Journal of Management
Vol. 44 No. 4, April 2018 1453–1478
DOI: 10.1177/0149206315604188

Archival Data in Micro-Organizational Research: A Toolkit for Moving to a Broader Set of Topics

Christopher M. Barnes
University of Washington
Carolyn T. Dang
University of New Mexico
Keith Leavitt
Oregon State University
Cristiano L. Guarana
University of Virginia
Eric L. Uhlmann
INSEAD Singapore

Compared to macro-organizational researchers, micro-organizational researchers have generally eschewed archival sources of data as a means of advancing knowledge. The goal of this paper is to discuss emerging opportunities to use archival research for the purposes of advancing and testing theory in micro-organizational research. We discuss eight specific strengths common to archival micro-organizational research and how they differ from other traditional methods. We further discuss limitations of archival research, as well as strategies for mitigating these limitations. Taken together, we provide a toolkit to encourage micro-organizational researchers to capitalize on archival data.

Keywords: *Big Data; archival research; organizational behavior; management; secondary data; research methods*

Acknowledgments: This article was accepted under the editorship of Patrick M. Wright. We appreciate help from Richard Watson in gathering articles in the early stages of our literature review.

Corresponding author: Christopher M. Barnes, University of Washington, 585 Paccar, Seattle, WA 98195, USA.

E-mail: chris24b@uw.edu

A recent estimate published by IBM analytic services suggests that 90% of the world's data have been generated within the last 2 years, with 80% of these data residing in unstructured (yet often readily accessible) formats (SINTEF, 2013). A great deal of these newly generated data, including social media and Internet search terms, capture individual-level (microlevel) attitudes and behaviors that occur in contexts relevant to organizational scholars. Broadly defined here as data initially collected and stored for purposes other than testing the researcher's specific hypotheses, archival research (also known as secondary field research) entails capitalizing on research data that are already in existence rather than generating new primary data.[1] In contrast to macro-organizational research, micro-organizational fields, such as organizational behavior, human resources, and applied psychology, do not appear to be realizing the promise of the "Big Data" revolution or archival data sources more generally.

Indeed, an exhaustive survey of organizational research found that only a small minority of articles include archival data (Scandura & Williams, 2000). For example, Antonakis, Bastardoz, Liu, and Schriesheim (2014) found that in the leadership domain, articles with archival data composed less than 10% of articles in high-impact social science journals. For the present review, we searched the 10 most recent years of articles in the *Journal of Applied Psychology* (a top journal focused almost exclusively on micro-organizational research) and found that only 12% of the articles included at least one study that utilized archival data. Moreover, when micro-organizational researchers have used archival methodologies, they have tended to focus on a relatively narrow set of archival measures, such as employee compensation, turnover, and other ratings from personnel records kept in human resources departments.

We suspect that the primary reason that micro-organizational researchers have underutilized archival research is because they believe the potential limitations (those related to construct validity and omitted psychological mechanisms) outweigh the benefits for studying individual and group behavior. Additionally, underutilization may be due to dogma about the appropriateness of using archival data for conducting micro-organizational research, a lack of awareness about the vast amounts of archival data that are now available, the real or perceived irrelevance of archival data to local settings, or the real or perceived inability to access and analyze archival data sets easily. Moreover, micro-organizational researchers may not recognize the unique strengths of archival research and may simply not be trained to conduct archival research. Thus, a discussion of how to conduct archival research in micro-organizational content areas should help to address some of the issues that are keeping the micro-organizational research community from capitalizing on the Big Data movement. We seek to challenge assumptions of micro-organizational researchers regarding the appropriateness of archival methods for micro-organizational research, namely, that issues of measurement and construct validity related to archival data are insurmountable for topics relevant to micro-organizational research and that the scope of archival data opportunities is generally restricted to commercially available data sets focused on firm-level analysis.

Rather than viewing the strengths and limitations of archival research as trade-offs that researchers must make if they choose to use archival methodologies, we argue that the limitations of archival methodology (mainly dealing with measurement and construct validity) can be addressed through specific strategies and by using archival research in ways that are supplemental to those already commonly used within micro-organizational research. That

is, we argue that archival approaches offer strengths that contribute to the goals of conducting full-cycle organizational research (Chatman & Flynn, 2005). Assuming the limitations surrounding measurement and construct validity are adequately addressed elsewhere, the strengths of archival research present a unique avenue upon which to expand and test micro-organizational theories. We contend that once hypotheses are supported using other methods (e.g., surveys, experiments), archival research provides the opportunity to explore phenomena in social, political, and/or cultural realms that are typically unsuitable for other forms of research (i.e., *Assuming that this hypothesis is true, in what ways does it manifest in the world?*). Thus, archival research can offer a unique avenue for tying psychological phenomena to important real-world outcomes. Accordingly, we present here the unique strengths and challenges of archival methods, focusing on issues pertinent to micro-organizational researchers.

We first discuss eight specific strengths that are common in archival research that can be specifically applied in the context of micro-organizational research. We discuss how micro-organizational researchers have recently utilized archival databases in innovative ways, providing examples of how such research capitalizes on each of these strengths. We then discuss common known limitations to archival research and describe how they might be overcome within a micro-organizational research context. We provide tables that show examples of archival research in micro-organizational research, as well as some examples of data sources that can be used for this purpose. In sum, we develop an initial toolkit for micro-organizational researchers that should enable them to more fully capture the value of the currently underutilized archival research approach.

Strengths of Archival Research

In this section, we discuss strengths of archival methodology for the purpose of advancing micro-organizational research. Although not all archival data sets uniformly offer these strengths, they generally represent strengths not readily afforded by more commonly used methods in micro-organizational research. We address each of these strengths, sharing examples from recent archival micro-organizational research. We begin with a novel type of question that is made possible by archival research: identifying unexpected manifestations of a theory in the real world.

Strength 1: Uncovering Unexpected Manifestations in the Real World

Archival research opens new worlds of research questions that are typically absent from the micro-organizational research literature but which may greatly improve external validity and relevance for policy decisions. Whereas micro-organizational research commonly uses experiments to assess the possible causal relationship between variables (i.e., asking, *Under certain conditions, can this hypothesis be true?*) and survey field methods to demonstrate that strong and possibly causal relationships exist in real applied settings (i.e., asking, *Under normal circumstances, is there evidence that this hypothesis is true?*), we argue that archival research can effectively demonstrate meaningful downstream implications of a hypothesis (i.e., asking, *Assuming that this hypothesis is true, in what ways does it manifest in the world?*). As such, we believe archival research presents a way to balance the limitations of

construct measurement/validity of process variables—on the basis of the assumption that the process is supported in previous research findings—with the benefits of meaningfully operationalizing important outcomes in the world.

By asking themselves, *Assuming that this hypothesis is true, in what ways does it manifest in the world?*, micro-organizational researchers can uncover new and unexpected domains in which a theory may hold true. For example, across a series of three studies using data from government agencies, Varnum and Kitayama (2010) found that babies received popular names less frequently in western regions of the United States compared to eastern regions. This pattern was replicated when comparing baby names in western versus eastern Canadian provinces. Explanations for this revolved around the voluntary settlement theory, which proposed that economically motivated voluntary settlement in the frontiers (e.g., western areas of the United States and Canada) fostered values of independence. Thus, in the case of baby naming—a behavioral act with strong personal and familial significance—babies born in western versus eastern regions receive less popular baby names.

Approaching research questions from this angle allows for overlooked novelty but also allows us to capture the broader domain space of a construct. Historically, focus within micro-organizational research on measurement properties (which typically entailed using scales with good psychometric properties) may mean that our studies cover only a narrow part of the domain space of a construct (Leavitt, Mitchell, & Peterson, 2010). In contrast, by examining proxies with archival data, we can show that the same pattern of results generalizes to other socially meaningful conceptualizations of a construct.

Strength 2: Measurement of Socially Sensitive Phenomena

Many organizationally relevant phenomena are socially sensitive and, thus, difficult to measure accurately through either self-report or in a laboratory in which participants are aware that they are under observation. Some behavior, such as unethical behavior, deviant work behavior, counterproductive workplace behavior, workplace harassment, incivility, work discrimination, abusive supervision, and workplace bullying, may be accompanied by social stigma and punishment (Bellizzi & Hasty, 2002; Cameron, Chaudhuri, Erkal, & Gangadharan, 2009; Gino, Shu, & Bazerman, 2010). Large literatures indicate that these behaviors are common in organizations (cf. Wimbush & Dalton, 1997), but employees may not be willing to admit to engaging in such behavior. A common solution to this is to have employees rate each other on these behaviors, such as supervisors rating subordinates (e.g., Barnes, Schaubroeck, Huth, & Ghumman, 2011). However, many socially sensitive behaviors may be conducted in private, with the intended goal of deliberately avoiding detection. In other words, both self-rated and other-rated measures of socially sensitive behaviors may be limited, capturing only some instances of what is generally socially sensitive behavior.

Archival data can be especially helpful in the context of socially sensitive phenomena. For example, counterproductive work behavior could come with the stigma of being perceived as a bad employee or with the possibility of being fired or being passed over for opportunities for career advancement or financial benefits. Detert, Treviño, Burris, and Andiappan (2007) conducted a study of counterproductive work behavior in 265 restaurants. Their measure of counterproductive work behavior was food loss, which was obtained from organizational archives. Although the archival data in this study did not link counterproductive work

behavior to individuals, the data did indicate at the work unit level an objective estimate of a specific counterproductive work behavior. Bing, Stewart, Davison, Green, and McIntyre (2007) similarly studied counterproductive behavior, operationalizing it as traffic violations stored in archives. Dilchert, Ones, Davis, and Rostow (2007) operationalized counterproductivity in police officers as incidents recorded in organizational archives, including racially offensive conduct, incidents involving excessive force, misuse of official vehicles, and damaging official property.

Harmful aggressive workplace behavior may similarly be viewed as undesirable in employees, with negative outcomes for employees perceived as engaging in such behavior. Larrick, Timmerman, Carton, and Abrevaya (2011) conducted a study of aggression and retaliatory behavior toward fellow employees by utilizing archival data from Major League Baseball, including 57,293 professional baseball games. They operationalized aggressive behavior as batters who were hit by a pitch (including controls such as pitcher accuracy), which is recorded in the detailed baseball archives. As with the Detert et al. (2007) study noted above, Larrick et al. measured their outcome variable in a manner that avoids the limitations of measuring socially sensitive behavior through self-report or supervisor report.

Work accidents may similarly be socially sensitive, in that they may call into question the competence of employees and organizations involved in the incident. In some organizational contexts, reporting workplace accidents is required by law or policy (cf. Barnes & Wagner, 2009), such that employees not reporting such accidents risk being fired. However, potential sanctions for injuries provide incentives for organizations to underreport accidents and injuries. In a study of 1,390 employees, Probst, Brubaker, and Barsotti (2008) compared data from Occupational Safety and Health Administration logs—the mechanism for reporting injuries and illnesses to regulating bodies—with medical claims data from an owner-controlled insurance program, which is not monitored by regulating bodies. Probst and colleagues found that the number of injuries reported in the owner-controlled insurance program was over three times higher than that reported by surveys to the official Occupational Safety and Health Administration.

In many of the cases noted above, socially sensitive behaviors are measured and archived indirectly or incidentally as part of a larger effort. However, organizations occasionally go through considerable efforts to measure socially sensitive behavior by using resources that are typically beyond the reach of most researchers. For example, credit agencies gather information from various sources, including a broad array of creditors, to feed into financial algorithms to generate credit scores. Bernerth, Taylor, Walker, and Whitman (2012) utilized this archival data from Fair Isaac Corporation to test their model of employee behavior and performance. Social media and marketing organizations may similarly present opportunities to draw upon multiple external sources to measure variables that are relevant to organizational behavior.

In all of the examples noted above, the archival sources of data avoid many of the limitations entailed by researchers directly asking research participants or their supervisors to report socially sensitive data to researchers. Although archival sources of data are not without their own limitations, many such sources are measured in a manner that is not salient to the research participants, minimizing the incentive or opportunity to distort their responses. Indeed, in examples such as the injury reports noted by Probst et al. (2008) and the credit scores noted by Bernerth et al. (2012), archival data are uniquely robust to the efforts of employees who may deliberately distort self-reported data.

Strength 3: Research Reproducibility

Management and micro-organizational research have recently increased focus on the topic of transparency in data and analyses (Briner & Walshe, 2013; Kepes & McDaniel, 2013; see also Nosek, Spies, & Motyl, 2012; Simmons, Nelson, & Simonsohn, 2011). This is due in part to concerns regarding research misconduct (Simonsohn, 2013) as well as differences in opinion regarding the most appropriate procedures for conducting analyses. Because archival data can typically be obtained by interested parties seeking to replicate analyses, these data allow transparency in a manner that is not as readily accessible in laboratory or field research. Indeed, many archival data sets are publicly available and often are free to access. This discourages research misconduct and allows for easier detection of misconduct. Furthermore, it enables open and transparent discussions regarding different approaches to data analysis for the same research question (e.g., compare the updated conclusions of Silberzahn, Simonsohn, & Uhlmann, 2014, to the original conclusions of Silberzahn & Uhlmann, 2013). Indeed, archival research will often allow for a form of replication in which the same data are used and the analyses and conclusions are examined by multiple independent parties (Sakaluk, Williams, & Biernat, 2014). Relatedly, access to publicly available data also "levels the playing field" for scholars at institutions with limited resources, increasing the number of scholars who can contribute to a research area.

Strength 4: Statistical Power

Small samples are subject to Type II errors, in that they may lack the statistical power to detect a relationship between constructs (Cohen, Cohen, West, & Aiken, 2003). This may lead to deficient theories that leave out important theoretical links. Small samples are also subject to Type I errors, in that sampling error may lead to results that provide false positives that will not replicate in other samples (Hollenbeck, DeRue, & Mannor, 2006). As Cohen et al. note, all else held equal, there is a trade-off between Type I and Type II errors, depending on cutoff choices in statistical analyses. However, when sample size is allowed to vary, Type I and Type II errors are not in such direct opposition to each other; large samples lower the frequency of both Type I and Type II errors.

A related issue is that with small samples, sampling error becomes a significant issue that biases the results. As clearly indicated by Hunter and Schmidt (2004), many studies of the same phenomenon will have results that appear different, and much of these differences will be due simply to sampling error. The solution proposed by Hunter and Schmidt is meta-analyzing literatures. Meta-analysis is a very helpful tool, but often literatures have to wait for years for enough studies to accumulate to meta-analyze. In large samples, there is much less of an issue of sampling error driving differences in effect sizes among different studies, enabling researchers to begin making strong inferences before sufficient articles have accumulated for a meta-analysis.

Although not all archival data sets are large, many archival databases contain very large data sets with sample sizes that are simply not found in other areas of research. Laboratory and field studies commonly have as many as a few hundred participants, but archival studies commonly have thousands of participants. Examples of archival research utilizing extremely large data sets include Barnes and Wagner (2009) with over 500,000 work injuries and Larrick et al. (2011) with over 4 million employee interactions. In sum, many archival

databases enable very high levels of statistical power that mitigate concerns of both Type I and Type II errors. Although this is not the case with every archival study, this occurs much more frequently with archival research than with more typical forms of research.

The statistical power provided by many archival data sets may invite concerns that researchers will find statistically significant but not practically significant effects. This is a reasonable concern, in that there may be detectable effects that are trivial in nature. However, if researchers focus on effect sizes in addition to statistical significance, effects that have practical significance can be identified. Additionally, relatively small effects may have profound aggregate impact over time, a phenomenon known as sensitive dependence. For example, an agent-based simulation by Martell, Lane, and Emrich (1996) found that a 1% gender bias (which may be otherwise viewed as trivial) led to nearly the same level of underrepresentation of women at the top levels of an organization as a 5% bias. While arithmetic simulations allow scholars to test for the downstream consequences of dynamic models (Fioretti, 2013), they rely heavily on starting points and careful model specification to make accurate predictions. We argue that archival research can similarly be used to examine sensitive dependence without requiring such a priori formalized assumptions. For example, Barnes and Wagner (2009) incorporated a weak quasimanipulation of sleep on a broad scale, showing evidence that even small amounts of lost sleep have work outcomes that are of practical significance. In this form of demonstration, the high statistical power that is often possible with archival research allows for examinations of small changes in predictors. Thus, the small effect of the spring change to daylight saving time (i.e., only 1 lost hour of sleep) would typically go unnoticed, but Barnes and Wagner detected a notable increase in workplace injuries of 5.6%. This work has subsequently been cited in policy arguments involving the value and harm caused by daylight saving time, whereas similar research conducted within the laboratory (reporting only effect sizes on momentary outcomes) would be unlikely to capture the concern of regulators (Barnes & Drake, in press; Mirsky, 2014).

Strength 5: Deriving Population Parameter Estimates

Not only do researchers aim to uncover the relationships among constructs but, often, researchers and practitioners also are interested in estimating the population value of an effect. Indeed, for practitioners, specific estimates can provide more useful data. For example, beyond simply indicating that the relationship between two variables is significantly different than zero, a population parameter estimate may indicate how much a manager should weight a specific decision cue. Typically, organizational behavior researchers utilize samples and estimate confidence intervals that should contain parameter estimates. As Cohen et al. (2003) note, standard errors influence confidence intervals, and larger samples decrease standard errors. Thus, larger samples enable better population parameter estimates. And as noted above, archival databases often provide opportunities to utilize very large samples.

Additionally, archival databases are sometimes constructed in order to be carefully representative of a specific population in a manner that is typically not possible in other forms of research. For example, the Bureau of Labor Statistics goes to great lengths to ensure the representativeness of the American Time Use Survey, utilizing a large staff to track down a very large and carefully sampled group of research participants. Barnes, Wagner, and Ghumman (2012) utilized this archival database to test a theoretical extension to the work-life conflict

literature. Although samples can never perfectly emulate the population from which they came, the American Time Use Survey provided a much better approximation than a convenience sample or even a systematically sampled group of a few hundred participants. The amount of resources required to generate a truly representative sample is typically beyond what is available to most micro-organizational researchers. Typically, researchers rely on meta-analyses to aggregate findings across different studies conducted in different settings. As indicated above, meta-analyses are extremely helpful tools, but they require time to accumulate enough of a body of studies to aggregate. Moreover, the aggregation of multiple nonrepresentative studies does not necessarily result in an aggregated sample that is representative of the population. Indeed, by definition, meta-analyses are composed of the individual studies they aggregate, and researchers have called into question the representativeness of samples in organizational behavior (Shen, Kiger, Davies, Rasch, Simon, & Ones, 2011).

A recent example of an empirical article estimating a population parameter estimate is the Probst et al. (2008) examination of workplace injuries and illnesses in construction companies. As noted above, Probst et al. estimated the actual occurrence of injuries by probing an owner-controlled insurance program. Probst et al. explicitly note that even their estimate is likely conservative, in that there are likely some injuries that are not reported even to non-government-monitored insurance programs. However, their estimate likely more closely approximates the true score of the injury rate than do other methods of investigation.

Strength 6: Examining Effects Across Time

Time is an important factor that may influence the proposed relationship between variables. Mitchell and James (2001) note that within organizational literature, few studies specifically address the theoretical and methodological issues caused by time. For example, in most X and Y relationships examined by management scholars—where X is theorized and empirically demonstrated to cause Y—issues relating to the duration of the effect or the trajectory of the effect over time are not specific. This may be due in part to costs (time, monetary, etc.) associated with conducting longitudinal studies.

Archival research may thus prove an especially useful tool upon which to examine issues of time in micro-organizational research. With regards to the duration of effects, many archival data sources have a longitudinal design, which allows researchers to examine the duration of an effect. Using data from the Dunedin Study—which is a longitudinal study of the health and behavior of individuals in Dunedin, New Zealand—Roberts, Caspi, and Moffitt (2003) found that work experiences affected personality traits over the course of an 8-year time span. This allowed for a more precise specification regarding (a) the existence of an effect between work experiences and personality development and (b) the duration with which work experiences affect personality development. Similarly, Judge and Kammeyer-Mueller (2012) studied ambition by using data from the Terman Life Cycle Study, which examines high-ability individuals over a 7-decade period.

The longitudinal nature of archival data sources could also speak to when effects occur. Using data from the German Socio Economic Panel, which tracks labor and demographic characteristics of German workers, Weller, Holtom, Matiaske, and Mellewigt (2009) found that personal recruitment methods had the strongest effect on turnover early on in employees' tenure with the organization. The effects diminished as employees' tenure with the organization increased. In addition to addressing the duration of effects, archival data sources can

speak to the trajectory of phenomenon over time. Indeed, some archival sources, such as the General Social Survey (which began in 1972) and the Current Population Survey (which began in 1940), are conducted annually or biennially, allowing researchers to examine whether phenomenon change over the course of long periods of time that are impractical in other forms of research. For example, Highhouse, Zickar, and Yankelevich (2010) used the General Social Survey to assess whether the American work ethic is in a state of decline or of increase or has remained constant from the period of 1980 to 2006 and whether that trend is similar to or different from the trend observed from the period of 1950 to 1980.

Strength 7: Differences in Relationships Across Sociopolitical Contexts

Micro-organizational researchers are increasingly focusing on cross-cultural research. Such research enables examination of differences between cultures in behaviors and relationships among constructs (e.g., Ng & Van Dyne, 2001) and of how people from different cultures interact together (Hinds, Liu, & Lyon, 2012) and provides diverse perspectives for theoretical innovation (Chen, Leung, & Chen, 2009). Sociopolitical contexts may be explicitly included in theoretical models or explored as potentially relevant moderators.

P. E. Spector and colleagues provide examples of such research with samples that are a hybrid between primary field research and archival research. Spector and C. L. Cooper developed the Collaborative International Study of Managerial Stress (CISMS), an ambitious field study entailing a variety of measures collected from a broad variety of participants across over 20 countries. Spector and colleagues utilized this large field sample as a database for multiple studies, in essence utilizing it as a large archival study that would enable many subsequent smaller research projects. For example, Spector et al. (2004) conducted a study of a subset of the CISMS database. Their sample was composed of over 2,000 managers from a broad variety of countries, enabling comparison of stress across different sociopolitical contexts. They found differences across Anglo, Chinese, and Latino participants with regards to work hours, job satisfaction, mental well-being, and physical well-being. In a different study also utilizing the CISMS database, Spector et al. (2002) found that locus of control beliefs contribute to well-being almost universally across 24 countries. Moreover, individualism/collectivism did not moderate this relationship. However, they suggest that how control is manifested can still differ across sociopolitical contexts. In a second phase of the CISMS study, Spector et al. (2007) found that country cluster moderated the relation of work demands with strain-based work interference with family, with the Anglo country cluster showing the strongest relationship. This suggests that research demonstrating that work demands influence strain-based work interference with family does not fully generalize to all countries, which is an important finding given that a large portion of organizational behavior research is conducted in Anglo countries.

This type of comparison of organizationally relevant phenomena across sociopolitical contexts is difficult in most primary studies. Indeed, the author string length of some such papers indicates the amount of resources this requires (e.g., 23 authors in Spector et al., 2007). Thus, archival databases including data from multiple sociopolitical contexts can be extremely valuable in advancing research in micro-organizational research, even if such databases originate from deliberate efforts to assemble large databases that will enable programmatic subsequent studies. Indeed, the first few studies from such data sets may be considered primary research, but researchers can later go back and utilize the same such databases

in a manner that would be considered using archival data. Thus, Spector and colleagues show how a primary database can be utilized later as an archival database to pursue multiple research projects. Moreover, they pursue research questions that are especially well suited for archival data.

Strength 8: Theory Extension and Testing at Higher Levels of Analysis

Micro-organizational research often includes multiple levels of nesting, with events nested within individuals nested within groups and teams nested within larger work units and divisions nested within organizations nested within national cultures. Measurement at higher levels of analysis typically involves measuring data at the individual level of analysis, providing justification for aggregation, and then aggregating (Chan, 1998; Kozlowski & Klein, 2000). However, there are important limitations in aggregating individual data to higher levels of analysis (Newman & Sin, 2009; van Mierlo, Vermunt, & Rutte, 2009), and groups may vary in the agreement among individuals (LeBreton & Senter, 2008). Indeed, an intraclass correlation, or ICC(1), value of .25 can be considered a large group-level effect (LeBreton & Senter; Murphy & Myors, 1998) even though a large portion of the variance is still at the individual level.

One way to avoid many of these limitations is to directly measure the construct at the unit of analysis at which it conceptually resides. Archival databases may provide opportunities for doing exactly this. Team performance is a commonly studied construct in micro-organizational research (cf. Ilgen, Hollenbeck, Johnson, & Jundt, 2005). Some organizational contexts capture and record ratings of team performance, oftentimes on the basis of objective outcome data rather than individual perceptions. For example, Keller (2006) examined 118 research and development project teams, obtaining data on profitability and speed to market as operationalizations of team performance from the archives of the sample organization.

Outside of the teams literature, there may be other levels of analysis for which archival data are well suited. As noted above, Detert et al. (2007) conducted a study of counterproductive work behavior at the work unit level of analysis. Their measure of counterproductive work behavior was food loss at the work unit level, which was obtained from organizational archives. Thus, both conceptually and empirically the variable was at the work unit level of analysis. Indeed, the strategic management literature largely focuses on larger units of analysis (especially firm level) and utilizes archival data extensively. Micro-organizational research may often intersect with strategic management in mesolevel research, for which archival data could be very useful. For example, Pritchard, Jones, Roth, Stuebing, and Ekeberg (1988) conducted a study of a productivity measurement and enhancement system. In this research, they implemented an intervention composed of goal setting, performance measurement, and incentives. Utilizing archival data from the U.S. Air Force, they found that this intervention positively influenced organizational productivity.

It is worth noting that it is important to control for omitted sources of variance due to nesting of data that violates assumptions of nonindependence of data (Antonakis, Bendahan, Jacquart, & Lalive, 2010; Bollen & Brand, 2010; Halaby, 2004). Researchers may err by using random-effects methods without ensuring that sources of variance from higher levels of analysis are accounted for. Thus, archival researchers should leverage the multilevel nature of archival data sources when appropriate and in other circumstances merely control for multilevel nesting.

Challenges and Solutions for Archival Research in the Micro-Organizational Domain

Selecting the Research Question

In conducting archival research, an important first step is selecting the research question. In research that takes a theory testing approach, theoretical models will serve as the source of the research question. Theory testing is an important means of advancing our science and is clearly an appropriate strategy for archival research. However, archival research can also be well suited to an exploratory approach that can serve as an important building block to subsequent confirmatory studies. In research that is more exploratory in nature, then, a gap between theories may often be a source of the research question. Wherever the question comes from, it is important to consider whether it is suitable for archival research.

Given the concerns we discuss below regarding how archival research may have construct validity limitations, it may be helpful to consider the likelihood that a particular construct from a research question is likely to be a good fit for archival research. With the exception of research databases constructed by researchers for the purpose of enabling future research—for example, the General Social Study (Highhouse et al., 2010) and the National Youth Longitudinal Study (Judge, Hurst, & Simon, 2009)—most archival databases will not include psychological states, such as affect or cognition. Archival databases are typically more useful for measuring behaviors, outcomes, and dimensions of the context. Thus, researchers will typically have more success in conducting archival research when the research question entails measuring behaviors, work outcomes, or context.

An additional question to consider is the primary goal in asking the research question. As noted above, archival research will commonly (but not always) suffer limitations in establishing causality. However, archival research is better suited than many other alternatives for some goals. Micro-organizational researchers will likely find their archival research efforts to be most successful when their research questions play to the strengths of archival research.

Concerns regarding archival methodologies primarily surround construct validity issues and causality issues. In the next section, we discuss these limitations and recommend potential strategies to mitigate these weaknesses, especially multistudy approaches that use complementary methods.

Construct Validity

Archival data can come from many sources (see Table 1). Some such sources will be large-scale surveys conducted for the purposes of enabling future studies that are yet to be determined and will utilize psychometrically sound measures. More common will be archival databases assembled for purposes other than conducting any sort of research. For example, sports databases (e.g., Barnes, Reb, & Ang, 2012; Larrick et al., 2011; Timmerman, 2007) are typically created for the purposes of recording history and giving sports fans another means of following games. Accident and injury databases are typically created and maintained not for research purposes but in order to enable the improvement of practice and to enable government oversight (e.g., Barnes & Wagner, 2009). Publically available, externally conducted customer service ratings, such as Trip Advisor, are typically assembled to guide consumer decisions (e.g., Conlon, Van Dyne, Milner, & Ng, 2004).

Table 1
Sources of Data

Sources of data	General description and characteristics	General strengths	General weaknesses
Company records	Data collected from companies and organizations for multipurpose use; examples include records from human resources department (e.g., performance reviews) or company profit reports	Includes variables directly relevant to management constructs (e.g., job performance, absenteeism, turnover); can track changes over time (e.g., quarterly, yearly); third-party ratings (e.g., supervisors rating employees)	Not publically available; access to data dependent on permission from company/ organization; single-item measures; limited constructs available in records
Academic records	Data regarding students collected by schools (e.g., colleges, high school); examples include academic performance (grade point averages); scores on exams (e.g., SATs); demographic information (gender, race)	Includes variables directly relevant to management constructs (e.g., general mental ability); can track changes over time (e.g., students' grade point average)	Not publically available; access to data dependent on permission from schools; single-item measures; limited constructs available in records; exclusively student/ prospective student sample
Previous research project data repurposed	Data collected by researchers, used by other researchers for different projects	Facilitates replication of previous findings	Not publically available; limited constructs available in data
Government research archives	Data collected by various government agencies for multipurpose use; examples include U.S. census data; U.S. Bureau of Labor Statistics data; American Time Use Survey	Publically available; large sample; sampling design facilitates inferences about data's findings to larger population (e.g., American population); can track changes over time	Variables less directly relevant to management constructs; single-item measures; limited constructs available in records
Sponsored research activities from agencies	Data collected by various agencies for multipurpose use; examples include survey data from Pew, Gallup, General Social Survey	Can be publically available; large sample; sampling design facilitates inferences about data's findings to larger population (e.g., American population); can track changes over time; contains modules that are relevant to management constructs (e.g., Pew's attitudes towards work survey)	Limited constructs available in data; single-item measures

(continued)

Table 1 (continued)

Sources of data	General description and characteristics	General strengths	General weaknesses
Sports data	Sports data collected by various sources; examples include statistics from major American sports leagues (e.g., National Football League, Major League Baseball, National Basketball Association, National Hockey League); information about athletes' contracts	Publically available; includes variables directly relevant to management constructs (e.g., job performance, turnover); can track changes over time	Limited constructs available in data; single-item measures
Historiometry	Texts about historical and contemporary figures and events; examples include books and articles written about historical/contemporary figures and events; speeches given by historical/contemporary figures	In-depth information about figures and events; chronicles rare and unique events (e.g., natural disasters)	Limited sample (e.g., analysis limited to single event/figure); limited constructs available to study; limited quantitative information
Media/communication	Information from media outlets or companies; examples include newspaper articles; video recordings; letters to shareholders	In-depth information about figures and events; can compare and contrast information from different outlets	Limited constructs available to study; limited quantitative information

Researchers should be cautious when considering such databases, carefully evaluating whether the measures included in the database can be said to accurately represent a given construct. In laboratory research, manipulation checks can help to ensure that the experimental procedure manipulates what was intended (Rosenthal & Rosnow, 1991). In primary field research, researchers can seek to establish construct validity by examining convergent and divergent validity, often using factor analysis as a tool in such examinations. However, in archival research, such tools may be limited given that measures may be single item (e.g., employee job performance is reported as a single-item score, such as sales) without additional items assessing the construct of interest.

As noted by Schwab (1980), reliability is necessary (but not sufficient) for construct validity. Almost all primary field research measures at least one form of reliability, such as test-retest reliability, internal consistency, split-half reliability, or group mean reliability (Kozlowski & Klein, 2000; Nunnaly & Bernstein, 1994). These tools are also often not available in archival research, making it difficult to empirically evaluate the reliability of variables included in many archival databases. Indeed, researchers should evaluate the potential of each archival source for issues such as missing data and inaccurately recorded data. There will likely be variance from database to database in the reliability of the data, and few archival databases will have multi-item measures that allow for typical measurement of reliability. Moreover, in some databases, it is not clear to what degree different data contributors vary in their coding approach or whether some people responsible for entering data were not diligent in doing so.

To an even greater degree than with most laboratory or primary field research, micro-organizational researchers conducting archival research must carefully evaluate construct validity and be willing to walk away from a data set rather than attempt to draw inferences when construct validity is low. Most potential sources of archival data are not actually suitable for research purposes. However, given the sheer number of data sources, if even only a fraction is suitable, this can add considerable value to micro-organizational research. Furthermore, many of the downstream outcomes of interest in archival research, including salaries paid, revenues generated, retail shrinkage, or lives lost, are actually better estimated within archival sources than they are with self-reported data, which are prone to bias from self-presentation biases, limitations of memory, and recency effects.

One avenue for conducting archival research involves taking qualitative data, such as recorded or transcribed speeches, and converting these data into quantitative data in a valid manner. Content analysis is a common approach, which often involves quantifying and categorizing the content of a sample of communication. There are automated software tools that count the occurrences of certain words on the basis of research linking those words to specific constructs (Duriau, Reger, & Pfarrer, 2007). Indeed, there are online repositories of information and social media that can feed into such tools (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011; Bligh, Kohles, & Meindl, 2004; Cambria, Schuller, Xia, & Havasi, 2013; Feldman, 2013; Kouloumpis, Wilson, & Moore, 2011; Pang & Lee, 2008). We recommend using validated algorithms for extracting such content. Alternatively, human independent raters can similarly code data, either on the basis of a count of specifically defined language or through the coding of other more holistic characteristics. When human coders are used, establishing the convergence amongst raters is a helpful step toward establishing the reliability of the measure.

Causality

According to Popper, "To give a causal explanation of an event means to deduce a statement which describes it, using as premises of the deduction one or more universal laws, together with certain singular statements" (1959: 38). In order to establish causality, researchers must establish temporal precedence, establish covariation, and eliminate alternative explanations, such as spurious correlations (Shadish, Cook, & Campbell, 2001). In many contexts, archival research may be as well suited as any other form of research for establishing covariation.

The ability to establish temporal precedence will vary from source to source. In some studies, there is clear temporal precedence (e.g., Pritchard et al., 1988). In contrast, some archival databases may be cross-sectional, making it difficult if not impossible to establish temporal precedence. In such cases, reverse causality is a clear possibility that cannot be ruled out. Similarly, unmeasured variables may play important roles. Endogeneity refers to the fact that some measured variables in a model may be influenced by variables not included in the model; this is potentially problematic because variables outside of the model and data may drive both the predictor and the outcome (Hamilton & Nickerson, 2003; Hitt, Boyd, & Li, 2004). Endogeneity is an especially concerning issue in archival research (Antonakis et al., 2014) and may cause spurious correlations. Although problems related to endogeneity can be minimized in laboratory experiments (Antonakis et al., 2010), it remains a challenge in archival data. Nonetheless, econometricians have developed methodological approaches to minimize the complications that arise with endogeneity. For example, researchers can identify instruments in their data set. Instruments are "exogenous variables and do not depend on other variables or disturbances in the system of equations" (Antonakis et al., 2010: 1100). Another approach to minimize endogeneity concerns is to use panel data. Panel data allow creating fixed effects to account for unobserved sources of variance in the cluster that predicts behavior (Hamilton & Nickerson). Strategy researchers have identified methodological approaches to minimize the effects of endogeneity on archival data (for a comprehensive list of recommendations, see Antonakis et al., 2010).

Also relevant is whether the construct of interest in the archival data set varies randomly in nature or whether the variable could be affected by simultaneity or omitted variables. Time (regular cycles such as daylight saving time) or geography (e.g., latitude) are less of a concern with regards to simultaneity or omitted variables. Constructs that vary randomly in nature, such as cognitive ability, may be driven by other variables (e.g., socioeconomic status) in a manner that is of greater concern. Others are clearly exogenous, such as temperature as the independent variable in Larrick et al. (2011) or mortality rates as the independent variable in Acemoglu, Johnson, and Robinson (2001), which cannot vary as a function of other variables in the specification or omitted variables from the model.

Even when carefully designing archival research to attempt to meet the Popper (1959) criteria for causality and follow procedures intended to minimize concerns about endogeneity, causal implications are *guaranteed* only by an experimental design. However, experiments typically suffer from low statistical power (Ioannidis, 2014) and imperfect generalizability to real organizational situations. A judicious multimethod approach that includes archival data as a complement to currently prevalent methodologies in microresearch allows for the strongest scientific inferences. It may be too burdensome to expect every paper to include both archival and experimental studies. However, across a program of

research in a given literature (which may include multiple papers), the inclusion of archival research to go along with experimental research should add complementary strengths. This will allow for researchers to capitalize on opposite ends of the continuum of rigor and realism.

Mitigating Limitations of Archival Research

The limitations of archival research often apply to other forms of research as well. Fortunately, there are ways to mitigate the limitations noted above (for some examples, see Table 2). An important step is the selection of the right archival database. Issues of temporal precedence can be addressed by selecting archival databases that are structured from data obtained at multiple time points, with a clear indication of which events preceded which. Accordingly, because critical variables are measured at multiple points in time, additional causal support can be drawn (and reciprocal causation ruled out) by showing a relationship exists between a proposed cause at Time 1 and the proposed outcome at Time 2 but not between the proposed outcome at Time 1 and the proposed cause at Time 2. Some alternative explanations can also be ruled out by selecting databases that include relevant control variables that can be utilized in subsequent analyses. Similarly, selecting carefully constructed archival databases can ameliorate issues of unreliability due to inaccuracy in collection and recording. Fortunately, many databases are painstakingly constructed, with strict procedures regarding data collection and recording. The Bureau of Labor Statistics is an example of an organization that puts great care into data collection and entry (cf. Barnes, Wagner, & Ghumman, 2012).

A related strategy for mitigating the limitations of archival research is to carefully select variables. One limitation noted above is that it is often difficult to establish construct validity. One strategy is to select variables from databases that are constructed by researchers using validated measures (e.g., Highhouse et al., 2010; Judge et al., 2009). Another is to select variables that capture unambiguous stimuli, such as workplace injuries (Barnes & Wagner, 2009), workplace aggression (e.g., Larrick et al., 2011; Timmerman, 2007), or theft (Detert et al., 2007).

Some archival databases are potentially useful even if they have a major limitation. Rather than missing out on the potential value of such archival databases, researchers can implement the strategy of combining multiple studies. The goal of the researcher should be to address one or more of the major limitations of a given archival database with another study. Thus, matching up complementary strengths and weaknesses of different studies can help researchers to gain a better understanding and vector in multiple tests of theoretical models.

One way to do this is with a different archival database. An example of this strategy is provided by Barnes and Wagner (2009). As noted above, their investigation of the influence of the clock changes associated with daylight saving time included a model proposing that sleep mediates the effect of the clock changes on workplace injuries. In Study 1, they utilized an archival database of mining injuries. However, Study 1 did not measure the proposed causal mechanism of sleep. To address this limitation, Barnes and Wagner conducted a second study, using the American Time Use Survey put together by the Bureau of Labor Statistics. In Study 2, they directly examined the effect of clock changes on sleep.

A second way to do this is to use a different method in the complementary study. Laboratory and primary field studies tend to have different strengths and limitations of archival research

Table 2

Examples of Articles Using Archival Data in the Most Recent Decade (2005–2014) in the *Journal of Applied Psychology*

Citation	Source of data	Example of construct(s) measured	Operationalization of construct(s)	Nature of IV (exogenous or endogenous)	Sample size	Time frame	Combined sources/ measures
Smith-Jentsch, Mathieu, & Kraiger (2005)	Government research archives (Federal Aviation Association)	Team mental models (IV) Performance	*Team mental models*: coding of semi-structured interviews and self-reported items from questionnaire *Performance*: flight delays	Endogenous	47 flight towers	2-year period	Yes (field data)
Sims, Drasgow, & Fitzgerald (2005)	Government research archives (U.S. Department of Defense)	Sexual harassment (IV) Turnover	*Sexual harassment*: self-report questionnaire *Turnover*: leave or stay	Endogenous	11,521	4.5 years	Yes (field data)
Keller (2006)	Company records	Charismatic leadership (IV) Job performance	*Charismatic leadership*: employees rated leader's behavior *Job performance*: products that went to market	Endogenous	118 project teams	1-year period	Yes (field data)
Lyness & Heilman (2006)	Company records	Job type/employee gender (IV) Job performance	*Job type*: employees' job title and gender *Job performance*: performance evaluations	Endogenous (job title) Exogenous (gender)	489	2-year period	Yes (field data)
Kuncel & Klieger (2007)	Academic records and media/ communication (*U.S. News* school ranking)	Performance (IV) Choice of law school	*Performance*: LSAT scores and student grade point average *Choice of law school*: application patterns across law schools when schools based on law school rankings	Endogenous	114 schools	Cross-sectional	Yes (combined academic records with media publication)
Timmerman (2007)	Sports data (Major League Baseball)	Geographic origin of pitcher (IV) Retaliation	*Geographic origin of pitcher*: coded whether pitchers were from a certain geographic location (southern U.S. state) *Retaliation*: plate appearance of every batter who appeared in the half inning after his own pitcher had hit an opposing batter	Exogenous	4,735,090 plate appearances	Cross-sectional	No

(continued)

Table 2 (continued)

Citation	Source of data	Example of construct(s) measured	Operationalization of construct(s)	Nature of IV (exogenous or endogenous)	Sample size	Time frame	Combined sources/ measures
Dierdorff & Ellington (2008)	Sponsored research activity (O*NET and General Social Survey)	Job characteristics (IV) Work-family conflict	*Job characteristics*: items taken from O*NET's job characteristics scores *Work-family conflict*: items taken from General Social Survey work-family module	Endogenous	1,367 individuals across 126 occupations	Cross-sectional	Yes (combined two sponsored research databases)
Wang, Zhan, Liu, & Shultz (2008)	Sponsored research activity (Health and Retirement Study data from the National Institute on Aging)	Individual attributes/ attitudes (age, education, job satisfaction, marital status) (IV) Bridge employment decision	*Individual attributes/attitudes*: items from survey *Bridge employment*: assessed from career changes in occupation codes	Endogenous	12,562	3 periods, 2 years apart	No
Barnes & Wagner (2009)	Sponsored research activity (Mine Safety and Health Administration; American Time Use Survey)	Daylight saving time (IV) Workplace injury Sleep	*Daylight saving time*: coded daylight saving day *Workplace injury*: reported number of injuries *Sleep*: self-reported hours of sleep	Exogenous	576,292 accidents reported; 14,310 for American Time Use Survey	24-year period for mining data; 3-year period for American Time Use Survey	Yes (combined two sponsored research databases)
Giebels & Taylor (2009)	Historiometry (audiotapes of crisis negotiations from police department) and previous research project data repurposed (Hofstede's collectivism-individualism country index scores)	Cultural differences (high vs. low context cultures) (IV) Message persuasion	*Cultural differences*: Hofstede's collectivism-individualism country index scores *Message persuasion*: coded audio transcripts for message persuasion	Endogenous	25 conflict negotiations	10-year period	Yes (combined research database with historiometry)
Judge, Ilies, & Dimotakis (2010)	Government research archives (Swedish Adoption Twin Study on Aging)	Socioeconomic status (IVs) Subjective well-being General mental ability and childhood	*Socioeconomic status*: income and education attainment *General mental ability*: test scores	Endogenous	398	4-year period	No

(continued)

Table 2 (continued)

Citation	Source of data	Example of construct(s) measured	Operationalization of construct(s)	Nature of IV (exogenous or endogenous)	Sample size	Time frame	Combined sources/ measures
Highhouse, Zickar, & Yankelevich (2010)	Sponsored research activity (General Social Survey)	Work ethic (IV) Job satisfaction	*Work ethic*: item about attitudes toward work *Job satisfaction*: item about job satisfaction	Endogenous	15,420	17-year period	No
Becker & Cropanzano (2011)	Company records	Job performance (IV) Turnover Tenure	*Job performance*: supervisor ratings *Turnover*: leave or stay and self-reported reasons *Tenure*: time spent in organization	Endogenous	2,570	5-year period	Yes (field data)
Barnes, Reb, & Ang (2012)	Sports data (National Basketball Association data)	Job performance (IV) Compensation	*Performance*: players' points per game *Compensation*: calculated from players' new and old contracts	Endogenous	131	2 National Basketball Association seasons	No
Wagner, Barnes, Lim, & Ferris (2012)	Sponsored research activity (Google.com search database)	Daylight saving time (IV) Cyberloafing	*Daylight saving time*: coded daylight saving time *Cyberloafing*: percentage of Internet searches conducted in the entertainment category	Exogenous	3,492 measurement points	3-day span, over course of 6 years	Yes (experimental)
Johnson & Allen (2013)	Sponsored research activity (Panel Study of Income Dynamics, O*NET)	Mother's job demands (IV) Child's health	*Mother's job demands*: items taken from O*NET's job characteristics scores *Child's health*: computed child's body mass index	Endogenous	654,696	4-year period	Yes (combined archival sources)
Minbashian & Luppino (2014)	Sports data (Association of Tennis Professionals)	Complexity of task (IV) Job performance	*Complexity of task*: quality of one's opponent, as indicated by ranking on the world tour *Job performance*: total number of points won by an individual as a percentage of the total points played in the match	Endogenous	393	9-year period	No

(continued)

Table 2 (continued)

Citation	Source of data	Example of construct(s) measured	Operationalization of construct(s)	Nature of IV (exogenous or endogenous)	Sample size	Time frame	Combined sources/ measures
Hirst, Van Knippenberg, Zhou, Quintane, & Zhu (2015)	Company records	Network efficiency (IV) Creativity	*Network efficiency*: proportion of an individual's direct ties that are not interconnected *Creativity*: self-report scale	Endogenous	11 divisions	Cross-sectional	Yes (field data)
Lee, Gino, & Staats (2014)	Government research archives (The National Climactic Data Center of the U.S. Department of Commerce) Company records	Weather conditions (IV) Employee productivity	*Weather*: precipitation levels *Productivity*: number of minutes a worker spent to complete the task	Exogenous	111 employees, completing 598,393 transactions	2-year period	Yes (survey and experimental data)
Dai, Milkman, Hofmann, & Staats (2015)	Company records	Hours spent at work (IV) Compliance with professional standards	*Hours spent at work*: items taken from records *Compliance with professional standards*: variable recording whether a caregiver washed his or her hands during a given hand hygiene opportunity	Endogenous	4,211	1-year period	Yes (field data)

Note: IV = independent variable.

that can be brought together in a complementary manner. Wagner, Barnes, Lim, and Ferris (2012) provide an example of this in their study of the influence of sleep on cyberloafing. Using a Google search database, Wagner and colleagues conducted an archival study of the clock changes associated with daylight saving time—known from the Barnes and Wagner (2009) study to be associated with lost sleep—on searches for entertainment Web sites. This archival study could not establish that such Internet searches occurred in a work setting and also did not directly measure sleep. Accordingly, Wagner and colleagues conducted an additional laboratory study that directly examined the effect of sleep on cyberloafing. The laboratory study was conducted in an artificial setting, which entails a different set of limitations. However, the two studies together provide stronger support for their model than the archival study alone. In considering complementary studies, one important question to ask is how the complementary study specifically addresses a limitation of the archival study. For example, if there were concern about endogeneity in a given relationship in an archival study, an especially helpful complementary study would be an experimental study involving a manipulation of the key independent variable.

Utilization of multiple types of studies is consistent with recent calls for full-cycle research. Chatman and Flynn define full-cycle organizational research as

> an iterative approach to understanding individual and group behavior in organizations, which includes: (a) field observation of interesting organizational phenomena, (b) theorizing about the causes of the phenomena, (c) experimental tests of the theory, and (d) further field observations that enhance understanding and inspire additional theorizing. (2005: 435)

Arguing that full-cycle organizational research is a powerful system for advancing our understanding of organizational behavior, Chatman and Flynn note that observational research has benefits that include (1) evidence that validates (or invalidates) assumptions about both whether phenomenon occur and whether hypothesized relationships occur in realistic settings, (2) determining the relevance of a phenomenon, and (3) identifying the complexity of the construct. We contend that archival research can be an additionally helpful tool for full-cycle research. Indeed, by including both a tightly controlled laboratory experiment and a large-scale study of naturally occurring behavior, Wagner and colleagues (2012) took a step in the direction of full-cycle research. This is consistent with the idea that by utilizing different types of tools (laboratory, field, archival research, and meta-analysis), we can triangulate the findings to converge on evidence that is more compelling than that generated by using only one tool.

The limitations regarding measurement and construct validity can thus be addressed through careful selection of archival data sources and variables and by complementing archival methodologies with methodologies that are more common in micro-organizational research (e.g., surveys, lab studies, field experiments). Indeed, we believe that the field of micro-organizational research has the luxury of adding archival studies in domains that are already supplemented by rigorous primary data approaches. Thus, assumptions about measurement of process variables can be relaxed to extend well-supported theory and uncover meaningful downstream implications. In other words, rather than being overly constrained by measurement issues that are inherent in archival research, micro-organizational research can take steps to address the limitations and in so doing open themselves to the benefits of archival research.

In combining different methods, it is worth noting that researchers should look beyond simply meta-analyzing the results of different studies. Although meta-analyses are appealing because they enable quantitative combination of multiple studies, Newman, Jacobs, and Bartram (2007) demonstrated that Bayesian analysis can provide more generalizable and accurate estimations when compared to meta-analytic or single study estimations. Whereas meta-analysis improves the precision of an estimation by calculating the mean validity across settings, single studies provide relatively imprecise estimations of the true local validity in a single setting (Newman et al.). One of the advantages of Bayesian analysis is its combination of random-effects meta-analytic estimates (Bayesian prior) with local validity study estimates. This new estimation is called Bayesian posterior (see Box & Tiao, 1973; Iversen, 1984). For a review of Bayesian analysis, see Newman et al.

Conclusion

As noted above, archival data have proven enormously useful in organizational research. However, micro-organizational researchers have been reluctant to consider archival data (Scandura & Williams, 2000). Historically, when micro-organizational researchers have used archival data, it has been for a relatively narrow range of topics related to compensation and turnover. There are many other potential sources of archival data, with new sources likely to appear in future years. The Big Data revolution is here, and micro-organizational researchers should make efforts to avoid being left behind. As Big Data picks up steam, the number of opportunities for relevant data should increase considerably. Thus, this discussion will hopefully provide some examples that will be useful to readers in considering what types of archival data sets are in existence and, perhaps more importantly, spark ideas for places to look for other sources.

Note

1. Although meta-analyses can be categorized as archival (e.g., Scandura & Williams, 2000), we consider this to be a different type of research that is already well leveraged in micro-organizational research.

References

Acemoglu, D., Johnson, S., & Robinson, J. A. 2001. The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91: 1369-1401.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. 2011. *Sentiment analysis of Twitter data*. In *Proceedings of the Workshop on Languages in Social Media:* 30-38. Stroudsburg, PA: Association for Computational Linguistics.

Antonakis, J., Bastardoz, N., Liu, Y., & Schriesheim, C. A. 2014. What makes articles highly cited? *The Leadership Quarterly*, 25: 152-179.

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

Barnes, C. M., & Drake, C. L. in press. Prioritizing sleep health: Public health policy recommendations. *Perspectives on Psychological Science*.

Barnes, C. M., Reb, J., & Ang, D. 2012. More than just the mean: Moving to a dynamic view of performance-based compensation. *Journal of Applied Psychology*, 97: 711-718.

Barnes, C. M., Schaubroeck, J. M., Huth, M., & Ghumman, S. 2011. Lack of sleep and unethical behavior. *Organizational Behavior and Human Decision Processes*, 115: 169-180.

Barnes, C. M., & Wagner, D. T. 2009. Changing to daylight saving time cuts into sleep and increases workplace injuries. *Journal of Applied Psychology*, 94: 1305-1317.

Barnes, C. M., Wagner, D. T., & Ghumman, S. 2012. Borrowing from sleep to pay work and family: Expanding time-based conflict to the broader non-work domain. *Personnel Psychology*, 65: 789-819.

Becker, W. J., & Cropanzano, R. 2011. Dynamic aspects of voluntary turnover: An integrated approach to curvilinearity in the performance–turnover relationship. *Journal of Applied Psychology*, 96: 233-246.

Bellizzi, J. A., & Hasty, R. W. 2002. Supervising unethical sales force behavior: Do men and women managers discipline men and women subordinates uniformly? *Journal of Business Ethics*, 40: 155-166.

Bernerth, J. B., Taylor, S. G., Walker, J. H., & Whitman, D. S. 2012. An empirical investigation of dispositional antecedents and performance-related outcomes of credit scores. *Journal of Applied Psychology*, 97: 469-478.

Bing, M. N., Stewart, S. M., Davison, H. K., Green, P. D., & McIntyre, M. D. 2007. Integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology*, 92: 722-744.

Bligh, M. C., Kohles, J. C., & Meindl, J. R. 2004. Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks. *The Leadership Quarterly*, 2: 211-239.

Bollen, K. A., & Brand, J. E. 2010. A general panel model with random and fixed effects: A structural equations approach. *Social Forces*, 89: 1-34.

Box, G. E. P., & Tiao, G. C. 1973. *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.

Briner, R. B., & Walshe, N. D. 2013. The causes and consequences of a scientific literature we cannot trust: An evidence-based practice perspective. *Industrial and Organizational Psychology*, 6: 269-272.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2): 15-21.

Cameron, L., Chaudhuri, A., Erkal, N., & Gangadharan, L. 2009. Propensities to engage in and punish corrupt behavior? Experimental evidence from Australia, India, Indonesia and Singapore. *Journal of Public Economics*, 93: 843-851.

Chan, D. 1998. Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83: 234-246.

Chatman, J. A., & Flynn, F. J. 2005. Full-cycle organizational psychology research. *Organization Science*, 16: 434-447.

Chen, Y., Leung, K., & Chen, C. 2009. Bringing national culture to the table: Making a difference through cross-cultural differences and perspectives. *The Academy of Management Annals*, 3: 217-249.

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

Conlon, D. E., Van Dyne, L., Milner, M., & Ng, K. Y. 2004. The effects of physical and social context on evaluations of captive, intensive service relationships. *Academy of Management Journal*, 47: 433-445.

Dai, H., Milkman, K. L., Hofmann, D. A., & Staats, B. R. 2015. The impact of time at work and time off from work on rule compliance: The case of hand hygiene in health care. *Journal of Applied Psychology*, 100: 846-862.

Detert, J. R., Treviño, L. K., Burris, E. R., & Andiappan, M. 2007. Managerial modes of influence and counterproductivity in organizations: A longitudinal business-unit-level investigation. *Journal of Applied Psychology*, 92: 993-1005.

Dierdorff, E. C., & Ellington, J. K. 2008. It's the nature of the work: Examining behavior-based sources of work-family conflict across occupations. *Journal of Applied Psychology*, 93: 883-892.

Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, J. D. 2007. Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology*, 92: 616-627.

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. 2007. A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods*, 10: 5-34.

Feldman, R. 2013. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4): 82-89.

Fioretti, G. 2013. Agent-based simulation models in organization science. *Organizational Research Methods*, 16: 227-242.

Giebels, E., & Taylor, P. J. 2009. Interaction patterns in crisis negotiations: Persuasive arguments and cultural differences. *Journal of Applied Psychology*, 94: 5-19.

Gino, F., Shu, L. L., & Bazerman, M. H. 2010. Nameless + harmless = blameless: When seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organizational Behavior and Human Decision Processes*, 111: 102-115.

Halaby, C. N. 2004. Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, 30: 507-544.

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

Highhouse, S., Zickar, M. J., & Yankelevich, M. 2010. Would you work if you won the lottery? Tracking changes in the American work ethic. *Journal of Applied Psychology*, 95: 349-357.

Hinds, P. J., Liu, L., & Lyon, J. B. 2012. Putting the global in global work: An intercultural lens on the process of cross-national collaboration. *The Academy of Management Annals*, 5: 1-54.

Hirst, G., Van Knippenberg, D., Zhou, J., Quintane, E., & Zhu, C. 2015. Heard it through the grapevine: Indirect networks and employee creativity. *Journal of Applied Psychology*, 100: 567-574.

Hitt, M. A., Boyd, B. K., & Li, D. 2004. The state of strategic management research and a vision of the future. *Research Methodology in Strategic Management*, 1: 1-31.

Hollenbeck, J. R., DeRue, D. S., & Mannor, M. J. 2006. Statistical power and parameter stability when subjects are few and tests are many: Comment on Peterson, Smith, Martorana, and Owens (2003). *Journal of Applied Psychology*, 91: 1-5.

Hunter, J. E., & Schmidt, F. L. 2004. *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M. D., & Jundt, D. K. 2005. Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology*, 56: 517-543.

Ioannidis, J. P. 2014. How to make more published research true. *PLoS Medicine*, 11(10): e1001747. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001747

Iversen, G. R. 1984. *Bayesian statistical inference*. Beverly Hills, CA: Sage.

Johnson, R. C., & Allen, T. D. 2013. Examining the links between employed mothers' work characteristics, physical activity, and child health. *Journal of Applied Psychology*, 98: 148-157.

Judge, T. A., Hurst, C., & Simon, L. S. 2009. Does it pay to be smart, attractive, or confident (or all three)? Relationships among general mental ability, physical attractiveness, core self-evaluations, and income. *Journal of Applied Psychology*, 94: 742-755.

Judge, T. A., Ilies, R., & Dimotakis, N. 2010. Are health and happiness the product of wisdom? The relationship of general mental ability to educational and occupational attainment, health, and well-being. *Journal of Applied Psychology*, 95: 454-468.

Judge, T. A., & Kammeyer-Mueller, J. D. 2012. On the value of aiming high: The causes and consequences of ambition. *Journal of Applied Psychology*, 97: 758-775.

Keller, R. T. 2006. Transformational leadership, initiating structure, and substitutes for leadership: A longitudinal study of research and development project team performance. *Journal of Applied Psychology*, 91: 202-210.

Kepes, S., & McDaniel, M. A. 2013. How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology*, 6: 252-268.

Kouloumpis, E., Wilson, T., & Moore, J. 2011. Twitter sentiment analysis: The good the bad and the OMG! In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*: 538-541. Palo Alto, CA: AAAI Press.

Kozlowski, S. W. J., & Klein, K. J. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions*: 3-90. San Francisco: Jossey-Bass.

Kuncel, N. R., & Klieger, D. M. 2007. Application patterns when applicants know the odds: Implications for selection research and practice. *Journal of Applied Psychology*, 92: 586-593.

Larrick, R. P., Timmerman, T., Carton, A., & Abrevaya, J. 2011. Temper, temperature, and temptation: Heat-related retaliation in baseball. *Psychological Science*, 22: 423-428.

Leavitt, K., Mitchell, T., & Peterson, J. 2010. Theory pruning: Strategies for reducing our dense theoretical landscape. *Organizational Research Methods*, 13: 644-667.

LeBreton, J. M., & Senter, J. L. 2008. Answers to twenty questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11: 815-852.

Lee, J. J., Gino, F., & Staats, B. R. 2014. Rainmakers: Why bad weather means good productivity. *Journal of Applied Psychology*, 99: 504-513.

Lyness, K. S., & Heilman, M. E. 2006. When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91: 777-785.

Martell, R. F., Lane, D. M., & Emrich, C. 1996. Male-female differences: A computer simulation. *American Psychologist*, 51: 157-158.

Minbashian, A., & Luppino, D. 2014. Short-term and long-term within-person variability in performance: An integrative model. *Journal of Applied Psychology*, 99: 898-914.

Mirsky, S. 2014. Workplace injuries may rise right after daylight saving time. *Scientific American*, March 9. http://www.scientificamerican.com/podcast/episode/time-change-injuries/. Accessed August 17, 2015.

Mitchell, T. R., & James, L. R. 2001. Building better theory: Time and the specification of when things happen. *Academy of Management Review*, 26: 530-547.

Murphy, K. R., & Myors, B. 1998. *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Mahwah, NJ: Erlbaum.

Newman, D. A., Jacobs, R. R., & Bartram, D. 2007. Choosing the best method for local validity estimation: Relative accuracy of meta-analysis versus a local study versus Bayes-analysis. *Journal of Applied Psychology*, 92: 1394-1413.

Newman, D. A., & Sin, H. 2009. How do missing data bias estimates of within-group agreement? Sensitivity of SDWG, CVWG, rWG(J), rWG(J)*, and ICC to systematic nonresponse. *Organizational Research Methods*, 12: 113-147.

Ng, K. Y., & Van Dyne, L. 2001. Individualism-collectivism as a boundary condition for effectiveness of minority influence in decision making. *Organizational Behavior and Human Decision Processes*, 84: 198-225.

Nosek, B. A., Spies, J. R., & Motyl, M. 2012. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7: 615-631.

Nunnally, J. C., & Bernstein, I. 1994. *Psychometric theory*. New York: McGraw Hill.

Pang, B., & Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2): 1-135.

Popper, K. 1959. *The logic of scientific discovery*. London: Routledge.

Pritchard, R., Jones, S., Roth, P., Stuebing, K., & Ekeberg, S. 1988. Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology*, 73: 337-358.

Probst, T. M., Brubaker, T. L., & Barsotti, A. 2008. Organizational injury rate underreporting: The moderating effect of organizational safety climate. *Journal of Applied Psychology*, 93: 1147-1154.

Roberts, B. W., Caspi, A., & Moffitt, T. E. 2003. Work experiences and personality development in young adulthood. *Journal of Personality and Social Psychology*, 84: 582-593.

Rosenthal, R., & Rosnow, R. L. 1991. *Essentials of behavioral research: Methods and data analysis*. Boston: McGraw Hill.

Sakaluk, J. K., Williams, A. J., & Biernat, M. 2014. Analytic review as a solution to the misreporting of statistical results in psychological science. *Perspectives on Psychological Science*, 9: 652-660.

Scandura, T. A., & Williams, E. A. 2000. Research methodology in organizational studies: Current practices and implications for future research. *Academy of Management Journal*, 43: 1248-1264.

Schwab, D. P. 1980. Construct validity in organizational behavior. *Research in Organizational Behavior*, 2: 3-43.

Shadish, W. R., Cook, T. D., & Campbell, D. T. 2001. *Experimental and quasi-experimental designs for generalized causal inference*. New York: Wadsworth.

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. 2011. Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology*, 96: 1055-1064.

Silberzahn, R., Simonsohn, U., & Uhlmann, E. L. 2014. Matched names analysis reveals no evidence of name meaning effects: A collaborative commentary on Silberzahn and Uhlmann (2013). *Psychological Science*, 25: 1504-1505.

Silberzahn, R., & Uhlmann, E. L. 2013. It pays to be Herr Kaiser: Germans with noble-sounding surnames more often work as managers. *Psychological Science*, 24: 2437-2444.

Simmons, J., Nelson, L., & Simonsohn, U. 2011. False-positive psychology: *Undisclosed* flexibility in data collection and analysis allow presenting anything as significant. *Psychological Science*, 22: 1359-1366.

Simonsohn, U. 2013. Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science*, 24: 1875-1888.

Sims, C. S., Drasgow, F., & Fitzgerald, L. F. 2005. The effects of sexual harassment on turnover in the military: Time-dependent modeling. *Journal of Applied Psychology*, 90: 1141-1152.

SINTEF. 2013. Big Data, for better or worse: 90% of world's data generated over last two years. *ScienceDaily*, May 22. http://www.sciencedaily.com/releases/2013/05/130522085217.htm. Accessed August 13, 2014.

Smith-Jentsch, K. A., Mathieu, J. E., & Kraiger, K. 2005. Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. *Journal of Applied Psychology*, 90: 523-535.

Spector, P. E., Allen, T. D., Poelmans, S. A. Y., Lapierre, L. M., Cooper, C. L., O'Driscoll, M., Sanchez, J. I., Abarca, N., Alexandrova, M., Beham, B., Brough, P., Ferreiro, P., Fraile, G., Lu, C. Q., Lu, L., Moreno-Veláquez, I., Pagon, M., Pitariu, H., Salamatov, V., Shima, S., Suarez Simoni, A., Ling Siu, O., & Widerszal-Bazyl, M. 2007. Cross-national differences in relationships of work demands, job satisfaction and turnover intentions with work-family conflict. *Personnel Psychology*, 60: 805-835.

Spector, P. E., Cooper, C. L., Poelmans, S., Allen, T. D., O'Driscoll, M., Sanchez, J. I., Siu, O. L., Dewe, P., Hart, P., Lu, L., de Moraes, L. F. R., Ostrognay, G. M., Sparks, K., Wong, P., & Yu, S. 2004. A cross-national comparative study of work/family stressors, working hours, and well-being: China and Latin America vs. the Anglo world. *Personnel Psychology*, 57: 119-142.

Spector, P. E., Cooper, C. L., Sanchez, J. I., O'Driscoll, M., Sparks, K., Bernin, P., Büssing, A., Dewe, P., Hart, P., Lu, L., Miller, K., Renault de Moraes, L., Ostrognay, G. M., Pagon, M., Pitariu, H., Poelmans, S., Radhakrishnan, P., Russinova, V., Salamatov, V., Salgado, J., Shima, S., Siu, O. L., Stora, J. B., Teichmann, M., Theorell, T., Vlerick, P., Westman, M., Widerszal-Bazyl, M., Wong, P., & Yu, S. 2002. A 24 nation/territory study of work locus of control in relation to well-being at work: How generalizable are western findings? *Academy of Management Journal*, 45: 453-466.

Timmerman, T. A. 2007. It was a thought pitch: Personal, situational, and target influences on hit-by-pitch events across time. *Journal of Applied Psychology*, 92: 876-884.

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. 2009. Composing group-level constructs from individual-level survey data. *Organizational Research Methods*, 12: 368-392.

Varnum, M. E., & Kitayama, S. 2010. What's in a name? Popular names are less common on frontiers. *Psychological Science*, 22: 176-183.

Wagner, D. T., Barnes, C. M., Lim, V. K. G., & Ferris, D. L. 2012. Lost sleep and cyberloafing: Evidence from the laboratory and a daylight saving time quasi-experiment. *Journal of Applied Psychology*, 97: 1068-1076.

Wang, M., Zhan, Y., Liu, S., & Shultz, K. S. 2008. Antecedents of bridge employment: A longitudinal investigation. *Journal of Applied Psychology*, 93: 818-830.

Weller, I., Holtom, B. C., Matiaske, W., & Mellewigt, T. 2009. Level and time effects of recruitment sources on early voluntary turnover. *Journal of Applied Psychology*, 94: 1146-1162.

Wimbush, J. C., & Dalton, D. R. 1997. Base rate for employee theft: Convergence of multiple methods. *Journal of Applied Psychology*, 82: 756-763.