Industrial and Organizational Psychology, 8(4), pp 509–575 December 2015.
Copyright © 2015 Society for Industrial and Organizational Psychology. doi:10.1017/iop.2015.75

Commentaries

The Big Duplicity of Big Data

Thomas J. Whelan and Amy M. DuVernet
Training Industry, Inc., Cary, North Carolina

As discussed in Guzzo, Fink, King, Tonidandel, and Landis's (2015) focal article, big data is more than a passing trend in business analytics. The plethora of information available presents a host of interesting challenges and opportunities for industrial and organizational (I–O) psychology. When utilizing big data sources to make organizational decisions, our field has a considerable amount to offer in the form of advice on how big data metrics are derived and used and on the potential threats to validity that their use presents. We've all heard the axiom, "garbage in, garbage out," and that applies regardless of whether the scale is a small wastebasket or a dump truck.

In this commentary, we expand on the groundwork of Guzzo et al. concerning caveats, limitations, and recommendations for the use of big data by adapting Shadish, Cook, and Campbell's (2002) framework to explore the types of validity threats that may be characteristic of the use of big data. Although big data promises to expand our understanding of organizational phenomena, it is critical that we remain mindful of the pitfalls that could threaten the validity of conclusions made from these kinds of datasets. For the purposes of this commentary, we've specifically focused on two categories of threats—namely, threats to statistical conclusion and construct validity. Though it may seem rudimentary to review such widely understood threats to validity (at least in our field), we feel that the anticipation of what can be accomplished with the volume, variety, and velocity of big data should not overpower the purposeful rigor of statistical analysis and interpretation that are paramount to I–O. Table 1 presents a summary of the validity threats discussed in this commentary.

Thomas J. Whelan and Amy M. DuVernet, Training Industry, Inc., Cary, North Carolina.

Correspondence concerning this article should be addressed to Thomas J. Whelan, 401 Harrison Oaks Boulevard, Suite 300, Cary, NC 27513. E-mail: twhelan@trainingindustry.com

https://doi.org/10.1017/iop.2015.75 Published online by Cambridge University Press

Table 1. Summary of Threats to Validity in Big Data Research

Threat category	Threat	Suggested practice
Statistical conclusion validity	Statistical power	Report effect sizes
	Violated assumptions of statistical tests	Analyze descriptive statistics; test distribution of variables for normality assumption
	Fishing and error rates	Report number of hypothesis tests; apply corrections for Type I error rates
	Heterogeneity of units	Assess systematic variance in outcomes; explore covariates based on theoretical rationale
Construct validity	Inadequate explication of constructs	Seek multiple operationalizations whenever possible
	Mono-operation/method bias	Replicate and cross-validate results
	Reactivity to the experimental situation	Conduct focus groups; socialize performance data collection

Statistical Conclusion Validity

Threats to statistical conclusion validity, described by Shadish et al. (2002), can be defined as the "reasons why researchers may be wrong in drawing valid inferences about the existence and size of covariation between two variables" (p. 45). Although Shadish et al. explained an array of threats, we focus on the four that we feel are especially relevant in big data analyses.

Statistical Power

As Guzzo et al. asserted, statistical power is a concern with big data, and traditional significance testing may be less relevant with such large datasets. The sample sizes for some big data variables may be so large that a Type I error rate set at the traditional $\alpha = .05$ will result in most hypothesis tests achieving statistical significance, leading to the identification of trivially small effects as significant. On the converse, big data at aggregate levels of analysis (e.g., team, department, organization), as recommended by Guzzo et al. in order to maintain anonymity, could be underpowered for the types of tests being applied. Thus, for both scenarios, we believe the examination of effect sizes (e.g., Bosco, Aguinis, Singh, Field, & Pierce, 2015; Cohen, 1988) rather than, or in supplement to, statistical tests of significance will paint a more complete picture of the relationships investigated.

https://doi.org/10.1017/iop.2015.75 Published online by Cambridge University Press

Violated Assumptions of Statistical Tests

Shadish et al. (2002) warned that "violations of statistical test assumptions can lead to either overestimating or underestimating the size and significance of an effect" (p. 45). Although some statistics, such as a *t*-test, are known to be relatively robust to violations of normality, this is not a universal property of the analyses commonly used and reported in our domain. For example, there are a number of assumptions that need to be met before applying a Pearson product moment correlation, including the correct level of measurement, the presence of outliers, and the normality of the relationship between two variables. Confusion about levels of analysis can result from data aggregation or the scale (e.g., nominal, ordinal, etc.) of the variables under study. The large number of variables available in big data sets can lead to increased difficulties associated with maintaining a complete understanding of their scale, range, and level of aggregation. Outliers, too, may be more problematic to handle as they are likely to occur more frequently given the large number of data points. Finally, although the distribution of big data variables may approach normality more easily than variables found in traditional datasets, this cannot be assumed simply because there are a large number of observations. Many big data variables are likely to exhibit nonnormal distributions; for example, data associated with time to click an object of interest (e.g., a supplemental learning activity in a self-paced training module) will likely present a skewed distribution. For these reasons, we believe it is critical for big data analysts to take appropriate precautions before conducting analyses, including a full exploration of the data to investigate the descriptive statistics, the outliers, and the distributions of variables of interest.

Fishing and Error Rates

Although Guzzo and colleagues (2015) hinted at the increased risk of fishing and error rates implicit in the use of big data, we believe this threat cannot be overstated. The fishing and error rate problem is defined as a situation where "repeated tests for significant relationships, if uncorrected for the number of tests, can artifactually inflate statistical significance" (Shadish et al., 2002, p. 45). Big data affords the ability to run a large number of tests; analysts may utilize stepwise regression, machine learning, and other exploratory techniques in a manner akin to "dustbowl empiricism." Significant post hoc findings may abound in big data, thus concurrent attention to theory-informed, hypothesis-driven analysis is essential. In this respect, I–O psychologists are especially well suited to understand and advise organizations on how to approach big data problems, how to analyze them appropriately, and how to act on the findings. We concur with Guzzo et al.'s recommendation that researchers must clarify the extent to which various relationships were

https://doi.org/10.1017/iop.2015.75 Published online by Cambridge University Press

investigated rather than solely reporting favorable results. However, we are concerned that space limitations may prevent a full description of all investigated relationships. Therefore, we echo Shadish et al.'s (2002) recommendation that in cases where a large number of relationships are investigated, corrections must be applied to account for the increased probability of Type I error rates.

Heterogeneity of Units

Shadish et al. (2002) cautioned that "increased variability on the outcome variable within conditions increases error variance, making detection of a relationship more difficult" (p. 45). This is especially true when comparing metrics across departments or business units, where heterogeneity between groups of employees may be ignored or overlooked in favor of utilizing the largest possible sample. In addition, unexplored interactions between variables under study and other characteristics of the organization, department, or other groups of employees are likely to produce heterogeneity in outcome variables. For example, technical service representatives may be evaluated in terms of their time to resolve customer calls. Individuals in service groups focusing on small electronics may produce dramatically different resolution times than those in groups focusing on troubleshooting large, complex electronics. In response to this threat, we believe it is critical to fully explore the dataset to ensure comparability of outcomes prior to investing time and effort toward theoretical testing of relationships. Further, we advocate for an investigation of covariates that could more fully explicate relationships; big data may make it easier to include covariates that can aid in elucidating systematic patterns in the dataset. However, as discussed in the next section, these covariates may also be subject to threats related to the underlying constructs being measured.

Construct Validity

Threats to construct validity concern "the match between study operations and the constructs used to describe those operations" (Shadish et al., 2002, p. 72). Although there may be additional threats to construct validity than those listed below, we again focus on those that we believe to be most pertinent to working with big data.

Inadequate Explication of Constructs

This is defined as an instance where "failure to adequately explicate a construct may lead to incorrect inferences about the relationship between operation and construct" (Shadish et al., 2002, p. 73). Because the current trend toward the use of big data often relies on archival data, a temptation to use available variables as proxies for focal constructs may arise. Despite a

https://doi.org/10.1017/iop.2015.75 Published online by Cambridge University Press

wealth of data points, the constructs purported to be correlated could lack precision and/or adequate coverage of the construct domain. It is critical that I–Os keep construct validity at the forefront when utilizing big data to investigate organizational phenomenon. As previously argued, we believe that a full descriptive exploration of big data variables is crucial; moreover, in reporting the results of big data analyses, it is pivotal that researchers clearly define construct operationalizations and fully specify the limitations of these operationalizations. We also believe that big data has the potential to offer solutions to this dilemma, as they may provide multiple operationalizations of a specific construct, allowing for tests of construct convergence and affording the potential for a richer exploration of variable relationships than would be permitted with traditional experimental or quasi-experimental research efforts.

Mono-Operation/Method Bias

Mono-operation bias occurs when "any one operationalization of a construct both underrepresents the construct of interest and measures irrelevant constructs, complicating inferences" (Shadish et al., 2002, p. 73). This, again, is related to the previous threats to construct validity. Relying too heavily on a single operationalization of a construct may fall short of capturing the variety of ways in which it can manifest in an organization. This is particularly pertinent to applications of big data, where the quantity of information may overshadow the relative quality when it comes to how constructs are represented. Mono-method bias occurs "when all operationalizations use the same method (e.g., self-report), [and] that method [becomes] a part of the construct actually studied" (Shadish et al., 2002, p. 73). As previously mentioned, for some big data applications, there may be a variety of methods used to collect data about the same construct. In these cases, we believe it is incumbent on analysts to utilize all potential operationalizations in their investigation. Where relationships vary across operationalizations, a thorough investigation of potential causes is warranted. For other types of information, a single source or method may be the only available operationalization due to factors such as the ease, cost, and intrusiveness with which it can be collected in relation to other methods. In these situations, we echo Guzzo et al.'s (2015) recommendation that replication and cross-validation with other sources and methods are critical precursors to conclusions that inform practice.

Reactivity to the Experimental Situation

As stressed in the focal article, issues regarding data privacy are inherent in big data analyses. These types of analyses not only present challenges associated with maintaining anonymity and obtaining informed consent but also

https://doi.org/10.1017/iop.2015.75 Published online by Cambridge University Press

pose threats by way of participant reactivity. Reactivity to the experimental situation is defined as a situation whereby "participant responses reflect not just treatments and measures but also participants' perceptions of the experimental situation, and those perceptions are part of the treatment construct actually tested" (Shadish et al., 2002, p. 73). In the context of big data, as information gathering practices are made more salient to employees—particularly in persistent data collection scenarios such as the "Internet of Things"—and as the amassed results of big data research begin to inform organizational practices, sensitization to the sources of this information may result. This may lead employees to attend to certain tasks or types of data collection, keeping the potential use of those data in mind, to ensure that the information trail of their activities reflects favorably on them. Though there are many ways in which this could happen, a behavioral data example might occur if the time that an employee is signed into a chat application becomes unreliable over time due to employee awareness or perceptions that the data are being used to gauge their time spent engaged in work. A further related example would occur if the content of chat communications were altered to reflect assumptions of monitoring. We believe responses such as these are inevitable, and it will be increasingly challenging to gauge and control how such monitoring interacts with individual differences to shape behavior (e.g., Watson et al., 2013). This represents an important avenue of further research, as a delineation of the boundaries of this threat to validity and continued assessment of perceptions of anonymity are needed to interpret and evaluate the results of big data analytics. Thus, a critical step in big data research is the conduct of focus groups and interviews in order to investigate the potential for participant reactivity. We also believe that socializing behavioral monitoring could produce a baseline, minimizing the potential for individual differences in anonymity perceptions to confound the relationships studied.

Conclusion

Big data has been a recurring and popular theme at the past several Society for Industrial and Organizational Psychology conferences; we hope this commentary adds to these works by providing guidance to I–O researchers and practitioners who have not yet had the opportunity to dip their big toe in the pool of big data. As a field, I–O must be cognizant of where applications of big data are likely to fall prey to such threats to validity, both to appropriately advise organizations and to intervene in scenarios where seemingly innocuous applications of big data are used to make problematic inferences that surpass the statistical understanding of analysts. Although the excitement over applications of big data is certainly warranted, the specter of acting on misleading results is not abated by the size of the data set. I–O expertise

https://doi.org/10.1017/iop.2015.75 Published online by Cambridge University Press

can and should play a leading role in helping organizations to capitalize on the benefits big data can reap.

References

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, *100*(2), 431–449.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Guzzo, R. A., Fink, A. A., King, E., Tonidandel, S., & Landis, R. S. (2015). Big data recommendations for industrial–organizational psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *8*(4), 491–508.

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

Watson, A. M., Thompson, L. F., Rudolph, J. V., Whelan, T. J., Behrend, T. S., & Gissel, A. L. (2013). When big brother is watching: Goal orientation shapes reactions to electronic monitoring during online training. *Journal of Applied Psychology*, *98*(4), 642–657.

Big Data Recommendations for Industrial–Organizational Psychology: Are We in Whoville?

Christopher T. Rotolo and Allan H. Church
PepsiCo, Inc., Purchase, New York

Guzzo, Fink, King, Tonidandel, and Landis's (2015) focal article was intended to be not a set of standards but instead "stepping stones" to "raise awareness and provide direction" (p. 492) to our field for working with big data. We believe that the work done by the authors successfully achieved those objectives, and we encourage the Society for Industrial and Organizational Psychology (SIOP) Executive Committee to keep advancing this work toward greater clarity and guidance. Having clear alignment among members of our field and guidelines for handling nebulous issues such as big data is an important aspect of the scientist–practitioner model. As such, we have no substantial debate over the content being proposed in the focal article. In fact, we would advocate for a full set of standards in this area. We do, however, propose that, as important as it is, this work as commissioned by the SIOP Executive Committee is a shortsighted endeavor and does not take

Christopher T. Rotolo and Allan H. Church, Global Talent Assessment and Development, PepsiCo, Inc., Purchase, New York.

Correspondence concerning this article should be addressed to Christopher T. Rotolo, PepsiCo, Inc., 700 Anderson Hill Road, Purchase, NY 10577. E-mail: christopher.rotolo@pepsico.com

https://doi.org/10.1017/iop.2015.75 Published online by Cambridge University Press