Journal of Business Venturing 35 (2020) 105938

Contents lists available at ScienceDirect

Journal of Business Venturing

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

Editorial: Enhancing the exploration and communication of quantitative entrepreneurship research☆

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

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

ARTICLE INFO

Keywords:
Research design
Publishing in *JBV*
Exploratory research

ABSTRACT

The purpose of this editorial is to discuss ways to enhance exploratory quantitative studies in entrepreneurship. We use examples from entrepreneurship research and other scientific fields to illustrate the advantages of graphical data display for both exploratory purposes and post hoc tests. We provide suggestions for authors, reviewers, and editors on ways to enhance the transparency, accuracy, and pedagogical presentation of quantitative data in papers with the explicit purpose of illuminating emerging and important entrepreneurship phenomena. Our hope is that we spark a conversation among entrepreneurship scholars about the state of our empirical work and the possibilities that lie ahead to enhance exploratory entrepreneurship research.

1. Introduction

A hallmark of entrepreneurship research has always been the exploration of novel and important phenomena. With entrepreneurship constituting a fundamental engine of economic, social, and cultural change, entrepreneurship scholars should not feel bounded by disciplinary traditions or methodological dogmas prescribing how to analyze or present material.

However, the implicit tradition in much of management and broader social science research to rely on summaries of descriptive statistics, a regression model or two, and perhaps one or two graphical displays of marginal effects from a fitted statistical model dominates much of quantitative entrepreneurship research. Graphical displays of raw or even bivariate data are still rather uncommon, and many papers in our field still lack proper descriptive statistics, such as the minimum and maximum values of included variables. Such conservative data presentation and analysis diminishes the usefulness of a paper for a) researchers primarily interested in the phenomenon at hand and not the theory being tested; b) for theorists seeking to understand the specific conditions under which the theory was tested; and c) for the research community at large because only papers with detailed descriptive data are readily usable in future meta-analyses. However, with the rapid advancement in easy-to-use, and often free, data visualization tools, new opportunities exist to enhance the insights provided by both exploratory research papers and theory-testing papers.

☆ We are grateful for comments from Herman Aguinis, Anna Dreber Almenberg, Per Davidsson, Frederic Delmar, Timothy Folta, Keith Hmieleski, Alan Johnson, Reddi Kotha, Jeffery McMullen, 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. All opinions expressed in this editorial are solely ours and do not represent 'rules' for JBV submissions. Wennberg acknowledges funding from the Royal Swedish Academy of Letters, History and Antiquities.

* Corresponding author.
E-mail addresses: karl.wennberg@liu.se (K. Wennberg), andersonbri@umkc.edu (B.S. Anderson).

https://doi.org/10.1016/j.jbusvent.2019.05.002
Received 6 March 2019; Received in revised form 26 April 2019; Accepted 30 May 2019
Available online 07 June 2019
0883-9026/

In this editorial, we provide our perspective on improving the practice of exploratory entrepreneurship research. While some of the discussion and recommendations proposed in this paper are not unique to entrepreneurship research, our discussion and recommendations address empirical challenges specific to entrepreneurship scholarship that merit special attention. Entrepreneurship researchers frequently deal with datasets containing stochastic predictor and outcome variables in the same model (Coad et al., 2015b), complex selection effects and survivor biases (Delmar and Shane, 2004), and skewness that often does not fit well in traditional linear modeling (Crawford et al., 2015). With these unique features of entrepreneurship research in mind, we revisit and discuss some fundamental notions of exploratory research practices in entrepreneurship focusing on *visualizing* and *re-expressing data* (for exploratory purposes) and on ensuring data is *resistant* to outliers and specific distributional properties (for exploratory and theory-testing purposes). We also provide examples of post hoc analyses in theory-testing papers in which interesting patterns emerge as part of the data analysis process.

The advantages of more rigorous and creative exploratory analyses lie in both the potential to unearth new and novel insights not immediately apparent or foreseen, and to explicate the properties of raw data in more detail before subsequent analysis. We believe that embracing exploratory analysis, also an important precursor to theory-testing papers, increases research opportunities and provides entrepreneurship scholars with an expanded toolkit. Through exploratory analyses, entrepreneurship research gains additional momentum and impact by enhancing both the creativity and the detail with which we present and analyze the data in our research.

2. The value of exploratory entrepreneurship research

Allowing the publication of high-quality exploratory research is pivotal for research progress because not every paper can involve testing existing theory. On the contrary, the proliferation of quantitative work aiming to “extend” or “add” new theoretical understanding in the management sciences—often without causal or comparative testing of competing theoretical explanations—has frequently been lamented as one of the key reasons for the slow rate of cumulative research progress in the field (e.g., Pfeffer, 1993; Van de Ven et al., 2015) and, consequently, the unsatisfactory practical impact of our research (Economist, 2016). The transformative power of entrepreneurship to change the course of a single person, an industry, or an entire nation places substantive importance on the research questions we ask and, *by extension*, on the answers we provide to these questions. Entrepreneurship research is thus significantly broader and more ambitious in scope than merely seeking to test and extend theory (Shepherd, 2015). Some phenomena may not easily lend themselves to the testing of theoretical conjectures but are nevertheless important. Some phenomena are new, scarcely documented, or ill understood, necessitating careful exploratory analysis. Indeed, we cannot theorize very well about phenomena we do not have a good description of in the first place (Weick, 1995).

2.1. Defining exploratory research

Exploratory analysis is the hunch, the notion, the possibility, the idea on the back of the napkin that makes the researcher ask, “I wonder if…?” The value of exploratory analysis is to probe new areas that illuminate entrepreneurial phenomena without regard to offering a specific a priori reason for doing so (Ketchen Jr and Shook, 1996; Van de Ven et al., 2015).

Pure exploratory analyses do not test hypotheses since there is little entrepreneurship theory to explain the phenomenon at hand. Such exploratory analyses are particularly useful for identifying new, misunderstood, and underexplored research areas and for constructing new models to guide future empirical work. This type of exploratory analysis is similar but not identical to qualitative methodologies, which researchers use to inductively generate new theory and propositions. In many ways, well-done exploratory research offers the greatest a priori opportunity to make a useful contribution to entrepreneurship scholarship.

To be clear, exploratory analysis is **not** downloading a random dataset and engaging in a “hunt for asterisks” (Bettis, 2012). To explore means to venture into the unknown. However, to be useful, such discovery must occur at the field level; if other scholars have a good understanding of a phenomenon but an individual researcher does not, that researcher's own elucidation of the phenomenon does not rise to usefulness for the field. When a true paucity of understanding exists among entrepreneurship scholars, however, then exploratory analysis is an immensely valuable tool.

2.2. Key considerations for exploratory entrepreneurship research

The first focus of this editorial, then, is exploratory research designs that employ a quantitative methodology to illuminate a new, largely unexplored phenomenon. The second focus is exploratory research designs that researchers use to illustrate and validate data before testing associated theoretical conjectures. In the former, the researcher uses various analyses in an exploratory study the outcome of which is ideally a standalone paper, for example, Mollick's (2014) exploratory study of crowdfunding. In the latter, the researcher uses exploratory analyses alongside other techniques to test a pre-defined hypothesis, for example, illustrating raw data against tested predictions of the hypothesis. However, in theory-testing work, exploratory analyses that seek to validate data before theory testing are arguably more complex and in need of careful delineation from the theory testing analysis. This is because while exploration is necessary prior to knowing what a dataset looks like and how it could be analyzed, exploration can easily slip into implicit pre-testing and changes being made to the theoretically argued hypotheses simply based on the patterns found in the exploratory analysis. Also called HARKing—hypothesizing after results are known—this practice can lead to hypotheses being fitted to the data rather than vice versa.[1] Such problems can also easily occur when reviewers and editors ask authors to add and/or change their original hypotheses, which should be discouraged (Anderson et al., 2019).

Regardless of whether a paper focuses predominately on exploration or on theory testing, there are several noteworthy aspects of exploratory data analysis relevant for entrepreneurship research in which data is frequently messy, noisy, and strangely distributed (Shepherd, 2015). For example, the traditional descriptive techniques propagated by Tukey (1980) and others, such as histograms, bar charts, boxplots, and scatterplots, are still rather uncommon in entrepreneurship research but remain immensely useful some 40 years after Tukey's recommendations. Today, researchers can easily combine such descriptive techniques with newer types of visualizations using easy-to-use and often freely available software packages. These packages provide readily interpretable visualizations that can help identify unexpected patterns, data anomalies, and selection of statistical tools for multivariate analyses. In the following section, we outline several types of exploratory research and focus on data visualizations as a key element of exploratory analyses. We then describe the role of model fitting and data description in the exploratory research process.

3. Advancing exploratory entrepreneurship research

3.1. Types of exploratory research

We would like to distinguish between exploratory analysis, which we refer to here, and qualitative methodologies, which can also generate new theory and propositions. Qualitative methodologies often accomplish the same general objective as our depiction of exploratory analysis, and *JBV*, like its peer entrepreneurship journals, encourages high-quality qualitative research. Qualitative papers, however, may include quantitative data and may thus equally benefit from careful exploration and descriptive visualizations that summarize and visualize research observations, verbal data, and timelines, in addition to traditional qualitative analyses. Muñoz et al.'s (2018) process-tracing analysis of sustainable venturing provides a recent, illustrative example.

For our purposes, exploratory quantitative analysis involves a range of tools, including descriptive statistics, cluster analysis, fuzzy-set QCA, Bayesian exploratory modeling,[2] various visualizations and other methodologies to identify patterns in data that merit further research (e.g., Cohen, 1994; Hollenbeck and Wright, 2017; Mahoney et al., 2013). Exploratory data analysis can—and should—be both descriptive and multivariate. Regardless of the actual tools used, Tukey's (1980) four classical suggestions for exploratory data analysis remain highly laudable: (1) *visualize* data patterns and the structure of the analyses to be conducted; (2) analyze *residuals* to explain what remains of the data after analysis; (3) *re-express* variables to simplify analyses using mathematical functions, such as the logarithm and the square root; and (4) ensure data *resistance* by examining outliers.

When it comes to multivariate forms of exploratory data analysis, classical tools for dimensionality reduction, such as multidimensional scaling and principal component analysis, have potential far beyond the construction of psychometrically measured latent variables, such as exploring large combined datasets drawn from text archives, social networks, or trade statistics. Further, Bayesian approaches offer substantial advantages to deal with very large datasets or data with substantial missing observations and noisy measures. There are also an increasing number of software packages for exploratory data analysis based on machine learning techniques, allowing scholars to fit optimal models using data without prior knowledge of theoretically expected relationships. Such analyses can lay the groundwork for inductive theorizing in exploratory papers and for the identification of plausible moderators in theory-testing papers (Aguinis et al., 2013a). Further, useful exploratory tools for post hoc analyses in theory-testing papers include Monte Carlo simulations and other tools that focus on data resistance or uncertainty in estimated variables.

We believe that the recent growth of "big data" sources lend themselves particularly well to exploratory analysis (e.g., Ng and Stuart, 2016). First, in both exploratory and theory-testing quantitative analysis, there is an increasing demand for datasets large enough to establish good power and allow for other best practices such as hold-out samples (Andersson et al., 2019). Also, a principal reason we suggest exploratory analysis and not theory-testing analysis for big data is that teasing out causal relationships from sources like governmental tax registers, Google searches, Tweets, Facebook posts, and Kickstarter campaigns is inordinately difficult (Angrist and Pischke, 2008; Street et al., 2015). Still, exploring large datasets and modeling relationships using, for example, cluster analysis, topic modeling, and network or diffusion analyses may yield interesting insights that provide unique value without showing causal evidence. Such "'Big Data'…entail a need for greater presentation of the methods for generating and pre-processing data" (Greve, 2018, p. 424).

3.2. Data visualization in exploratory entrepreneurship research

An overlooked aspect of entrepreneurship research is the usefulness of data visualization. New tools for summarizing complex and high-dimensional data make it easier for researchers to communicate key insights and findings (Moody and Healy, 2014). Exploratory research—particularly studies that combine multiple, potentially disparate databases—benefits from visualizations that help "tell the

[1] Most scholars are aware of the dangers of adapting their theory and research design during exploration so that one samples on statistically significant results. For example, by running a few "exploratory" regressions models, implicitly or explicitly looking for low *p*-values for fitted coefficients, perhaps adding interaction terms and/or nonlinear terms to those coefficients, and finally writing a contingency-type story focusing primarily on those interaction terms with low *p*-values. There is a great deal written on the dangers of "research degrees of freedom" in such analyses (Gelman and Loken, 2013) and on the "pseudoscientific practice" that results (Greve, 2018).

[2] Of course, researchers may use several of these methods for theory-testing research.

story" of the data. As recently noted by entrepreneurship and management scholar Henrich Greve, "showing the data advances science because significant but weak findings can be ignored, unexplained variation can be analyzed, and new ideas can be generated from a dataset by all readers of a paper, not just the authors" (Greve, 2018, 424). We encourage entrepreneurship researchers with exploratory studies to make data visualizations the central communication tool of their papers and researchers with theory-testing studies to use visualizations to illustrate both their raw data and their predictions for their hypotheses. Including multiple visualizations (charts, graphs, images, and interactive graphics posted online) give life to a paper, enhance interpretation of results, and reinforce the importance of the focal topic to entrepreneurship scholars. Below, we offer some examples of such research.

3.2.1. *Univariate visualizations*

Researchers commonly use univariate plots to show the distribution of key variables, providing key considerations for subsequent model selection. Such plots may be especially relevant in entrepreneurship research with high variation and potential non-linearities in key variables. A frequent variable in entrepreneurship research is "growth intention" or "growth aspirations," often operationalized as expected growth in sales or employees. Such a variable could be differentially operationalized as a count (number of staff or increase in staff from today), or as a ratio (difference in number of staff at a point in the future compared to today), or could be re-expressed in logarithmic (for counts or the sum of the ratio) or log-log forms (for the numerator and denominator of the ratio) (e.g., Autio and Acs, 2010; Estrin et al., 2013; McKelvie et al., 2017). The specific operationalization may naturally have a large influence on model specification and the results obtained, and investigations of distributions beyond tabulations provide important information.

For example, Crawford et al. (2015, p. 706) plotted founders' expected number of employees after 12 months of operations from Wave 1 in the CAUSEE[3] dataset, showing a highly skewed figure in which most founders expect to employ between 0 and 10 employees (the leftward "spike" in Fig. 1), some expect to employ between 10 and 100 employees, and a small number expect to employ several hundred or even thousands of employees. Such plots were essential both to determine model specification (e.g., as Crawford and colleagues argue, a power-law model) and to consider whether those expecting to employ several thousand employees after 12 months of operations were outliers that should be handled differentially in subsequent modeling.

3.2.2. *Bivariate visualizations*

Visualizations are also ideal for showing bivariate patterns such as the correlation a key phenomenon/outcome variable of interest and one or more explanatory variables. Such evidence is worthwhile both in purely exploratory papers and in theory-testing papers as part of showing "the raw effect" of some key predictor been introduced. Visualizations of bivariate evidence should include margins of uncertainty (e.g., confidence intervals). We provide an example in Fig. 2 below, which shows the (bivariate) survival rates of new ventures engaged in innovativeness and those that are not displayed as bar charts with confidence intervals marked as black vertical lines (from Hyytinen et al., 2015, p. 573). Such descriptions provide tentative evidence of important variation that can be subsequently theorized and modeled in theory-testing papers, such as in the paper by Hyytinen et al. (2015), or discussed as important novel findings in exploratory papers.

Bivariate plots are also useful in post hoc analyses when interesting patterns emerge as part of data analysis. An example is Mollick's (2014, p.10) study on crowdfunding. After describing various crowdfunding projects on Kickstarter and modeling predictors of crowdfunding success, Mollick plotted pie charts of project success and failure across US metropolitan statistical areas (MSAs). Mollick's map, reproduced in Fig. 3 below, shows different-sized pie charts according to the number of projects launched in the MSAs, with the light green areas representing the fraction of unsuccessful projects and the dark green areas representing the fraction of successful projects. It is clear from the map that success ratios vary highly across MSAs and that MSAs with more projects seem to have higher success ratios.

Finally, researchers may use contour plots and similar figures to illustrate bivariate and multivariate results, such as correlations between a variable over time or between different variables in a dataset. Such plots are analogous to raw correlations (for two variables), but with the advantage that both researchers and readers can easily estimate *where* in the distribution of the two variables the correlation resides. An example—reproduced in Fig. 4—is Coad et al.'s (2015a) study of firm growth, which clearly highlights (1) the high frequency in the data of firms that are neither growing nor declining (the black area), (2) the tendency of "growth spurts" or "regression toward the mean" indicating that firms that grew or declined in one period tend to remain the same in the subsequent period (the horizontal pattern to the left and right of the black center), and (3) gradually sparser observations of firms' growth or decline in one period and their growth or decline in a subsequent period (the light grey and white areas).[4]

Given that many readers may not be familiar with how to construct visualizations, we encourage researchers to post code and descriptions for creating the visualizations presented in a paper or in online repositories like ResearchGate, Github or the Open Science Framework. Excellent visualization tools are also available in most software packages, such as SPSS, SAS, and Stata, and are freely available in open-source packages, such as R. Google Charts also offers users various visualization options, including advanced flowcharts, such as Sankey diagrams.

Researchers can also take advantage of innovative data-plotting practices from other disciplines. The lion's share of visualizations

[3] "Comprehensive Australian Study of Entrepreneurial Emergence".

[4] In Figure 4, the colored areas correspond to the frequency of the number of firms in each growth-rate quadrant categorized into logarithmic bins. The black quadrant thus illustrates that the number of firms experiencing no growth or slow growth at both t-1 and t represent most of the data compared to firms experiencing rapid decline or growth in sales.

Fig. 1. Founder's expected number of employees after 12 months of operations, Wave 1 in the CAUSEE dataset (Crawford et al., 2015).
Note: Reproduced from *Journal of Business Venturing* 2015, 30, p. 706.

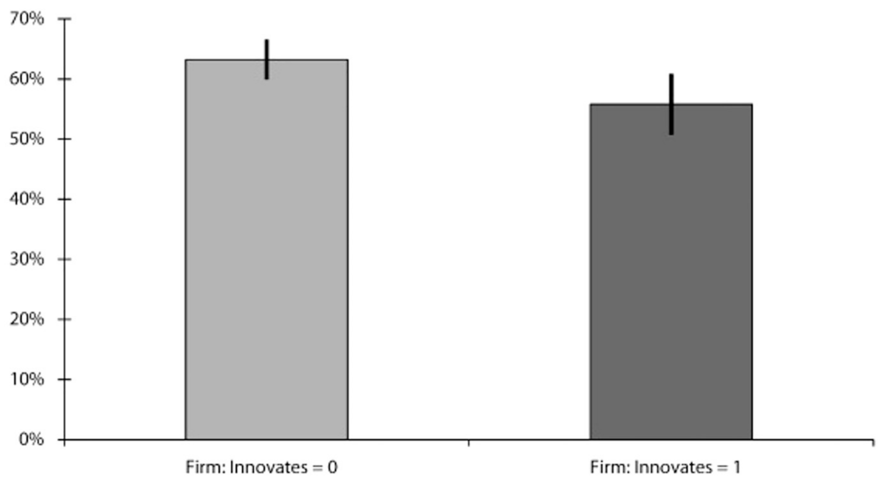

Fig. 2. Description of new ventures' survival rates (Hyytinen et al., 2015).
Note: Reproduced from *Journal of Business Venturing* 2015, 30, p. 573.

used in entrepreneurship research as well as in the broader social sciences still focuses on describing static statistical relationships between variables (Moody and Healy, 2014). As Tufte (1997, p. 10) explains, for visualizations to be *explanatory*, they must be "about the representation of mechanisms and motion, of process and dynamics." Multilayered visualization methods in the form of maps (Vilhena et al., 2014), graphs of diffusion processes (Gibbons, 2004), differing strengths of causal effects (Börner et al., 2018), and plots of estimated causal effects under differing conditions (Alabdulkareem et al., 2018) are beginning to appear in the social sciences. We foresee much development in this area and the diffusion of ideas and methods from other disciplines. We provide a recent example in Fig. 5. This rather detailed visualization shows Börner et al.'s (2018) analysis of whether skills mentioned in research publications are "Granger-causing" skills advertised in job postings or vice versa. The arrows in Fig. 5 represent skills with significant Granger causality (P value < 0.05). The direction and thickness of each arrow indicate the F-value strength and direction, highlighting skills in "Immunology" and "Facebook" as key skills related to the advertised jobs in the data.

3.3. Explanation versus prediction and the role of model fitting

If flexibility and freedom from theoretical and dogmatic paradigms are the virtues of exploratory analysis, they are also its limitations. Because the observed relationships are purely a function of the data and not an existing theoretical expectation, with any kind of exploratory work, there is the very real possibility that the identified effect is purely an artifact of the data and sample (Meyer et al., 2017; Simmons et al., 2011). It is also impossible with exploratory analysis to eliminate alternative explanations for the

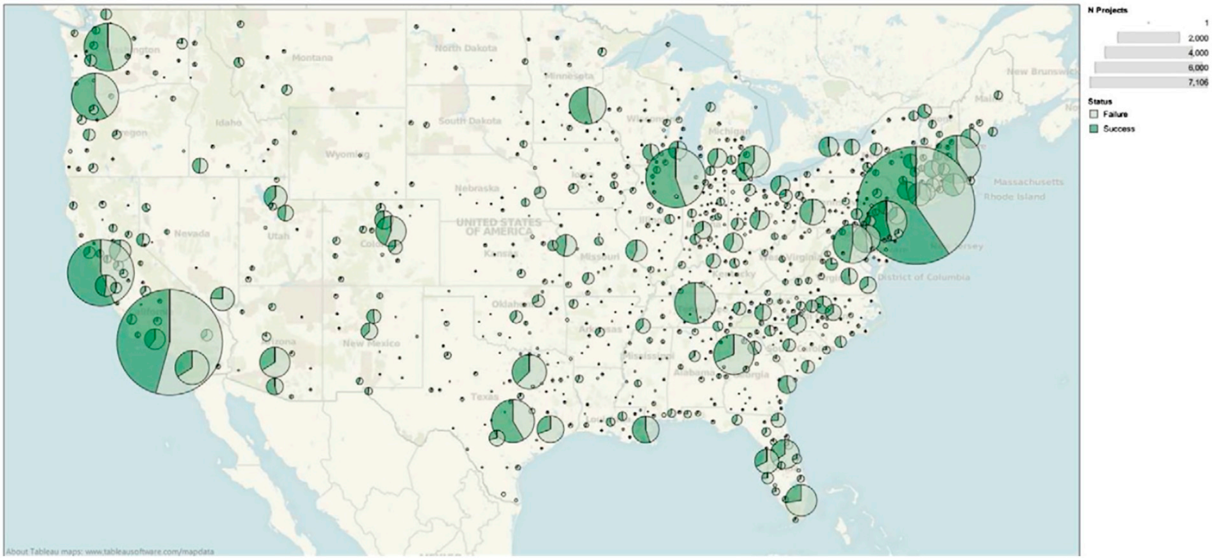

Fig. 3. The geography of crowdfunding (Mollick, 2014).
Note: Reproduced from *Journal of Business Venturing* 2014, 29, p. 10.

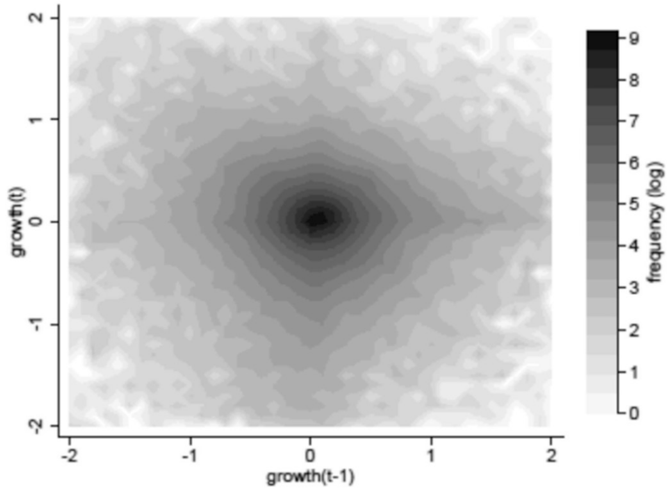

Fig. 4. Contour plot for growth rates (from Coad et al., 2015a, 2015b, p. 18).
Note: Reproduced from Coad et al., 2015a, HUI Working paper, 112.

observed relationships because unknown variables likely underlie the identified, but tentative, research model (Rubin, 1974, 2005). For these reasons, transparency in analysis and disclosure of intent are critical to demonstrating statistical conclusion validity and to minimizing the impact of researchers' degree of freedom in subsequent theory-testing research (Anderson et al., 2019).

Scholars using machine learning algorithms highlight that there is often a valid tradeoff between causal evidence and prediction accuracy (Salganik, 2017). While theory-testing quantitative social science research seeks to refute one explanation in favor of another in explaining what caused an outcome of interest, knowledge of the past may not yield strong predictions of the future. Hold-out samples (within-study replications), comparisons with other studies, simulations, and a focus on having more accurate measures rather than maximizing explained variance may lead to better predictive models. Many entrepreneurs, and perhaps researchers, are more concerned with prediction accuracy—"throw mud against the wall and see what sticks"—than a precise causal explanation for a specific action or strategy (e.g., whether a business plan leads to a higher likelihood of funding or not).

In these types of papers, the model—the collection of variables and the estimator used to derive the prediction—takes center stage. However, in exploratory work, there may be multiple valid estimation approaches; least squares, structural equation modeling (SEM), dynamic panel models, and so forth. There may also be multiple variable operationalizations and transformations appropriate for the data and for the estimator. The key consideration is to disclose why the researcher chose which models and variables, and the results of those models. For no other reason, these disclosures help future scholars avoid analytical "dead ends" that consume time and resources. Further, in model fitting, there are numerous approaches to deal with missing data and outliers. The method itself

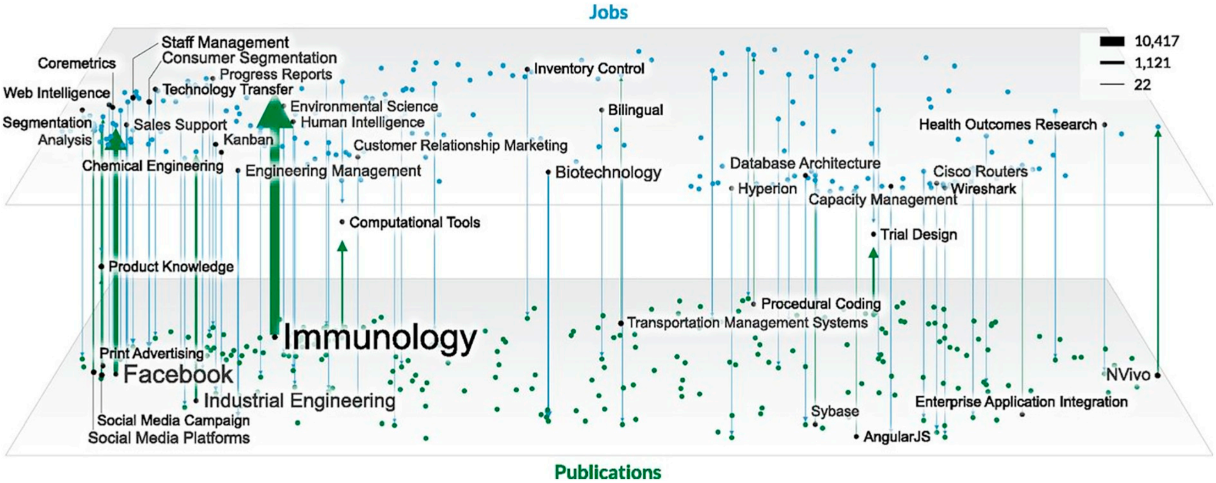

Fig. 5. Strength of influence mapping among skills mentioned in research publications and skills mentioned in job advertisements (Börner et al., 2018).
Note: Reproduced from Börner et al., 2018; PNAS, 115:12630-12637.

(e.g., list-wise deletion, imputation, and so forth) is not as important as clearly disclosing the decision rules and the impact of those decision rules on the reported results.[5] As is apparent from Fig. 1, outliers are frequent in entrepreneurship datasets, and whether those outliers are included or excluded or whether the data is re-represented by variable transformation may have a sizeable impact on the results obtained in multivariate testing.

While exploratory research is not able to make causal claims, a valuable approach for evaluating the predictive accuracy of these types of studies is to use ex ante hold-out samples. Such samples are particularly useful for “big data” studies, where the data may reveal complex interactions, but the results are more likely to be idiosyncratic to the sample itself. With a hold-out sample, the researcher randomly splits the data into two or more new samples ideally before conducting the initial analyses. The researcher uses one sample to explore the data and build a predictive model and then uses the additional samples for model validation. To be clear, in the first sample, p-hacking is fine assuming appropriate disclosure, but p-hacking would not be appropriate in the validation samples.

3.4. Data description and visualization in theory-testing entrepreneurship research

Theory-testing papers benefit from transparent and explicit exploratory analyses either as part of the data description or as part of post hoc investigations of patterns unearthed during the theory-testing stage.[6] Many data analysis texts recommend that researchers create extensive descriptive and exploratory plots before undertaking more advanced multivariate modeling (Singer and Willett, 2003). Such exploratory data plotting data can reveal general patterns, provide insight into the functional form of variables, and identify outliers, all of which are important for subsequent theory-testing research.

Researchers should employ exploratory analytical techniques in theory-testing entrepreneurship research because many key variables (e.g., entrepreneurial growth aspirations, crowdfunding campaigns, and other resource-mobilizing efforts) and entrepreneurial outcome measures may be highly skewed. Further, the development of very early-stage ventures may be difficult to detect even in detailed datasets due to restricted ranges in continuous variables (e.g. revenue and employee count), or in innovation outputs (e.g. patents or new products) that are often at or near zero (McKelvie et al., 2017). In a recent paper, Johnson et al. (2018) recommended using descriptive plots for such outcomes and ideally combining the use of both continuous and discrete outcome variables (1) to get to know data before using more sophisticated multivariate modeling; and (2) to provide tentative evidence of whether similar results may be expected for continuous or discrete outcome variables.

As with univariate and bivariate visualizations, researchers construct visualization plots for multivariate results when two or more factors are jointly assessed to illustrate or explain some outcome. An example of the latter approach is Jara-Figueroa et al.'s (2018) paper on the growth and survival of all startups in Brazil (reproduced in Fig. 6). This paper highlighted that *new venture members'* industry- and occupation-specific knowledge strongly predict venture growth and survival but that these associations are more pronounced under specific combinations of industry- and occupation-specific knowledge, observable as the dark blue squares in Panels A, B, and D of Fig. 6.

[5] In our opinion, this disclosure extends beyond the sample reported in a published paper to decisions made in research design and data collection. We also recommend similar disclosures for range restrictions made on the data (e.g., excluding firms of specific age or size) during the design or analysis stage.

[6] We strongly encourage an explicit separation of data exploration and theory testing—for example, by pre-registering hypotheses before data analyses—to avoid the risk of tailoring hypotheses to emerging data patterns and to avoid the risk of hypothesizing after results are known (HARKing). For a more thorough discussion and examples of free easy-to-use pre-registration platforms, see Anderson et al. (2019).

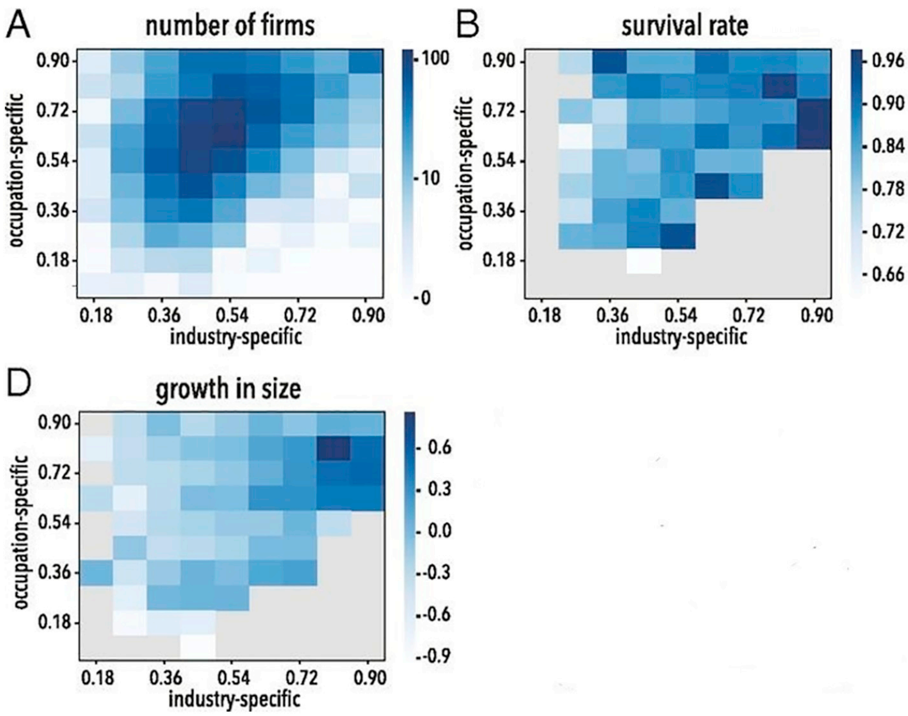

Fig. 6. Multivariate plots of *new venture members'* industry- and occupation-specific knowledge for (A) number of firms, (B) new venture survival, and (D) new venture growth (Jara-Figueroa et al., 2018).
Note: Reproduced from Jara-Figueroa et al., 2018; PNAS, 115:12646-12653.

To the benefit of researchers and our field, new data visualizations and statistical tools add multiple ways to uncover and illuminate the nuance and contextual differences so common to entrepreneurship phenomena. However, like all tools, there is the potential for misuse and abuse. In the concluding section, we offer a series of best practices and recommendations for exploratory research to avoid these pitfalls.

4. Some practices for exploratory research

Critical to the contribution of any exploratory analysis is clearly communicating that the purpose of the analysis was to explore a new phenomenon and thereby generate new insights, and not to test theory if the latter follows the former in the same study. This helps ensure that researchers are not engaging in HARKing—generating hypotheses after conducting statistical analyses—a particularly troubling threat to a study's validity (Bosco et al., 2016). Researchers writing exploratory papers should embrace the researcher degrees of freedom used to find the study's insights and be open about it. While the specific disclosure for a given paper is a function of the data and methods used for the analysis, the following sample disclosure statement is a useful starting point for authors to include:

> We present these results as exploratory analyses. In the process of identifying these relationships, we used multiple statistical techniques, made multiple comparisons, and exercised our judgement in the analysis process to best reveal the phenomenon. We caution readers not to infer causality from our model.

Entrepreneurship scholars seeking to cover new ground with exploratory-oriented papers need to carefully clarify *why* and *how* their exploratory analysis moves the field forward, ideally providing examples of gaps in the field that are addressed by the exploratory analyses. We also encourage authors of exploratory papers to be explicit about their paper's purpose and contribution in the cover letter to the editor accompanying the paper submission. Greater clarity about why the field benefits from exploring the phenomenon under study enables the editor to select reviewers who are knowledgeable about the research question and who have experience with exploratory methodologies.

We also strongly encourage authors of exploratory papers to take advantage of online code and data repositories to allow reviewers and readers to follow along with the analyses presented in these papers, understand the assumptions made, and see all the manipulations (e.g., variable transformations) made to the data. When working with proprietary data that cannot be publicly released, the codebooks and code used for analyses could and should be publicly available. Editors and reviewers should judge the usefulness of an exploratory paper by how well the paper empowers researchers to ask new questions or to explore existing questions using a new perspective or method. Such papers are necessarily forward looking and enable future research.

4.1. Replication and transparency in exploratory analysis

With its rapid exploration of new phenomena and research domains, entrepreneurship research is increasingly cumulative in nature and is maturing as a field (Landström et al., 2012). To take advantage of our ability to "stand on the shoulders of others," and being mindful of future generations of scholars seeking to build on the work we produce, we encourage exploratory analyses—-whether in an exploratory paper or as part of a theory-testing paper—to provide sufficient information for reproducibility. The purpose of requiring descriptive statistics is to give other researchers the ability to reproduce the results presented in a paper (Bergh et al., 2017). Cross-sectional descriptive statistics may be difficult to interpret with the type of longitudinal data that represents an increasingly large share of all published entrepreneurship research. We thus encourage authors with longitudinal (multilevel) datasets to examine and show both cross-sectional and time-series (nested) correlations. We also encourage authors using latent variable structural equation modeling to display full indicator-level covariance matrices. Further, whenever possible, we strongly recommend making the full data and code available post-publication, which is also an ideal way to enhance the long-term relevance of a research paper.

We also urge scholars to enhance both exploratory and theory-testing analyses by describing their data using summary statistics beyond means and standard deviations. Importantly, scholars should also show the maximum and minimum values for all variables in the data to identify outliers, and for authors and readers to be able to calculate resampling statistics such as jackknife and bootstrap tests.[7] Outliers could represent errors in need of correction (e.g., implausible values), but could also represent influential data points that may need to be trimmed or excluded in robustness tests if the emphasis is on gauging the average association between predictor and outcome (Aguinis et al., 2013b). Further, because entrepreneurship research frequently has highly skewed data (e.g., innovation or growth rate), outliers may represent observations of interest worthy of specific investigation. For such phenomena, descriptions of the median and quartiles of variables allow readers to better assess the potential skewness or heavy-tailed distributions common in entrepreneurship research (Crawford et al., 2015).

4.2. Summary of recommendations

By no means does this editorial contain an exhaustive list of considerations for exploratory analyses. We do think, however, that entrepreneurship scholars benefit from a renewed emphasis on exploratory analysis, and that the suggestions and examples provided in this editorial serve as important considerations for these analyses. Below, we summarize our key recommendations for exploratory entrepreneurship research.

4.2.1. Describe data using both figures and plots

The value added of graphical displays, such as box-plots, increases when researchers measure variables on different scales. For example, a standard correlation measure for variables on a continuous or interval scale, such as the Pearson's correlation coefficient (r), is different from correlation measures for ordinal or binary data. Ordinal-scaled variables are commonly correlated using the Spearman's correlation coefficient (rho), whereas researcher should evaluate correlations between an interval-scaled variable and a binary variable with a point-biserial correlation (r_{pb}). Many datasets contain a mixture of variables measured on different scales, and in such instances, raw correlations may be less useful as an inspection tool. As such, we encourage researchers to supplement correlation matrixes with an appropriate visualization (e.g., Jara-Figueroa et al., 2018). We also suggest that in advance of standard data inspection before multivariate analysis—seeking to detect coding errors and uncover outliers—entrepreneurship researchers would benefit from employing plotting procedures to compare observation scatterplots with average patterns in key variables. Researchers should describe and ideally provide these analyses in a paper or in supplemental online material, to help readers understand the functional form of relationships in the data and why specific variables may need to be re-expressed or transformed (Hoaglin, 2003).

4.2.2. Distinguish pure exploratory papers from exploratory-before-theory-testing papers

While any quantitative-oriented study benefits from exploratory analysis, we strongly encourage the explicit separation of "purely" exploratory research and exploratory research done in the initial stages of or as post hoc tests for theory-testing papers. As preparation for theory-testing research, exploratory data plotting can help reveal general patterns, identify outliers, and provide insight into the functional form of variables. Classical approaches, such as John Tukey's four suggestions, are equally useful in both types of papers: (1) *visualize* data patterns and the structure of the analyses to be conducted, (2) analyze *residuals* to explain what remains of the data after analysis, (3) *re-express* variables to simplify analyses, and (4) ensure data *resistance* by examining outliers. Researchers may even seek to combine approaches by deriving hypotheses from an exploratory analysis of observational data and then testing those hypotheses with an experimental design or with a new random sample.

4.2.3. Explore time dimensionality

In addition to cross-sectional variation, variation over time is useful and valuable to explore and display. Sometimes interesting patterns in a phenomenon of interest—as well as in predictor variables potentially used to explain variation in that

[7] These standard tools are especially useful in small-N studies but researcher should interpret these results cautiously if the distribution is non-normal, such as in power-law distributions (Crawford et al., 2015).

phenomenon—fluctuate over time. Further, simple summary statistics and correlations may conceal an important time-trend, that a visualization would help reveal. Researchers should therefore plot these variables over time to examine the existence of such patterns. We also encourage authors with longitudinal (multilevel) datasets to be mindful of examining and showing both cross-sectional and time-series (nested) correlations.

4.3. Considerations for reviewers and editors

Lastly, we think it is worthwhile to discuss the usefulness of our recommendations for reviewers and editors. Beyond the recommendations we offer to authors, which reviewers could use as a starting point to evaluate exploratory work, we simply want to encourage reviewers and editors to be open to exploratory studies. These papers will necessarily look and feel different than theory-testing papers. Rarely, if ever, would these papers offer hypotheses, and the common standard of "making a theoretical contribution" does not apply in this case. These papers are harder to write and harder to review in several ways because the creativity necessary to craft exploratory research does not lend itself to the formalism found in theory-testing work. However, as we can see from the many citations that many exploratory papers in entrepreneurship receive, exploratory work benefits both authors and the field.

We do not mean to suggest that these papers should have a lower bar in the review process—far from it. Further, exploratory papers are not inherently devoid of theory. Ideally, exploratory papers highlight critical gaps in entrepreneurship theory or illustrate contexts in which our current theoretical understanding does not apply. Following Tukey's (1962) classic recommendation for reaping benefits from careful data exploration, we recommend that reviewers start from the perspective of *usefulness*—does the paper offer new insights, new ideas, or new ways of looking at an entrepreneurship phenomenon that other scholars will find useful (Anderson et al., 2019)? As Tukey (1962, pp. 5–6) argued half a century ago, there is little scientific novelty among scientists working on familiar problems, using familiar analytical frameworks, and using familiar processes for data analysis. For progress to be made, researchers must address new problems, address old problems using new or different analytical frameworks, and summarize and analyze new and old data in new ways.

5. Conclusion

As we look to accelerate entrepreneurship research, we need to be mindful that a central facet of the field has always been the exploration of novel and important phenomena. We suggest it is time to go beyond the standard practice of merely relying on summaries of descriptive statistics and graphical displays of marginal effects from fitted statistical models. Much of entrepreneurship research can benefit from additional explorative analysis to visualize and re-express data, ensuring data is resistant to outliers and specific distributional properties, and can also advance the use of explorative post hoc analyses after testing theoretical propositions. The advancement of easy-to-use data visualization tools also provides new opportunities to enhance the insights provided by both exploratory and theory-testing entrepreneurship research. In short, entrepreneurship scholarship has much to gain—and little to lose—by embracing and encouraging exploratory research.

References

Aguinis, H., Forcum, L.E., Joo, H., 2013a. Using market basket analysis in management research. J. Manag. 39, 1799–1824.
Aguinis, H., Gottfredson, R.K., Joo, H., 2013b. Best-practice recommendations for defining, identifying, and handling outliers. Organ. Res. Methods 16, 270–301.
Alabdulkareem, A., Frank, M.R., Sun, L., AlShebli, B., Hidalgo, C., Rahwan, I., 2018. Unpacking the polarization of workplace skills. Sci. Adv. 4, eaao6030.
Anderson, B., Wennberg, K., McMullen, J., 2019. Editorial: Enhancing quantitative theory-testing entrepreneurship research. J. Bus. Ventur (in press). https://www.sciencedirect.com/science/article/abs/pii/S0883902619300886.
Angrist, J.D., Pischke, J.-S., 2008. Mostly Harmless Econometrics: An empiricist's Companion. Princeton university press.
Autio, E., Acs, Z., 2010. Intellectual property protection and the formation of entrepreneurial growth aspirations. Strateg. Entrep. J. 4, 234–251.
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.
Bettis, R.A., 2012. The search for asterisks: compromised statistical tests and flawed theories. Strateg. Manag. J. 33, 108–113.
Börner, K., Scrivner, O., Gallant, M., Ma, S., Liu, X., Chewning, K., et al., 2018. Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. Proc. Natl. Acad. Sci. 115, 12630–12637.
Bosco, F.A., Aguinis, H., Field, J.G., Pierce, C.A., Dalton, D.R., 2016. HARKing's threat to organizational research: evidence from primary and meta-analytic sources. Pers. Psychol. 69, 709–750.
Coad, A., Daunfeldt, S.-O., Halvarsson, D., 2015a. Bursting into life: firm growth and growth persistence by age. HUI working paper. 112. https://swopec.hhs.se/huiwps/abs/huiwps0112.htm.
Coad, A., Frankish, J.S., Roberts, R.G., Storey, D.J., 2015b. Are firm growth paths random? A reply to "Firm growth and the illusion of randomness". Journal of Business Venturing Insights 3, 5–8.
Cohen, J., 1994. The earth is round ($p < .05$). Am. Psychol. 49, 997–1003.
Crawford, G.C., Aguinis, H., Lichtenstein, B., Davidsson, P., McKelvey, B., 2015. Power law distributions in entrepreneurship: implications for theory and research. J. Bus. Ventur. 30, 696–713.
Delmar, F., Shane, S., 2004. Legitimating first: organizing activities and the survival of new ventures. J. Bus. Ventur. 19, 385–410.
Economist, 2016. Management Theory is Becoming a Compendium of Dead Ideas. Economist (Dec 17th).
Estrin, S., Korosteleva, J., Mickiewicz, T., 2013. Which institutions encourage entrepreneurial growth aspirations? J. Bus. Ventur. 28, 564–580.
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. Working paper, Columbia University.
Gibbons, D.E., 2004. Network structure and innovation ambiguity effects on diffusion in dynamic organizational fields. Acad. Manag. J. 47, 938–951.
Greve, H.R., 2018. Show me the data! Improving evidence presentation for publication. Manag. Organ. Rev. 14, 423–432.
Hoaglin, D.C., 2003. John W. Tukey and data analysis. Stat. Sci. 311–318.
Hollenbeck, J.R., Wright, P.M., 2017. Harking, sharking, and tharking: making the case for post hoc analysis of scientific data. J. Manag. 43, 5–18.
Hyytinen, A., Pajarinen, M., Rouvinen, P., 2015. Does innovativeness reduce startup survival rates? J. Bus. Ventur. 30, 564–581.

Jara-Figueroa, C., Jun, B., Glaeser, E.L., Hidalgo, C.A., 2018. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. Proc. Natl. Acad. Sci. 115, 12646–12653.
Johnson, A.R., Masyn, K.E., McKelvie, A., 2018. Descriptive and Exploratory Plots of Growth and Survival Outcomes for New Ventures: A Contingent Approach to Incubation in an Ecosystem Context. Working Paper. Nord University.
Ketchen Jr., D.J., Shook, C.L., 1996. The application of cluster analysis in strategic management research: an analysis and critique. Strateg. Manag. J. 17, 441–458.
Landström, H., Harirchic, G., Åström, F., 2012. Entrepreneurship: exploring the knowledge base. Res. Policy 41, 1154–1181.
Mahoney, J., Goertz, G., Ragin, C.C., 2013. Causal models and counterfactuals. In: Handbook of Causal Analysis for Social Research. Springer, pp. 75–90.
McKelvie, A., Brattström, A., Wennberg, K., 2017. How young firms achieve growth: reconciling the roles of growth motivation and innovative activities. Small Bus. Econ. 49, 273–293.
Meyer, K.E., van Witteloostuijn, A., Beugelsdijk, S., 2017. What's in ap? Reassessing best practices for conducting and reporting hypothesis-testing research. J. Int. Bus. Stud. 48, 535–551.
Mollick, E., 2014. The dynamics of crowdfunding: an exploratory study. J. Bus. Ventur. 29, 1–16.
Moody, J.W., Healy, K., 2014. Data Visualization in Sociology. Annu. Rev. Sociol. 40, 105–128.
Muñoz, P., Cacciotti, G., Cohen, B., 2018. The double-edged sword of purpose-driven behavior in sustainable venturing. J. Bus. Ventur. 33, 149–178.
Ng, W., Stuart, T.E., 2016. Of Hobos and Highfliers: Disentangling the Classes and Careers of Technology-Based Entrepreneurs. Unpublished Working Paper. .
Pfeffer, J., 1993. Barriers to the advance of organizational science: paradigm development as a dependent variable. Acad. Manag. Rev. 18, 599–620.
Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66, 688.
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.
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.
Singer, J.D., Willett, J.B., 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, New York.
Street, A., Murray, T.A., Blitzer, J., Patel, R.S., 2015. Estimating voter registration deadline effects with web search data. Polit. Anal. 23, 225–241.
Tufte, E.R., 1997. Visual and statistical thinking: displays of evidence for making decisions. Graphics Cheshire.
Tukey, J.W., 1962. The future of data analysis. Ann. Math. Stat. 33, 1–67.
Tukey, J.W., 1980. We need both exploratory and confirmatory. Am. Stat. 34, 23–25.
Van de Ven, A.H., Ang, S., Arino, A., Bamberger, P., LeBaron, C., Miller, C., et al., 2015. Welcome to the academy of management discoveries (AMD). Academy of Management Discoveries 1, 1–4.
Vilhena, D.A., Foster, J.G., Rosvall, M., West, J.D., Evans, J., Bergstrom, C.T., 2014. Finding cultural holes: how structure and culture diverge in networks of scholarly communication. Sociological Science 1, 221.
Weick, K.E., 1995. What theory is not, theorizing is. Adm. Sci. Q. 40, 385–390.