists. Moreover, there is little reason to believe that entrepreneurship is immune to its causes. Therefore, as editors of the leading journal of a rapidly growing field, we face a choice: inform our constituents about best practices they can employ to enhance the rigor of their research, or run the risk of contributing to a crisis that plagues the broader academy.

The purpose of this editorial is to take this growing obligation seriously by calling for a renewed commitment to continuous research improvement. We believe that entrepreneurship's best and brightest contributions to the academy, to public policy, and to practice are in front of us, but only if we commit to bringing as much methodological rigor as possible to some of the most relevant research questions that both business and social science have to offer. Therefore, we propose a number of best practice recommendations intended to help authors sharpen and strengthen their papers, and to help guide reviewers and editors in helping authors to improve their work.

Because *JBV* draws researchers from different backgrounds—psychology, economics, sociology, and more—with different methodological traditions, we have sought, whenever possible, to make our advice applicable to the broad, multi-disciplinary research found in *JBV*. But with methodological approaches being as varied as they are, we had to start somewhere and chose to focus this editorial on mainstream, quantitative theory-testing research that tests a pre-defined hypothesis using either experimental or observational data (Bergh et al., 2017; Munafò et al., 2017). Thus, our target audience for this editorial consists primarily of scholars submitting theory-testing papers to *JBV*. But to be clear, *JBV* strongly values exploratory quantative research—studies designed to generate or otherwise illuminate new hypotheses (Van de Ven et al., 2015)—and research using qualitative and case study designs (Shankar and Shepherd, 2018), and *JBV* is likely to devote subsequent editorials to each topic in its own right.

Because entrepreneurship is a multi-disciplinary, multi-functional, multi-contextual, and multi-level field of study, we also believe that a one-size-fits all approach to methodological policy is inconsistent with the inclusive ethos of the journal and the field. Therefore, our recommendations should not be interpreted legalistically as official editorial policy for *JBV*. Instead, it is our hope that entrepreneurship researchers will voluntarily employ these guidelines as a means of rejecting the false trade-off between rigor and relevance, embracing new technologies and approaches for transparency in data collection and analysis, and encouraging replication and triangulation of results. In so doing, we believe that, collectively, entrepreneurship scholars can accelerate entrepreneurship research while also extending the longevity of their scholarship (Berglund et al., 2018).

2. Useful quantitative theory-testing research

JBV's editorial policy states that "…submitted articles contribute increased understanding of entrepreneurial phenomena. Articles can be either rigorous theoretical contributions or theory-driven empirical contributions." For the purposes of this paper, we define contribution in practical terms—the value that another researcher receives as a function of the published work. This definition places a paper in a value creating framework, such that the paper derives value and hence usefulness from what it offers other researchers (Bacharach, 1989; Edwards and Berry, 2010). As *JBV* editor-in-chief, Jeff McMullen, has noted at doctoral consortiums for several years, a paper is useful because, for example, it helps another researcher answer a question that he or she could not answer before, or it helps another researcher ask a question that he or she could not ask before. In our view, a paper is useful because it adds value—it addresses a meaningful problem in a meaningful way (George et al., 2016).

We posit three factors that jointly define the usefulness of a theory-testing paper. We depict our usefulness function in Fig. 1, with three axes representing (1) motivating the research question, (2) improving causal inference, and (3) reducing researcher degrees of freedom. We measure usefulness as the volume of the cube formed by these intersecting attributes. Depicting usefulness this way recognizes that there are several ways to make a useful contribution to the entrepreneurship literature. The researcher has three axes through which to conceive, execute, and frame an empirical study and thus define a study's usefulness.

2.1. Motivating the research question

The first necessary factor for a paper to be useful is a well-motivated research question in which other entrepreneurship scholars are interested (Hambrick, 2007). This is by far the most subjective element. Excellent guidance already exists on the centrality of the research question in crafting a paper, and we do not feel that we have much to offer beyond these authors (e.g. Alvesson and Sandberg, 2011; Edmondson and McManus, 2007; Ghoshal, 2005). We would emphasize, however, that of the three axes, asking and motivating a question that others perceive as interesting is the most important consideration to evaluate a paper's usefulness. No degree of empirical rigor can take precedence over asking interesting, important, and relevant research questions. A hallmark of entrepreneurship research is the close connection between the questions we ask and the phenomena we study (e.g. Berglund and Wennberg, 2016; Shepherd, 2015). Entrepreneurship research flourishes when the rigor of the empirics rises to the relevance of the question. But it would be a mistake for entrepreneurship research to devolve into a method-driven publication model, where the quest for ever-increasing precision in statistical inference supplants the theoretical and practical utility of the research question itself.

[1] We refer interested readers to these references and related discussions on potential causes for the replication crisis.

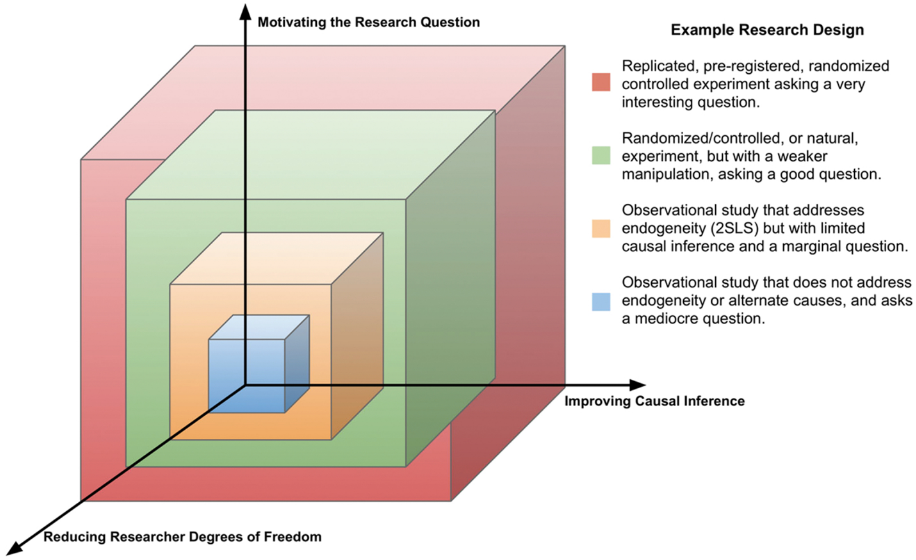

Fig. 1. The research usefulness function.

2.2. Improving causal inference

A second necessary factor for usefulness is the steps taken to maximize causal inference. We focus on casual inference because theory testing entrepreneurship scholarship is generally interested in understanding why a phenomenon occurred and what will happen if it occurs again. We believe that the next frontier in theory testing research involves improving our ability to make causal predictions about entrepreneurship phenomena. Doing so, however, requires greater emphasis on conceiving and executing research designs that improve causal inference (Pearl, 2009).

Scholars often cite randomized controlled trials as the gold standard for drawing causal inference and we agree, though it is worth noting that poorly conceived experiments with small samples and weak manipulations have limited usefulness, despite randomization (Shadish et al., 2002). Some entrepreneurship research questions yield themselves well to randomized controlled experiments. For example, Johnson et al. (2018) investigated the impact of gender bias in the evaluation of an entrepreneur's perceived trustworthiness and competence in a lab setting, where the researchers could create a realistic manipulation for a gender condition. But for some entrepreneurship research questions an experimental design is neither feasible, nor ethical. Consider research on entrepreneurial failure (Cardon et al., 2011); a researcher cannot assign an entrepreneur to a “failed” and “successful” treatment condition any more than, in another example, a researcher could assign firms to a “high” entrepreneurial orientation or a “low” entrepreneurial orientation (Anderson et al., 2015). In these cases, entrepreneurship researchers must seek alternate approaches to design a study making a causal claim.

It is worth noting that in the process of soliciting feedback on this editorial, we observed differences among entrepreneurship scholars on research designs that can, and cannot, establish causal claims. Generally, these differences split between scholars coming from a psychology methodological tradition and those coming from an economics methodological tradition. We do not claim to resolve the difference in perspectives on causal inference from entrepreneurship scholars drawing from these—and other—methodological traditions, nor do we think that it is necessary. One of the reasons for *JBV's* success as the leading entrepreneurship journal is our multi-disciplinary, multi-functional, multi-contextual, and multi-level approach to entrepreneurship research. Establishing causal claims is challenging across methodological traditions, and we discuss strengths and weaknesses of approaches to causal inference later in this editorial. Our point is simply that the usefulness of a theory testing paper depends on the extent to which the research design and analytical approach facilitates causal inference.

2.3. Reducing researcher degrees of freedom

The third necessary element are the steps taken by the researcher to minimize researcher degrees of freedom—the judgment calls and decisions made—when designing a study, conceiving the measures, collecting the data, and conducting the analysis that unintentionally, or intentionally, influence a paper's results (Gelman and Loken, 2014; Simmons et al., 2011). Some researcher degrees of freedom concerns fall under the rubric of questionable research practices that apply across social science disciplines (O'Boyle Jr et al., 2017), while others reflect the judgment calls inherent to studying entrepreneurship phenomena. For the remainder of the editorial, we focus our attention on the issues specific to entrepreneurship, assuming the researcher has motivated the paper with an interesting research question, and he or she now seeks to maximize a study's rigor by improving causal inference and reducing researcher degrees of freedom (the remaining axes of Fig. 1).

3. Theory testing research

3.1. The value and limitations of null hypothesis testing

Simplicity and ease of use are the principal values of the null hypothesis testing framework that dominates most empirical work in entrepreneurship (Cohen, 1994). Simplicity in the sense that often our hypotheses are often binary statements, such as: 'entrepreneurial orientation positively influences firm performance,' or 'entrepreneurial self-efficacy positively influences entrepreneurial action.' Supporting our hypotheses then is simply one of statistical—not practical—significance, and typically with a $p < .05$ standard for rejecting the null hypothesis using frequentist statistics and inference based on the *p*-value (Gelman and Stern, 2006). The null hypothesis framework is remarkably flexible and easy to use in most research settings, including in entrepreneurship research (Simonsohn et al., 2014).

The simplicity and ease of use of null hypothesis testing are also among its many limitations. Central to evaluating the statistical significance of a finding using the *p*-value is the assumption that the null hypothesis is true; the p-value simply being the probability that the observed effect or a larger effect is due to random variation under the assumption of a true null hypothesis (Greenland et al., 2016; Wasserstein and Lazar, 2016). Consider the examples in the previous paragraph. Before beginning a new study, we must assume that there is no relationship between entrepreneurial orientation and performance, and no relationship between entrepreneurial self-efficacy and entrepreneurial action, to infer statistical significance correctly based on the *p*-value. This makes null hypothesis testing using *p*-values an odd perspective with which to test theoretical relationships in entrepreneurship, and arguably across the social sciences (Gelman and Stern, 2006). For both examples, the null hypothesis is highly unlikely, and substantial scholarship already exists highlighting the implausibility of the null. Unfortunately, under null hypothesis testing there is a low threshold to demonstrate evidence in favor of one's hypothesis, and little expectation to provide evidence of the practical likelihood of the null hypothesis being true (Cohen, 1994).

We discuss the limitations of null hypothesis testing not necessarily to dissuade its use, although we believe that alternative approaches such as Bayesian inference and other methods offer untapped promise for entrepreneurship research (Hedström and Wennberg, 2017; Johnson et al., 2015). Our point is that the null hypothesis paradigm requires several theoretical assumptions necessary for inference that many entrepreneurship research questions and settings are not likely to satisfy (Kruschke and Liddell, 2018). Despite its limitations, however, null hypothesis testing remains the most popular approach to theory testing (Meyer et al., 2017), and that is certainly true among quantitative studies published in *JBV*. In the following sections, we discuss best practices for theory-testing empirical work appearing in *JBV*. Our recommendations apply to research designs relying on null hypothesis testing and frequentist (e.g., *p*-value) statistics, but many also apply to research designs using Bayesian inference or other approaches. As a starting point, we strongly encourage researchers to consider during the research design stage the plausibility of the null hypothesis for their research question. The less plausible the null, the more likely the researcher is to observe statistically significant—and potentially spurious—results (Dienes and Mclatchie, 2018; McShane and Gal, 2017).

3.2. An ounce of prevention

The single best thing we as entrepreneurship scholars can do in our theory testing research is to invest in the research design process. Although this editorial does not cover the range of research designs found in entrepreneurship work, there is an ample, and excellent, body of scholarship on research designs for causal inference (e.g., Angrist and Pischke, 2008; Pearl, 2009; Morgan and Winship, 2007; Shadish et al., 2002). We simply wish to point out that it is difficult to improve a paper's usefulness *after* designing the study, collecting the data, and writing the paper.

As part of the research design process for theory-testing studies, we encourage researchers to consider pre-registration (van't Veer and Giner-Sorolla, 2016). Pre-registration involves thinking through the design and hypotheses for a study before collecting data, and then documenting those choices on a public forum; for example, the *Open Science Framework* (www.osf.io). Pre-registration suits both experimental and observational designs and does not have to be onerous; simply disclosing which hypotheses, which variables, and which manipulations (if an experiment) the researcher intends to test improves the usefulness of theory-testing work. Greater disclosure—such as that required by most university human subjects research boards—is better, but a short, one-page description of the study's basic design and hypotheses suffices. Pre-registration provides transparency on the researcher degrees of freedom inherent to all theory-testing work, and provides an accountability mechanism to discourage researchers from changing analytical approaches or variable combinations if initial results do not turn out as expected. In the case of the latter, the researcher is free to make research design changes, but with the pre-registration, would do so in the context of reporting the initial study and then the changes made in a subsequent study or as post-hoc tests to the original hypotheses posed.

3.3. Best practices for theory-testing entrepreneurship research

A common problem in the social sciences is implicitly or explicitly "hunting for asterisks" (Bettis, 2012). In years past, hunting for asterisks was by and large an accepted approach to the research and publication process (O'Boyle Jr et al., 2017), and we too were guilty of the behavior. Fortunately, we know better now. Researchers can publish good entrepreneurship studies asking interesting questions using rigorous research designs regardless of whether he or she identifies statistically significant results. Numerous scholars in numerous disciplines amply discuss the p-hacking, HARKing, and multiple comparison problem, and we do not expand on them here (Aguinis et al., 2018; Bettis et al., 2016; Gelman and Loken, 2013; Meyer et al., 2017; Simmons et al., 2011).[2] We believe that

Table 1
Best practice recommendations for theory-testing entrepreneurship research.

Replication & transparency

- *Self-replicate your study* – Consider direct replications to present in the same paper. If not feasible, consider a mixed-method approach to provide another test of the research question (Bettis et al., 2016; Davidsson, 2016).
- *Be transparent, share code, and ideally share data* – Be open and honest about what you did and why you did it; make use of appendices or other formats to be as complete as reasonable about the study, its design, and its measures. While legal, privacy, or contractual obligations may limit data sharing, there are no such restrictions on sharing code (Aguinis et al., 2018).
- *Ensure sufficient information for reproducibility*– Provide enough information on the data and measures that another scholar could reproduce the results, ideally with the data itself but at a minimum by using the reported correlation or covariance matrix (Bergh et al., 2017).

Causal inference

- *Take steps to maximize causal inference at the design stage*– Make use of causal frameworks to evaluate threats to causal inference before beginning a study; it is very difficult to tackle causality problems after collecting data (Angrist and Pischke, 2008; Pearl, 2009; Shadish et al., 2002).
- *Take endogeneity seriously*– Endogeneity can occur even in experimental research, and should always be an assumption in observational research. Tackle endogeneity problems starting with research design supplemented by appropriate statistical tools (Antonakis et al., 2010).
- *Disclose researcher degrees of freedom that may impact causal inference*– Disclose all the manipulations, treatments, and instruments in a study, and discuss differences in model results from different analytical choices (Simmons et al., 2011).

Interpretation & statistical inference

- *Provide and interpret coefficient estimates, marginal effects, and confidence intervals*– Go beyond reporting the point estimate to provide a discussion of the effect relative to the dependent variable, the marginal effects if appropriate, and confidence intervals (Cohen et al., 2003).
- *Place the p-value in context*– Report exact *p*-values and remember that the p-value is not evidence in favor of or against the null hypothesis—it is a single data point to help evaluate analyses. Consider the plausibility of the null hypothesis; the less likely the null, the more likely an analysis yields a statistically significant result (Cohen, 1994; McShane and Gal, 2016).
- *Judge inference relative to sample size*– Big claims require big evidence; small samples are noisy, and in the case of a false positive, generally inflate effect sizes. Conversely, large samples may reveal statistically significant, but trivial effect sizes (Kelley and Maxwell, 2003; Loken and Gelman, 2017).

entrepreneurship researchers, like scholars in our sister disciplines, have begun to clean house by sweeping the 'hunting for asterisks' methodology to the dustbin of history.

In the next three sub-sections we highlight several best practices for advancing theory-testing entrepreneurship research. We provide a summary of our recommendations in Table 1, including references for additional information, and explain each point in more detail below. Any given study is unlikely to "check all the boxes." Therefore, our objective here is not to demand execution of the perfect study; there are weaknesses in every study and every paper. Instead, our intent is to help entrepreneurship researchers who are seeking to answer important questions learn how they might increase the rigor of the design and analyses of their theory-testing papers.

3.3.1. Replication and transparency

We cannot stress enough the importance of replication to theory-testing research. This is particularly important for entrepreneurship scholarship using publicly available datasets such as the Panel Study of Entrepreneurial Dynamics (PSED), the Kauffman Firm Study (KFS) and the Global Entrepreneurship Monitor (GEM). One could expect studies using these or similar datasets to suffer from a multiple comparison problem, especially if they are modeling the same outcome (Bettis, 2012). Under the $p < .05$ standard, we would expect 1 in 20 studies to yield an incorrect nomological conclusion purely by chance (Bettis et al., 2016). A single study using a single sample is an N of 1; a replication doubles the data points! Given the noise inherent in the study of entrepreneurship phenomena even with minimal researcher degrees of freedom and maximal transparency, it remains possible for observed results merely to be an artifact of the data.

Our recommendation is self-replication. Common now in psychology and behavioral economics, a self-replication uses the same research design, the same variables, and ideally the same code, on a different random sample and reported in the same paper (Bettis et al., 2016; Davidsson, 2016). Self-replication attempts to hold as much of the original research constant such that the only variation is the sample itself. Experimental designs and survey research lend themselves easily to self-replication, with the researcher simply adding additional rounds of data collection from a new random sample of the population of interest to a project. Research using publicly-available data sources (e.g., NLSY, PSED, PSED II, GEM) is more challenging to self-replicate. These data generally come from large-scale, multi-year research projects out of the scope for a single researcher or author team.

If self-replication is not feasible, we encourage authors to adopt a mixed-method approach to ask the same research question using two or more research designs. Most entrepreneurship and management journals, including *JBV*, strongly encourage this type of research. Called a *conceptual replication*, researchers may use different estimators, different variable operationalizations, and different sampling approaches to test the same research question as an earlier study and report these results in the same paper (Eshima and Anderson, 2017). The goal is to evaluate nomological consistency, where the researcher evaluates whether different approaches to the research question yield consistent insights (Bergh et al., 2017). In thinking through a conceptual replication, researchers should also consider appropriate sample sizes and statistical power for each study in the paper independently. For example, the researcher may be using an experimental design for study one and a survey design for study two—a sample of two hundred respondents may be

[2] P-hacking refers to selecting variables and analyses based on statistical significance; HARKing refers to constructing a theoretical explanation for a set of results after conducting data analysis; multiple comparisons refers to indiscriminately running statistical tests until achieving statistical significance or another desired result (see Aguinis et al., 2018).

appropriately sized for the former, but not for the later. Ultimately, unless the study period is very long (e.g., NLSY), or the researcher explores a sub-sample where external conditions are less likely to change drastically across context, single sample research tends to lower usefulness.

Regarding transparency, a challenge for entrepreneurship researchers is that multiple definitions and measures may exist for the same construct. In these cases, transparency and specificity is critical to evaluate a study's usefulness and to place a study's findings in context. Consider entrepreneurial orientation, for example. Rather than stating that a study “tests the relationship between entrepreneurial orientation and performance,” a researcher could write that the study “tests the relationship between the Lumpkin and Dess (1996) entrepreneurial orientation conceptualization and sales growth rate.” Another transparency consideration for entrepreneurship scholars is the choice of sampling strategy. Self-selection into entrepreneurship, firms self-selecting a strategic posture, and survival bias each create conditions that can make estimated effects a function of the chosen sample rather than representative of a robust underlying effect (Sine et al., 2006). Convenience sampling, self-selection by responding to surveys, or respondents choosing a secondary respondent (common when researchers mail or email surveys to founders and senior managers) are all examples of disclosable and often addressable, selection effects common in entrepreneurship research (Certo et al., 2016).

An additional transparency consideration for entrepreneurship scholars is to consider carefully the study context in terms of region, time-period, and the population characteristics from which the researcher samples. Given the often context-specific patterns of entrepreneurial phenomena, such details are pivotal to interpret a study's results (Aldrich, 2009). Researchers increase the usefulness of theory-testing research by discussing how and why a study used a given sample, any range restrictions on that sample (for example, dropping firms below a given size or age), selection effects influencing the sample, and the generalizability of the sample to the population of interest.

Regarding experimental designs, under ideal conditions, replicating a previous experiment helps establish ecological validity, or the extent to which a finding generalizes to another sample (Shadish et al., 2002). But as high-profile replication efforts in psychology and behavioral economics show, it is difficult to conduct ideal experiments (Camerer et al., 2016; Open Science Collaboration, 2015). As such we strongly encourage self-replication of experimental designs, and reported in the same paper. For transparency, we also encourage researchers to report all manipulations used in an experiment, and to provide as much documentation as possible on those manipulations to enable replication. While tempting to report only those manipulations resulting in a statistically significant effect, transparency and replicability favor reporting all attempted manipulations, whether ‘successful’ or not (Simmons et al., 2011). In cases where it may not be feasible to report all manipulations owing to paper length, we encourage researchers to leverage online repositories and appendices. In effect, what we are asking is that the researcher engages in a dialogue with the reviewers and editor. If an initial study in the lab yields a statistically significant result, but a self-replication does not, there may be an unknown boundary condition at work, a statistical power problem, or a measurement error problem, or another explanation; determining which is far easier with help. Researchers and the field have much to gain by self-replicating and maximizing transparency, especially when we openly share the discoveries made and lessons learned during the journey.

Lastly, we would like to encourage researchers to consider making their data and code publicly available. Although data and code sharing are becoming more prevalent across the social sciences, neither is yet a norm in our field. Many researchers may also be using point-and-click software, and thus not documenting their analyses and code in an easily sharable way. Nonetheless, the trend towards making data and code publicly available is on the rise, and is likely to be an important progression for our field as well. Authors can post data and code on their personal or organizational website, or make use of the growing number of online options. The ResearchGate (www.researchgate.net) platform, for example, allows researchers to assign a *Document Object Identifier* (DOI) to data that is citable by other researchers. This tool allows data owners to receive recognition for the difficult work and costs associated with collecting and processing data. Authors can share code (preferably in the form of plain-text files to avoid proprietary formatting) on ResearchGate, personal or organizational websites, or specific code-sharing platforms such as GitHub (www.github.com).

3.3.2. Causal inference

As Holland (1986) noted, there is no causation without manipulation. Given that randomized controlled experiments are the gold standard for drawing causal inference, we join other scholars in encouraging more experimental designs—natural or laboratory—in entrepreneurship research (Hsu et al., 2017; Williams et al., 2019). Experimental designs are not without their challenges, and substantive threats to causal inference remain even in the best executed experiment, as we will discuss later (Antonakis et al., 2010). What we would like to focus on first though is causal inference in observational designs—those that do not involve manipulating a predictor variable—and specifically the endogeneity problem in these designs.

Endogeneity exists when the disturbance term in a model equation correlates with one or more of the independent variables (Antonakis et al., 2010). When this happens, the resulting coefficient estimate for the relationship between the independent variable and the dependent variable is no longer consistent. No matter how large the sample becomes, the coefficient estimate will not converge to its population value (Angrist and Pischke, 2008; Antonakis et al., 2010; Pearl, 2009). Endogeneity may attenuate or accentuate estimates and impacts virtually all estimators (e.g., OLS and related least squares estimators, multilevel estimators, and maximum likelihood-based estimators such as logit and probit models, hazard models, structural equation models, and so forth). Endogeneity is not a red-herring, nor is it something that a researcher simply assumes away. Because endogeneity arises from so many potential sources, it is not feasible when using observational data to collect enough variables to control for all possible endogeneity problems (Morgan and Winship, 2007). Selection effects, measurement error, common methods variance, failing to account for clustered data, and omitting alternate causal variables may all induce endogeneity (Antonakis et al., 2010).

The best way to avoid the endogeneity problem is in the research design phase, where data collection should be as close to the experimental ideal (in the lab or in the field) as possible. For observational studies, a valuable approach is to compare the proposed research design with an 'ideal' experimental design, and then think through how deviations from that ideal create inferential problems (King et al., 1994). During the design stage for an observational study, a researcher could identify an instrument variable, exogenous to the error term of the equation, that mimics or otherwise "stands in" for random assignment in a laboratory setting (Angrist and Pischke, 2008). Entrepreneurship researchers may also employ a difference-in-differences design by assigning treatment and control groups based on a naturally occurring phenomenon, such as a change in public policy (Davidsson and Gordon, 2016). Regression discontinuity designs also compare the effect of an intervention where researchers can distinguish between treatment and control groups based on a naturally occurring cut-off (Decramer and Vanormelingen, 2016).

At the data analysis stage, there are several approaches to help entrepreneurship scholars address endogeneity problems. For example, Heckman correction models with strong and valid exclusion criteria help address sample selection effects (Delmar and Shane, 2003). Two-Stage Least Squares (2SLS) and related instrument variable methods are also useful, so long as one can identify strong and valid exogenous instruments (Lee, 2018; Tang and Wezel, 2015). Importantly, researchers should identify instruments as part of the research design process—before collecting data. If clever ideas emerge during analysis to improve an instrument or identification strategy, the researcher should report changes, manipulations, or substitutes to any instrument, along with the original analysis. Additional approaches include propensity score matching and new Bayesian estimators, such as the `causalImpact` time series analysis package for the R language (Brodersen et al., 2015).

Turning to experimental designs, perfect randomization ensures that the manipulated variable remains exogenous to the outcome measure (Shadish et al., 2002). Unfortunately, perfect randomization is rarely achievable, and even in the lab, endogeneity remains a salient concern (Antonakis et al., 2010). We would like specifically to highlight the endogeneity problem in a common experimental design where the researcher tests a mediation hypothesis, $x \rightarrow m \rightarrow y$, but measures the mediator. For example, the researcher might randomly assign participants to a treatment condition for x, but measure m with a Likert-style scale. There are at least two endogeneity concerns here. Randomization only ensures exogeneity of the path from $x \rightarrow m$; because m remains measured, the researcher cannot assume that m is exogenous to y. Omitted variable bias impacting the path from $m \rightarrow y$ is therefore the first endogeneity concern for this model. In this case, one solution is to find an instrument variable for m that allows the researcher to recover a consistent estimate of the effect of $m \rightarrow y$ (MacKinnon and Pirlott, 2015).

Measurement error is the second endogeneity concern in the model above. Measurement error is a problem across virtually all models in the social sciences, although we want to focus specifically on measurement error in models with latent constructs, which are common in entrepreneurship research. Measurement error is itself a source of endogeneity, which often goes overlooked by entrepreneurship researchers (Antonakis et al., 2010). Researchers that simply take the mean of a set of measures for a latent construct implicitly assume measurement error equals zero; unfortunately, this assumption ignores the possibility that measurement error induces a correlation between the summated scale and the disturbance term of the equation (Kline, 2015). The simplest solution—and the one we strongly recommend for any model making use of latent constructs—is to use an estimator that explicitly models measurement error, such as latent variable structural equation modeling (SEM; see Kline, 2015). SEM offers researchers a host of benefits beyond modeling measurement error, such as easily incorporating instrument variables as in the case of the mediation example above, and we strongly encourage its use when appropriate (Eshima and Anderson, 2017).

Irrespective of research design and experimental or observational data, it seems that the field is passing the point where it is appropriate to submit research that tests causal, or even associational, hypotheses without a discussion of alternative explanations and endogeneity, along with providing an appropriate method to minimize their impact. A theory-testing paper lacking acknowledgement of endogeneity and an appropriate approach to address it is demonstrably less useful than it could be with these relatively straightforward changes.

3.3.3. Interpretation and inference

It is critical for authors to correctly interpret the coefficient estimates for their models. For example, it is common to see results of a logit model incorrectly reported as 'effect sizes' rather than as log odds (Wiersema and Bowen, 2009). Similarly, researchers often report marginal effects 'holding all other variables constant' at their mean value, which is fine if the variable(s) is continuous, but incorrect if a variable is categorical or heavily skewed. Consider another example: in a panel model with higher-order (j) fixed effects, the coefficient estimate is the expected within-entity (i) change in the dependent variable over repeated observations of the independent variable, and not the 'overall effect' of the independent variable on the dependent variable (Certo et al., 2017; Cohen et al., 2013). Along with correctly interpreting estimates, we encourage researchers to calculate and discuss marginal effects with theoretically salient values, such as at the median and for a select few standard cases (Hoetker, 2007). Further, in multivariate models including a mix of continuous and categorical predictors, researchers should make use of tools to present the estimated marginal effects in plain language to make their results more accessible (Krasikova et al., 2018). We also encourage researchers to report exact p-values, standard errors, confidence intervals, and related parameters appropriate to the specified model.

As statistical software becomes even more powerful and user-friendly, researchers face the challenge of explaining complicated statistical and mathematical assumptions to readers who may not be familiar with a study's method. Similarly, point-and-click software makes it easy to specify complicated models without the researcher necessarily understanding the assumptions made by the method. Not every researcher is an expert in every method, nor should that be an expectation. But we need to be cognizant of the need to align the assumptions of the method with the data at hand, and to use an alternate approach if a preferred method is not feasible. Ultimately, the researcher bears the responsibility for explaining why he or she chose the data analysis method for a study, what assumptions the method makes, why the data aligns with those assumptions, and what alternative methods the researcher tried in the analysis process.

When interpreting model results, we also need to acknowledge that most effect sizes in the social sciences, and particularly in entrepreneurship, are likely to be small (Martin et al., 2013; Rosenbusch et al., 2011). We suggest that authors discuss their results, particularly in large samples with small standard errors and small coefficient estimates, in ways that make an effect size relatable. For example, rather than reporting 'we find support for our hypothesis ($\beta = .015$; $p < .05$),' authors might say 'holding all other variables in the model constant, we find that increasing research and development by $100,000 results in 15 more patents per one-thousand employees ($\beta = .015$; st. err. = .005; $p = .024$).' The better the explanation, the better the reader can judge the usefulness of the paper's findings. As Davidsson (2016) noted, entrepreneurship scholarship could benefit from a robust discussion of effect sizes, including articulation of the anticipated effect size when specifying a hypothesis. Formulating and discussing hypotheses that provide more detailed predictions—rather than simply a deviation from zero (as with the null hypothesis)—ultimately helps our field develop theories with greater precision (Edwards and Berry, 2010).

Entrepreneurship research with small samples merits additional consideration. Researchers often think that small samples make it difficult to detect statistically significant relationships (Cohen et al., 2003). That many small samples in entrepreneurship identify a statistically significant effect is, somewhat paradoxically, our primary concern here—large effects in small samples may not be revealing any 'real' effect at all (Kelley and Maxwell, 2003). Often referred to as the 'that which does not kill statistical significance makes it stronger' fallacy (Loken and Gelman, 2017), it is easy to assume that, because a statistically significant appears in a small sample, the observed effect is quite robust or prominent. Unfortunately, small samples may yield spurious results solely as a function of the increased variance—noise—of the sample itself (Gelman and Weakliem, 2009). Weak, underpowered samples increase the likelihood of Type I errors (false positives) and may yield estimates *in the opposite direction* of a true, underlying effect (Button et al., 2013). The signal-to-noise ratio is also problematic in small sample studies testing mediation or moderation. Unfortunately for both mediation and moderation, sample variance—particularly in small samples—increases the likelihood of spurious or inflated results, and measurement error compounds the problem of testing complex models in small samples (Murphy and Russell, 2017).

We believe the signal-to-noise ratio problem with small samples is particularly troubling for entrepreneurship research because many times small samples are the only option available; it is difficult to collect data on practicing entrepreneurs or senior decision-makers in organizations. Although the simple solution is to seek as large a random sample as possible, from a practical perspective, a self-replication on a new random sample reported in the same paper is the most viable solution to the sample size challenge. Admittedly, small, well-defined theoretical or homogenous samples may be better than large, noisy samples to identify complex mechanisms (Aldrich, 2009; Davidsson, 2016). But in these cases, the researcher should address concerns that the variability inherent to small samples is not a salient concern for the model, along with motivating the study context (e.g., country, region, time window) and necessity for a small sample (Welter, 2011). Researchers employing small samples may also benefit from using alternative modeling strategies such as Bayesian modeling that better contextualize uncertainty and variance (Depaoli and van de Schoot, 2017).

4. Putting our recommendations into practice

4.1. Evaluating the usefulness of our research

We would like to revisit how our best practices from above accelerate entrepreneurship research and increase the usefulness of our work. What we present below is a rubric to evaluate the usefulness of a future study while incorporating our recommendations. The objective is to inform the researcher's decisions about research design and analytical approach *before* data collection.

If we return to the usefulness graphic in Fig. 1 and assuming the researcher is asking a well-motivated question, the challenge becomes how to maximize the volume of the usefulness cube along the other two axes. The most useful research answers the research question by identifying a causal mechanism derived from theory, verified on data, and generalizable to similar settings and situations (Hedström and Wennberg, 2017). Multiple replications of a pre-registered randomized controlled experiment in which the researcher directly manipulates the independent variable(s) with well validated and reliable measures is the ideal (Ioannidis, 2016). This research is also the most difficult to conduct, and we make no pretense for the cost, complexity, time, knowledge, and expertise necessary to develop and execute randomized controlled experiments. We do feel, however, that experimental work in entrepreneurship offers substantial opportunity to improve the usefulness of our research, and to accelerate the impact of entrepreneurship scholarship on practice (as a case in point, we call attention to the recent special issue of *JBV* on the topic [Williams et al., 2019]).

In many ways, however, experimental designs represent an aspirational target for entrepreneurship researchers. Many of the phenomena we study do not lend themselves to experimental manipulation and we must rely on observational data and associated research designs. But when a researcher identifies an exogenous source of variation (e.g., a natural experiment [Morgan and Winship, 2007]), meant to mimic assignment to a treatment versus control condition, it is possible to draw causal inference from observational data (Angrist and Pischke, 2008; Pearl, 2009). We recognize that not all entrepreneurship scholars share our perspective, and we concur with the primacy of the randomized controlled experiment. But it would be unnecessarily limiting to entrepreneurship research to summarily dismiss the usefulness of well-designed and well-executed observational studies, particularly when the research question rules out a randomized controlled experiment.

Importantly, despite the rapid development of new statistical tools to model observational data, a statistical method in and of itself does not allow a researcher to make a causal claim (Pearl, 2009). Establishing a causal relationship—whether in the lab or in the field—is foremost a function of research design and a deep understanding of the research question. Useful frameworks for thinking through the necessary conditions for causality include the Rubin (2005) causal model, (i.e., potential outcomes [Angrist and Pischke, 2008]); structural causal modeling (Pearl, 2009); or the conditions laid out by Shadish et al. (2002). A researcher may prefer a specific framework depending on his or her methodological background, but each is valid. Regarding statistical tools that help model causal relationships with observational data, difference-in-differences models, propensity score matching, and regression discontinuity are all useful approaches, assuming the research design itself facilitates causal inference (Hildebrand et al., 2016; Kagan et al., 2017).

As the volume of the cube shrinks so does a paper's usefulness; for example, theory-testing papers that may be statistically sophisticated and highly transparent, but with a research design that does not allow the researcher to infer causality. There is often a valid trade-off between causal evidence and evaluating theoretical models holistically; for example, testing models with multiple dependent variables or models with many predictors simultaneously (Salganik, 2017). Many entrepreneurs, and perhaps researchers, are more concerned with a holistic perspective than a precise causal explanation for a specific action or strategy. Studies like these generally do not manipulate the independent variable, nor randomly assign participants to a treatment condition. Still, this research should make use of statistical approaches (e.g., instrument variable/2SLS approaches [Angrist and Pischke, 2008]) to help address biased estimates and other endogeneity problems. Additional tools that help address bias include tree regression, penalized regression, and Bayesian model averaging (Varian, 2016). Many of these tools also offer the potential of 'backwards forecasting,' which is useful to explain uncertain relationships that are highly time and context-dependent, and are also useful in modeling macro-level outcomes, such as economic development, from micro-level data such as from firms or individuals (Humphries, 2017).

Shrinking the volume of the cube further, we find limited usefulness in theory-testing research that lacks manipulation of the independent variable, lacks random assignment to a treatment condition, uses a single sample, and makes no attempt to address endogeneity concerns or uses a flawed endogeneity correction (e.g., specifying a model from $x \rightarrow y$ and then from $y \rightarrow x$ to address 'reciprocal causality' [see Antonakis et al., 2010]). Studies in this category may use cross-sectional designs, employ time lags with loose theoretical justification (e.g., using a one-year lag because that was available in the data), model noisy measures, use proxy variables that only loosely relate to the theoretical construct, and employ small or convenience samples. This research may be useful for exploring potential hypotheses for new areas or research ideas, but it is better suited to early stage exploratory work. That said, one way to improve the usefulness of this type of research would be to present the exploratory study as preliminary, but then also report a second confirmatory study in the same paper. Because early stage exploratory studies with promising results may encourage more rigorous examination, we hope that researchers using these types of studies will make use of online data and code repositories to share results and insights with the scholarly community.

4.2. Considerations for editors and reviewers

We would like to address how reviewers and editors can use our recommendations. Particularly for theory-testing research, we think the considerations outlined in Table 1 provide a helpful checklist for reviewers and editors to evaluate the usefulness of a theory-testing paper. By no means is this an exclusive list of considerations during the review process, and whether a paper motivates an interesting research question is foremost a question for the editor and reviewers. But to the list in Table 1, we would like to add three additional considerations for reviewers and editors.

4.2.1. Encourage sufficient information for reproducibility

The advantage of requiring descriptive statistics (mean, standard deviation, and correlations) is to give another researcher the ability to reproduce the results presented in the paper (Bergh et al., 2017). Tools such as `corr2data` for Stata and the `simstudy` package for R allow readers to generate data from descriptive statistics, which a researcher can use to reproduce the original study's findings. There are, however, limitations to this approach, and this is where editors and reviewers should probe authors to provide more information. For example, in panel/multilevel data, a correlation matrix pools the within (i) and between (j) variance, making it impossible to reproduce the result of a multilevel analysis from the correlation matrix alone. In this case, reviewers and editors should encourage code and, ideally, data sharing. In latent variable structural equation modeling, reporting only the correlation matrix between latent constructs ignores the measurement model and inhibits replicability. In this case, authors should provide the original data or the indicator-level covariance matrix. The critical point is that a study should provide enough information—data, code, model, and estimation approach—for another researcher to reproduce the results.

4.2.2. Guard against HARKing in the review process

One valuable way to combat HARKing and p-hacking is for reviewers and editors to avoid recommending that an author drop, add, or replace a hypothesis in a paper. We also do not recommend post hoc hypothesis testing; for example, testing a three-way interaction effect after originally hypothesizing only two-way interaction effects, and presenting the analysis as confirmatory. In both cases, reviewers and editors are explicitly asking the author to engage in p-hacking and potentially HARKing, which limits the usefulness of a paper. While often well-intentioned, the practice substantially degrades a paper's statistical conclusion validity. One solution, if a paper requires revisiting or fundamentally changing hypotheses, is to encourage a second, self-replicating study in the paper testing only the new or modified hypotheses, along with an appropriate disclosure. It is worth noting though that post hoc exploratory analysis, such as the example of a post hoc three-way interaction effect, can be a useful addition to a paper if presented just as an exploratory post hoc analysis motivated by results in the main analyses.

4.2.3. If you suspect p-hacking, say so

U.S. Supreme Court Justice Potter Stewart famously quipped, 'I know it when I see it.' Justice Potter was referring to pornography, but the phrase can apply to describing something without clear definitional boundaries. We suggest reviewers and editors constructively challenge authors presenting analyses and models that seem either too good to be true, or that seem too complex to have been conceived before data collection. This is, admittedly, a delicate subject, because the editor and reviewer are effectively implying that the researcher has done something unethical. We believe, however, that tackling the p-hacking and HARKing problem requires forthrightness and openness. If you as an author selected those hypotheses based on associations that were statistically significant, please, inform the editor during the review process. The paper may still be useful and capable of continuing the review process, assuming appropriate disclosure, but it requires transparency.

5. Conclusion

As we look to accelerate entrepreneurship research, it will take time for new methods and new ways of approaching research design and analyses to permeate the publishing process and inform doctoral student training. Our perspective is that step-by-step, the collective advancements made by entrepreneurship research in the past few decades will accelerate with scholars adopting the practices outlined in this paper. Any single paper will never be 'perfect' from a research design and analysis perspective. But we benefit as a field by pushing the standards for theory-testing quantitative research forward. We believe that entrepreneurship scholars have just scratched the surface of what we have to offer our stakeholders. By improving the rigor of our methods to approximate the relevance of our research questions, we further the promise of entrepreneurship research to better the condition of entrepreneurs, managers, firms, and even nations.

References

Aguinis, H., Ramani, R.S., Alabduljader, N., 2018. What you see is what you get? Enhancing methodological transparency in management research. Acad. Manag. Ann. 12, 83–110.

Aldrich, H., 2009. Lost in space, out of time: How and why we should study organizations comparatively. In: King, B.G., Felin, T., Whetten, D.A. (Eds.), Studying Differences between Organizations: Comparative Approaches to Organizational Research. Research in the Sociology of Organizations. Emerald Group, Bingley, pp. 21–44.

Alvesson, M., Sandberg, J., 2011. Generating research questions through problematization. Acad. Manag. Rev. 36, 247–271.

Anderson, B.S., Kreiser, P.M., Kuratko, D.F., Hornsby, J.S., Eshima, Y., 2015. Reconceptualizing entrepreneurial orientation. Strateg. Manag. J. 36, 1579–1596.

Angrist, J.D., Pischke, J.-S., 2008. Mostly Harmless Econometrics: an empiricist's Companion. Princeton University Press.

Antonakis, J., Bendahan, S., Jacquart, P., Lalive, R., 2010. On making causal claims: a review and recommendations. Leadersh. Q. 21, 1086–1120.

Bacharach, S.B., 1989. Organizational theories: some criteria for evaluation. Acad. Manag. Rev. 14, 496–515.

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. Strateg. Organ. 15, 423–436.

Berglund, H., Wennberg, K., 2016. Pragmatic entrepreneurs and institutionalized scholars? On the path-dependent nature of entrepreneurship scholarship. In: Parhankangas, A., Fayolle, A., Riot, P. (Eds.), Landstrom, H. Routledge, Challenging Entrepreneurship Research, pp. 37–52.

Berglund, H., Dimov, D., Wennberg, K., 2018. Beyond bridging rigor and relevance: the three-body problem in entrepreneurship. J Bus Ventur Insights 9, 87–91.

Bettis, R.A., 2012. The search for asterisks: compromised statistical tests and flawed theories. Strateg. Manag. J. 33, 108–113.

Bettis, R.A., Ethiraj, S., Gambardella, A., Helfat, C., Mitchell, W., 2016. Creating repeatable cumulative knowledge in strategic management. Strateg. Manag. J. 37, 257–261.

Brodersen, K.H., Gallusser, F., Koehler, J., Remy, N., Scott, S.L., 2015. Inferring causal impact using Bayesian structural time-series models. Ann. Appl. Stat. 9, 247–274.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., 2016. Evaluating replicability of laboratory experiments in economics. Science 351, 1433–1436.

Cardon, M.S., Stevens, C.E., Potter, D.R., 2011. Misfortunes or mistakes?: cultural sensemaking of entrepreneurial failure. J. Bus. Ventur. 26, 79–92.

Certo, S.T., Busenbark, J.R., Woo, H.s., Semadeni, M., 2016. Sample selection bias and Heckman models in strategic management research. Strateg. Manag. J. 37, 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. Strateg. Manag. J. 38, 1536–1556.

Cohen, J., 1994. The earth is round ($p < .05$). Am Psychol 49, 997–1003.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2003. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Erlbaum, Maywah, NJ.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2013. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Routledge.

Davidsson, P., 2016. The Power of Replication, Researching Entrepreneurship. Springer, pp. 247–284.

Davidsson, P., Gordon, S.R., 2016. Much ado about nothing? The surprising persistence of nascent entrepreneurs through macroeconomic crisis. Enterp. Theory Pract. 40, 915–941.

Decramer, S., Vanormelingen, S., 2016. The effectiveness of investment subsidies: evidence from a regression discontinuity design. Small Bus. Econ. 47, 1007–1032.

Delmar, F., Shane, S., 2003. Does business planning facilitate the development of new ventures? Strateg. Manag. J. 24, 1165–1185.

Depaoli, S., van de Schoot, R., 2017. Improving transparency and replication in Bayesian statistics: the WAMBS-checklist. Psychol. Methods 22, 240.

Dienes, Z., Mclatchie, N., 2018. Four reasons to prefer Bayesian analyses over significance testing. Psychon. Bull. Rev. 25, 207–218.

Edmondson, A.C., McManus, S.E., 2007. Methodological fit in management field research. Acad. Manag. Rev. 32, 1246–1264.

Edwards, J.R., Berry, J.W., 2010. The presence of something or the absence of nothing: increasing theoretical precision in management research. Organ. Res. Methods 13, 668–689.

Eshima, Y., Anderson, B.S., 2017. Firm growth, adaptive capability, and entrepreneurial orientation. Strateg. Manag. J. 38, 770–779.

Gelman, A., Loken, E., 2013. The Garden of Forking Paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. Columbia University, Department of Statistics.

Gelman, A., Loken, E., 2014. The Statistical Crisis in Science Data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. Am. Sci. 102, 460.

Gelman, A., Stern, H., 2006. The difference between "significant" and "not significant" is not itself statistically significant. Am. Stat. 60, 328–331.

Gelman, A., Weakliem, D., 2009. Of beauty, sex and power: too little attention has been paid to the statistical challenges in estimating small effects. Am. Sci. 97, 310–316.

George, G., Osinga, E.C., Lavie, D., Scott, B.A., 2016. Big data and data science methods for management research. Acad. Manag. J. 59, 1493–1507.

Ghoshal, S., 2005. Bad management theories are destroying good management practices. Acad. Manag. Learn. Educ. 4, 75–91.

Goldfarb, B., King, A.A., 2016. Scientific apophenia in strategic management research: significance tests & mistaken inference. Strateg. Manag. J. 37, 167–176.

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., Altman, D.G., 2016. Statistical tests, P. Eur. J. Epidemiol. 31, 337–350.

Hambrick, D.C., 2007. The field of management's devotion to theory: too much of a good thing? Acad. Manag. J. 50, 1346–1352.
Hedström, P., Wennberg, K., 2017. Causal mechanisms in organization and innovation studies. Innovation 19, 91–102.
Hildebrand, T., Puri, M., Rocholl, J., 2016. Adverse incentives in crowdfunding. Manag. Sci. 63, 587–608.
Hoetker, G., 2007. The use of logit and probit models in strategic management research: critical issues. Strateg. Manag. J. 28, 331–343.
Holland, P., 1986. Statistics and causal inference. J. Am. Stat. Assoc. 81, 945–960.
Honig, B., Lampel, J., Siegel, D., Drnevich, P., 2017. Special section on ethics in management research: norms, identity, and community in the 21st century. Acad. Manag. Learn. Educ. 16, 84–93.
Hsu, D.K., Simmons, S.A., Wieland, A.M., 2017. Designing entrepreneurship experiments. Organ. Res. Methods 20, 379–412.
Humphries, J.E., 2017. The Causes and Consequences of Self-Employment over the Life Cycle. Yale University.
Ioannidis, J.P., 2016. Why most clinical research is not useful. PLoS Med. 13, e1002049.
Johnson, A.R., van de Schoot, R., Delmar, F., Crano, W.D., 2015. Social influence interpretation of interpersonal processes and team performance over time using Bayesian model selection. J. Manag. 41, 574–606.
Johnson, M.A., Stevenson, R.M., Letwin, C.R., 2018. A woman's place is in the… startup! Crowdfunder judgments, implicit bias, and the stereotype content model. J. Bus. Ventur. 33, 813–831.
Kagan, E., Leider, S., Lovejoy, W.S., 2017. Ideation–execution transition in product development: an experimental analysis. Management Science Forthcoming. https://doi.org/10.1287/mnsc.2016.2709. Available at:.
Katz, J.A., 2018. Doctoral Programs in Entrepreneurship. https://sites.google.com/a/slu.edu/eweb/doctoral-programs-in-entrepreneurship.
Kelley, K., Maxwell, S.E., 2003. Sample size for multiple regression: obtaining regression coefficients that are accurate, not simply significant. Psychol. Methods 8, 305–321.
King, G., Keohane, R.O., Verba, S., 1994. Designing Social Inquiry: Scientific Inference in Qualitative Research. Princeton University Press.
Kline, R.B., 2015. Principles and Practice of Structural Equation Modeling. Guilford publications.
Krasikova, D.V., Le, H., Bachura, E., 2018. Toward customer-centric organizational science: a common language effect size Indicator for multiple linear regressions and regressions with higher-order terms. J Appl Psychol. https://doi.org/10.1037/apl0000296. In press.
Kruschke, J.K., Liddell, T.M., 2018. Bayesian data analysis for newcomers. Psychon. Bull. Rev. 25, 155–177.
Lee, Y.S., 2018. Government guaranteed small business loans and regional growth. J. Bus. Ventur. 33, 70–83.
Loken, E., Gelman, A., 2017. Measurement error and the replication crisis. Science 355, 584–585.
Lumpkin, G.T., Dess, G.G., 1996. Clarifying the entrepreneurial orientation construct and linking it to performance. Acad. Manage. Rev. 21, 135–172.
MacKinnon, D.P., Pirlott, A.G., 2015. Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. Personal. Soc. Psychol. Rev. 19, 30–43.
Martin, B.C., McNally, J.J., Kay, M.J., 2013. Examining the formation of human capital in entrepreneurship: a meta-analysis of entrepreneurship education outcomes. J. Bus. Ventur. 28, 211–224.
McShane, B.B., Gal, D., 2016. Blinding us to the obvious? The effect of statistical training on the evaluation of evidence. Manage. Sci. 62, 1707–1718.
McShane, B.B., Gal, D., 2017. Statistical significance and the dichotomization of evidence. J. Am. Stat. Assoc. 112, 885–895.
Meyer, K.E., van Witteloostuijn, A., Beugelsdijk, S., 2017. What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. J. Int. Bus. Stud. 48, 535–551.
Morgan, S.L., Winship, C., 2007. Counterfactuals and Causal Analysis: Methods and Principles for Social Research. Harvard University Press, Cambridge.
Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., du Sert, N.P., Simonsohn, U., Wagenmakers, E.-J., Ware, J.J., Ioannidis, J.P., 2017. A manifesto for reproducible science. Nat. Hum. Behav. 1, 0021.
Murphy, K.R., Russell, C.J., 2017. Mend it or end it: redirecting the search for interactions in the organizational sciences. Organ. Res. Methods 20, 549–573.
Nuijten, M.B., Hartgerink, C.H., van Assen, M.A., Epskamp, S., Wicherts, J.M., 2016. The prevalence of statistical reporting errors in psychology (1985–2013). Behav. Res. Methods 48, 1205–1226.
O'Boyle Jr., E.H., Banks, G.C., Gonzalez-Mulé, E., 2017. The chrysalis effect: how ugly initial results metamorphosize into beautiful articles. J. Manag. 43, 376–399.
Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349, aac4716.
Pearl, J., 2009. Causality: Models, Reasoning and Inference, 2nd Ed. Cambridge University Press, Boston.
Rosenbusch, N., Brinckmann, J., Bausch, A., 2011. Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. J. Bus. Ventur. 26, 441–457.
Rubin, D.B., 2005. Causal inference using potential outcomes: design, modeling, decisions. J. Am. Stat. Assoc. 100, 322–331.
Salganik, M.J., 2017. Bit by Bit: Social Research in the Digital Age. Princeton University Press.
Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-experimental Designs for Generalized Causal Inference. Houghton-Mifflin, Boston.
Shankar, R.K., Shepherd, D.A., 2018. Accelerating strategic fit or venture emergence: different paths adopted by corporate accelerators. J. Bus. Ventur. https://doi.org/10.1016/j.jbusvent.2018.06.004.
Shepherd, D.A., 2015. Party On! A call for entrepreneurship research that is more interactive, activity based, cognitively hot, compassionate, and prosocial. J. Bus. Ventur. 30, 489–507.
Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22, 1359–1366.
Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. P-curve and effect size: correcting for publication bias using only significant results. Perspect. Psychol. Sci. 9, 666–681.
Sine, W.D., Mitsuhashi, H., Kirsch, D.A., 2006. Revisiting burns and stalker: formal structure and new venture performance in emerging economic sectors. Acad. Manag. J. 49, 121–132.
van't Veer, A.E., Giner-Sorolla, R., 2016. Pre-registration in social psychology—a discussion and suggested template. J. Exp. Soc. Psychol. 67, 2–12.
Tang, Y., Wezel, F.C., 2015. Up to standard?: market positioning and performance of Hong Kong films, 1975–1997. J. Bus. Ventur. 30, 452–466.
Van de Ven, A., Ang, S., Arino, A., Bamberger, P., LeBaron, C., Miller, C., Milliken, F., 2015. Welcome to the academy of management discoveries. Acad Manag Discov 1, 1–4.
Varian, H.R., 2016. Causal inference in economics and marketing. PNAS 113, 7310–7315.
Wasserstein, R.L., Lazar, N.A., 2016. The ASA's Statement on P-Values: Context, Process, and Purpose. Taylor & Francis.
Welter, F., 2011. Contextualizing entrepreneurship—conceptual challenges and ways forward. Enterp. Theory Pract. 35, 165–184.
Wiersema, M.F., Bowen, H.P., 2009. The use of limited dependent variable techniques in strategy research: issues and methods. Strateg. Manag. J. 30, 679–692.
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. J. Bus. Ventur. 34 (1-??).