

Using Administrative Data for Social Science and Policy



ANDREW M. PENNER AND KENNETH A. DODGE

Several years ago, the director of social services in Durham, North Carolina, came to our Durham Connects Program research team with a request for collaboration. He administered several dozen “pockets” of dollars allocated for families with young children for programs such as childcare subsidies, housing loans, and Early Head Start. His administrative data systems detailed which families received the dollars, but the files were not linked with each other and had not been used to improve accountability or impact. On our side, we had been collecting interview data from the entire population of families at birth and therefore knew what resources families needed to support their infants’ healthy development. We teamed up with him, merged and analyzed the combined data files, and gained many insights (for example, some families were receiving many thousands of dollars in support but still living in poverty).

The findings were often translated into public policy. For example, the population’s need for substance abuse treatment for mothers of infants far exceeded the capacity of the community’s allocated resources in this domain,

and so the juxtaposition of the need and the capacity was presented to the county commissioners, who then increased their allocation. In another example, a high no-show rate at the community health clinic was discovered and thought to impede delivery of health care and cost many dollars in inefficiency; interviews with mothers showed that they failed to show up because the bus did not stop in the housing project and was too expensive. So the City Council changed the bus route to be more accessible and passed an ordinance to allow anyone carrying a baby to ride the city bus for free. No-shows at the clinic declined. A bonus of this collaboration was that in return for providing useful research findings to the community, university scholars were allowed to use the data files for their research.

This case study illustrates the tremendous potential of a university-community collaboration using administrative data. In addition to the profound human service gains, administrative data infrastructure has important potential to improve cutting-edge social science research. Not only do administrative data facilitate the

Andrew M. Penner is professor of sociology at the University of California, Irvine. **Kenneth A. Dodge** is Pritzker Professor of Early Learning Policy Studies and professor of psychology and neuroscience at Duke University.

© 2019 Russell Sage Foundation. Penner, Andrew M., and Kenneth A. Dodge. 2019. “Using Administrative Data for Social Science and Policy.” *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 1–18. DOI: 10.7758/RSF.2019.5.3.01. We are grateful to Nicole Deterding, Thad Domina, David Grusky, Paul Hanselman, NaYoung Hwang, Ryan Lewis, Silvia Melzer, Matthew Snipp, and participants in the Russell Sage Foundation administrative data conference for useful comments and discussions. Direct correspondence to: Andrew M. Penner at penner@uci.edu, University of California, Irvine, 4181 Social Science Plaza, Irvine, CA 92697.

Open Access Policy: *RSF: The Russell Sage Foundation Journal of the Social Sciences* is an open access journal. This article is published under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

evaluation of policies and interventions, they also enable researchers to address novel questions. Administrative data in many countries are routinely collected and used to improve policy and contribute to science to a much greater degree than in the United States. In this double issue, we hope to improve collaborations between American communities and academic scholars.

The value of “big data” is becoming increasingly apparent. In the current data revolution, firms’ profits are determined by their ability to capture and analyze data. The capacity to collect and analyze data determines not only the winners and losers in the economy, but also which societies can best educate their citizens, train their workforce, keep their population healthy, and promote the well-being and flourishing of their citizens. Schools, hospitals, and many other organizations now routinely collect data that allow them to serve their students, patients, and other stakeholders more effectively. Understanding the lessons contained in these data, and how best to extract them, plays an important role in helping practitioners, administrators, and policymakers understand the challenges that the American public faces and the effects of current practices and policies.

Recent federal initiatives such as the Murray-Ryan Evidence-Based Policymaking Commission Act of 2016 and the resulting Foundations for Evidence-Based Policymaking Act of 2018 suggest that administrative data will become increasingly central in U.S. social science and policy. However, efforts to leverage administrative data in the social sciences are uneven. In domains such as public health and demography, administrative data are used routinely, while they are less frequently used in mental health, social behavior, and other areas. But even where administrative data are commonly used, current capacity constraints hinder their utility. Data that could be used to inform decisions are often unavailable to those making the decisions, leaving them with incomplete information.

One clear example is in education. Despite their focus on preparing students who are “college- and career-ready,” schools have historically struggled to obtain data on the practices that will prepare their students to be successful because widespread links between

students in K–12 educational systems and higher education outcomes have become available only recently, and links between K–12 data systems and the labor market remain relatively rare. These data linkages are important to understand the efficacy of school-based vocational programs, dropout recovery interventions, college readiness programs, and advancement placement course policies. But schools, like other organizations, typically lack the capacity and expertise to build this infrastructure and analyze the resulting data.

This double issue opens with a section that includes an article by Sean Reardon showing the potential value of examining nationwide data on student achievement (2019). He amasses an unprecedented administrative data file and then breaks it down in many ways that are useful to policy decision making as well as scientific understanding of child development and the impact of education. This file, and others like it, can be used to address numerous important issues that were previously unavailable to rigorous empirical scrutiny. For example, Megan Austin, Joseph Waddington, and Mark Berends show how administrative data can inform the important policy question of the impact of school vouchers on student achievement (2019).

Administrative data research seeks to bring the data-driven approach increasingly used in fields such as micro-targeting (where commercial firms individualize advertising according to one’s buying patterns) and precision medicine to help answer important societal questions, providing schools, clinics, and others serving the public good with the information they need to work more effectively. To the degree that public-sector institutions are unable to build the infrastructure needed to understand the effectiveness of their programs, it seems likely that this lack of data infrastructure will result in inefficiencies (particularly relative to private sector institutions with fewer restrictions regarding data).

In this article and the double issue that follows, we seek to highlight the promise of administrative data to answer pressing questions for both science and policy. Our goal is not to be comprehensive, but to provide an indication of the potential that is currently barely being

tapped, offer suggestions for advancing this infrastructure, and highlight some of the challenges currently facing the field. Likewise, our goal is to spark future work by highlighting the promise of these data by showcasing a few exciting administrative data initiatives and research projects.

THE CASE FOR ADMINISTRATIVE DATA INFRASTRUCTURE

Administrative data have many benefits for research that seeks to advance science and inform policy. These data typically originate not for research purposes, but instead as client-level service records or from administrators who are accountable for documentation of implementation of services. In many cases, the information available in administrative datasets and the analyses of these data will be useful to policymakers precisely because these are the variables for which policymakers are accountable. Furthermore, these datasets often attempt to reach the entire population of relevant participants and suffer less from selection bias than many original research datasets. Thus, not only does administrative data infrastructure provide important opportunities to understand the social world better, but it does so in contexts where stakeholders are ready to act based on the results of research. As a result, the analysis of administrative data has the potential to bridge basic and applied research.

Appropriate data infrastructure allows broad and equitable access, lessens the burden on the organizations providing the data, standardizes expectations around availability and privacy, and allows for best practices in data security. A robust administrative data infrastructure also lowers the marginal cost of conducting additional research, allowing researchers to address important issues using existing data that do not incur the effort or pecuniary costs of new data collection. Although the startup cost of establishing this infrastructure is not cheap, once established, it raises the quality of the research it permits and allows for more research by changing the cost curve for high-quality additional research. Furthermore, accessible administrative data files enable replication of data analyses and findings across independent research teams using the same and different files,

which improves the robustness of the field. In a world where data analytics and knowledge are increasingly central to society, strong administrative data infrastructure is therefore not only important for knowledge creation, but also efficient.

Administrative data systems operate at scale. This feature allows researchers to consider questions such as whether all individuals benefit equally from a particular policy. When combined with research designs that allow for causal inference, these population-level data files afford answers about whether any individuals or groups are negatively affected and other questions of treatment-effect heterogeneity that are often prohibitively expensive to consider without the scale that administrative data provide. Having data on the entire universe of relevant individuals also allows for comparisons of small groups that are otherwise extremely difficult to isolate, for example, comparing employees with people doing the same job in the same establishment (Petersen and Morgan 1995). Operating at scale means that one can consider treatment not just at the individual level, but also at the community level, examining how spillover effects, feedback loops, and other dynamics might cause policies to work differently when implemented broadly. Finally, of course, having the population of participants available for analysis minimizes (but certainly does not eliminate) problems that survey researchers confront, such as biased attrition and missing data.

We next highlight several useful features of administrative data that are instrumental to advancing science and policy.

A Long-Term Perspective

Recent research has highlighted the utility of administrative data for understanding the long-term outcomes associated with a variety of interventions. Whereas much policy is necessarily driven by research and evaluations examining relatively short-term outcomes, administrative data can provide a longer-term perspective on the effects of past policies and thus a more accurate account of a policy's costs and benefits. Beyond following individuals who were exposed to interventions over time, administrative data can provide an important tool

for understanding multigenerational cycles of advantage and disadvantage, allowing researchers to trace the descendants of individuals from different backgrounds as well as the multigenerational effects of antipoverty policies and interventions. Administrative data also provide opportunities to connect researcher-collected data (for example, observations of student behavior) with future administrative records to help understand the long-term implications of researchers' observations (such as longer-term cost outcomes related to early school behavior problems).

Administrative data research has pushed the boundaries of what we know about longer-term temporal processes, highlighting the important implications of understanding longer-term policy effects, as well as the intergenerational transmission of advantage and disadvantage more broadly. Standard intervention research typically ends data collection at the end of the study and provides an evaluation at that point. In some large-scale studies, follow-ups track outcomes for multiple years post-intervention (see, for example, Ludwig et al. 2013). Insofar as participants in interventions continue to be captured in administrative records after the end of the study, administrative records provide opportunities for an efficient approach to understanding the benefits of interventions decades later. Given the importance of understanding the long-term impacts of policies aimed at ameliorating poverty (Bailey et al. 2017), and the goal of changing not just individuals' current circumstances but also their longer-term trajectories, one of the most exciting features of administrative data is their ability to turn any study into a *de facto* longitudinal study. Recent work in this vein, for example, has highlighted the adult income effects of having an experienced or effective kindergarten teacher and higher-achieving classmates (Chetty et al. 2011; Chetty, Friedman, and Rockoff 2014a, 2014b), the increases in adult civic participation arising from a psychosocial intervention in elementary school (Holbein 2017), and the benefits of cash transfers to poor families on their children's educational outcomes, mortality, and income (Aizer, Eli et al. 2016).

Administrative data research has also extended a long tradition of survey-based re-

search highlighting the intergenerational transmission of advantage and disadvantage (Blau and Duncan 1967; Ganzeboom, Treiman, and Ultee 1991). Survey-based research has examined parent-child transmission processes (for example, Campbell et al. 2014; Poulton et al. 2002), and in some cases, grandparents (Wightman and Danziger 2014). Administrative data allow researchers to push back on the time horizon of these studies. Using Swedish registry data, for example, Martin Hällsten examines the effects of not only grandparents, but also great-grandparents (2014). Likewise, work by Hyunjoon Park uses Korean historical registry information to examine the outcomes of ethnically Korean slaves, showing substantial effects on their descendants many generations after the abolition of slavery (2014). Multigenerational data are somewhat more rare in the United States, the Utah Population Database being a notable exception (see, for example, Temby and Smith 2014). A more detailed discussion of the opportunities provided by administrative data for understanding multigenerational processes is available from Xi Song and Cameron Campbell (2017). These data provide not only exciting new opportunities to advance the science of the multigenerational transmission of advantage, but also the long-term costs of stigmatized identities, poverty, and disadvantage. As data infrastructure matures, we expect to see more studies examining how social policies aimed at one generation affect their children and grandchildren (for example, Meghir, Palme, and Schnabel 2012).

Key Sites and Populations in the Production of Inequality

Inequality in opportunities and outcomes across groups and persons is one of the most vexing problems facing contemporary society. These inequalities are often produced in spaces that are difficult to examine using surveys or experiments. In the hiring process, for example, correspondence studies can examine who receives responses from potential employers but cannot help us understand how applicant pools are created, which interviewees receive offers, or whether pay differences exist among those who receive offers. Research on inequality in other institutional spheres such as health care,

housing, and higher education faces similar challenges. Inequality in institutions is often shaped by processes that determine who is included and excluded, and how individuals are ranked within an organization. Administrative data on hiring pipelines, performance ratings, promotion decisions, and decisions about termination can provide valuable insights into the decisions made by gatekeepers at key sites regarding entry and exit from organizations, as well as important processes governing intra-organizational inequality regimes. One useful feature of workplace administrative records is that they allow researchers to compare individuals with others who have the same occupation working in the same establishment. Research making such comparisons suggests that much of the wage inequality observed across gender and race is created by sorting processes, as individuals doing the same work for the same employer receive largely similar pay (Petersen and Morgan 1995; Tomaskovic-Devey 1993).

Building on this insight, a long tradition of case studies uses administrative records from company human resource departments to understand inequality in these sorting processes. Trond Petersen and Ishak Saporta argue that in the current institutional context the opportunity for firms to discriminate is greatest at the point of hire (as opposed to discrimination in promotion or termination practices) and show that hiring is where the largest differences between men and women are observed (2004). Roberto Fernandez and his collaborators use human resource data in a series of papers that provide important insights into race and gender inequality in the hiring process; they show, for example, the importance of referrals (Fernandez and Greenberg 2013) and the importance of supply-side adjustments to perceived demand-side constraints (Fernandez and Friedrich 2011; Fernandez and Campero 2017). Recent work in this vein highlights the perhaps surprising egalitarian influence of an executive search firm (Fernandez-Mateo and Fernandez 2016), and in the current double issue Fernandez and Brian Rubineau use their extraordinary hiring data to provide novel analysis of network recruitment efforts and their impact on the gender-based glass ceiling in the biopharma industry (2019).

Beyond the labor market, administrative data from other sources also provide important insights into key sites for generating inequality. Research on NIH funding decisions, for example, uses detailed records to document the existence of gender inequality in the NIH review process (Li 2012). Research examining the criminal justice system uses administrative records from juvenile courts to estimate the effects of juvenile incarceration on later criminal justice and school outcomes (Aizer and Doyle 2015). One of the supposed keys to combating inequality and reaching life success is, of course, education. Although we know that dropping out of high school breeds failure and college graduation brings success, much less clear is the value of post-high school associate degrees, vocational diplomas, certificates, and partial college. In this double issue, ChangHwan Kim and Christopher Tamborini merge school administrative data files with earnings files to examine the long-term earnings that accrue from these post-high school accomplishments (2019).

Janelle Downing and Tim Bruckner use housing foreclosure administrative records and birth records to highlight yet another source of inequality (2019). They show that housing foreclosures (and presumably the stress they cause) contribute to premature births and increase inequality in birth outcomes across race and ethnic groups.

Administrative data also allow us to understand small, often difficult to access, populations that are theoretically important. For example, research using large administrative datasets has shown that millionaire tax flight does not occur at levels that are socially meaningful (Young et al. 2016), and that top earners are increasing isolated from the rest of the population (Godechot 2013). Insofar as many large administrative datasets include information on the whole population, these data allow researchers to examine relatively small and theoretically important groups (for example, those that are hard to capture in a probability sample without an explicit oversample) without compromising representativeness (Liebler, Bhaskar, and Porter 2016). From a local policy perspective, the ability to identify small groups of people is helpful because it allows policymak-

ers to ensure that the policies they implement are having their intended effects for all stakeholders, and to determine where adjustments to existing policies are needed (Howard et al. 2019).

Understanding Individuals in Their Social Contexts

The importance of context is a truism in social science research. From network influences to cultural factors to questions about positional goods, contextual considerations play a profound role in shaping an individual's outcomes. Despite this, interventions and policies have historically operated from a baseline that presupposes constant universal effects that operate at scale if implemented with fidelity (Dodge 2011). The density of information in administrative data is useful in providing opportunities to examine important sources of heterogeneity (particularly contextual sources, but also individual-level factors), as well as providing opportunities to investigate policies and interventions at scale. Durham Connects, for example, takes advantage of administrative data by assigning all newborns in Durham born on even days to receive a nurse home visit (Dodge et al. 2014). This design allows for children to be followed in administrative records throughout their lives in an ethical manner without the requirement of individual consent (because data can be de-identified before being analyzed but retain the essential characteristic of assignment to intervention) without necessitating additional data collection.

Further, thinking about changes to the social system more broadly (for example, moving bus stops to help mothers travel to local clinics) shifts research and policy discussions away from a methodological individualism that focuses on the effects of treatments on individuals, and toward considering how programs and policies affect social systems more broadly (see, for example, Denice and Gross 2016). These systems-level approaches are important for both science and policy, as they address an important shortcoming of much social science research. Research often asks what would happen if everything was held constant and only one consideration was changed. This can be instructive, but it fails to take into account the

myriad of ways that people and their social worlds are interconnected. Analyses of how community-level policies and interventions shape not only individuals' outcomes, but also society more generally, allow researchers to capture the complicated feedback loops and spillover effects that occur when interventions and policies operate at scale, providing insight into how policies might change society more broadly (Dodge 2009; Penner et al. 2015). Although in theory such analyses are possible without using administrative data, in practice the existence of administrative records greatly facilitates them.

Research using administrative data can also provide a more complete account of certain aspects of context, including the government services context. For example, Robert Goerge and Emily Wiegand examine the overlap in families' access of government services across multiple agencies to show how some families are accessing many agencies whereas others seem to be underaccessing resources that might benefit them (2019). Goerge and Wiegand show that these differences vary across geographical locations within Illinois, suggesting that local practice might contribute to, and mitigate, any biases. Lanikue Howard and colleagues examine the relation between parents' payment of child support and children's involvement in the child welfare system (2019). Agustina Laurito and colleagues combine multiple administrative datasets to show how school climate and neighborhood crime levels affect student achievement (2019).

Administrative data can also afford studies that are simply not plausible through original data collection. For example, it is difficult to imagine survey data tracking all of the classmates that a student had or all of the co-workers over an employee's career, but educational administrative data and linked employer-employee datasets often include this information (Abowd, Haltiwanger, and Lane 2004). Administrative records can thus provide information not only about an individual research subject but also about the environment surrounding them. Christopher Candelaria exploits such data in education, where he disentangles the long-term effects of a third-grade teacher, the medium-term effects of middle-

school teachers, and the short-term effects of an eighth-grade teacher (2015).

Linking administrative data files across levels also provides innovative opportunities to understand individual behavior in broader context. Elizabeth Ananat and her colleagues, for example, link county-year-level administrative data about community-level job loss with individual student educational administrative data files in order to discover the impact of local economic downturns on student academic progress (2017). Further, as noted previously, administrative data allow for individuals to be placed in a multigenerational familial context. In providing dense coverage of populations, these data allow researchers to examine whether policies had spillover effects (either positive or negative) on those around the targeted populations, and to examine questions around how context moderates the effectiveness of treatment. Heterogeneity in treatment effectiveness is important not only for contributing to scientific understanding regarding the mechanisms through which interventions work, but also because it has important implications for generalizability and scalability (Domina et al. 2016).

An Iterative Policy Design, Implementation, and Evaluation Cycle

A common goal of scientific research is to improve societal outcomes. Social scientists often seek to do this by evaluating and informing existing policies, and it is not uncommon to hear researchers bemoan the lack of policy responsiveness to research. Although obtaining access to administrative data can be time consuming in some contexts, a potential advantage of administrative data analysis in many contexts is that the institution generating the data is also making and implementing the policies being evaluated, so that there is an audience that is positioned to make decisions about practice and policy based on researchers' findings (Howard et al. 2019). This is particularly true in researcher-practitioner partnerships, where researchers partner with organizations to help them use their administrative data in better ways to answer questions of interest to decision makers and stakeholders.

The research-policy link in social science is

often conceptualized as one in which research informs or evaluates policies, but in research-practitioner partnerships, policy implementation and research can have a bidirectional synergistic relationship. These partnerships provide data researchers could not otherwise access. This unusual opportunity holds not only for companies' human resource data, but also for educational data, where laws protecting the privacy of student data allow data to be shared with organizations conducting research on behalf of educational agencies to help improve instruction. Further, researchers not only have the opportunity to study important policies in real-world settings but can often inform the implementation process. This relationship not only allows for policies that draw on researchers' expertise, but can also lead to opportunities for better research because researchers can help implement policies in ways that facilitate high-quality evaluations (such as introducing lottery-based assignments for oversubscribed programs or thresholds in assignment scores that enable regression discontinuity-based designs). In contexts that facilitate the timely incorporation of feedback, data collected can be used to inform Bayesian adaptive designs to help improve interventions in real time (Finucane, Martinez, and Cody 2017).

By bringing researchers into the policymaking process, rigorous research becomes part of the iterative process of policy implementation and adjustment and the policy adjustments made in implementation are better captured by research (Howard et al. 2019). This approach can provide both better policy and better research and serves as a model for how to accumulate and incorporate knowledge beyond simply conducting a series of single-policy evaluations. Although shortening the feedback loop among implementation, research, and redesign is likely to be positive overall, one challenge of this linkage is that long-term outcomes by definition cannot be observed quickly, and many policies—particularly those aimed at implementing organizational changes—have effects that take time to emerge or vary across different stages of the implementation cycle (Mills and Wolf 2017; Sun, Penner, and Loeb 2017).

EXPANDING ADMINISTRATIVE DATA INSIGHTS

Existing administrative data have allowed researchers to address a variety of important questions, and in many contexts, administrative data provide the best opportunities to answer important policy questions (Austin, Waddington, and Berends 2019). Current infrastructure, however, constrains research and limits the ability to answer questions that are important for science and policy. In the section that follows, we highlight important frontiers for administrative data research, highlighting noteworthy exemplars of work in these areas. We frame the points as strengths, but they could also be conceptualized as ways to address the potential weaknesses of administrative data.

Combining Datasets Across Sources

Although current administrative data research focuses on contemporary data sources, there is a long tradition of using administrative records in archival and historical research (see, for example, Kessler-Harris 1982; Wilde 2004). Research using administrative data has much in common with history and archeology, insofar as it observes the tracks that individuals leave as they move through society and draws lessons from these glimpses into their lives. A key difference is that when records outlast people, opportunities for supplementing and triangulation through interviews, surveys, or ethnography decline, leaving scientists to reconstruct meaning from the traces people have left behind. Although administrative data researchers using contemporary data draw conclusions from the traces left behind in current records in a similar manner, research using contemporary records has the potential to incorporate information directly from individuals through surveys and observations to supplement the data in administrative records.

Given their origin in a particular institutional context, administrative records are typically fragmented, and these data are often not linked to other data that would be useful for research and policy. Hospitals, for example, collect detailed information about patients' health, schools regularly collect information about student development, and employers often keep records not only about the perfor-

mance of employees, but also about applicants who were ultimately not offered positions. Although various combinations of these data can provide important insights, they are typically compartmentalized. Likewise, given their origin, administrative records often lack certain kinds of information that are less likely to be collected in these records. For example, information about attitudes, affinities, and motives are not often collected in administrative records. Combining administrative data with records from other sources—either by linking administrative records across sources or by making administrative records available to be linked to data collected via other means—is thus central to building administrative data infrastructure.

Linking Administrative Data Records Across Domains

By virtue of how they come into existence, administrative data are typically focused on one facet of an individual's life, and data and insights are often siloed. Given that the potential for insight grows exponentially as data are integrated, combining administrative data across domains is of vital importance, and enables researchers to trace connections between settings like schools, criminal justice institutions, health organizations, and employers, and see how inequalities compound across these domains. Our introduction describing Durham Connects highlights the power of these insights for understanding the needs of families across diverse domains, and others likewise underscore the utility of linking administrative records across domains to understand the challenges facing families in poverty (Goerge and Wiegand 2019). Research linking data across domains documents how inequality in one domain shapes outcomes in others, highlighting, for example, the health consequences of foreclosure (Downing and Bruckner 2019) and how air pollution shapes mortality risk (Di et al. 2017). Other research in this vein allows us to understand the broad effects of policies, showing how lead abatement efforts lower children's blood lead levels and improve student achievement (Aizer, Currie et al. 2016), and how Superfund site cleanup improves children's later educational outcomes (Persico, Figlio, and Roth 2016).

Beyond helping us understand disadvantage better at any given point in time, linking data across domains can also open opportunities to follow individuals as they move through different institutional settings. Administrative records from birth, education, criminal justice, labor market, and mortality often capture different points in an individual's life; combining data across these stages allows us to understand how inequalities unfold over the arc of an individual's life. For example, research linking educational records with IRS records highlights the long-term income benefits associated with high-quality teachers (Chetty, Friedman, and Rockoff 2014a, 2014b) as well as the link between college major choices and later life income (Kim and Tamborini 2019). Similarly, research on the school-to-prison pipeline in Texas links education and justice records to trace the juvenile justice involvement of students suspended from school (Fabelo et al. 2011).

Much of the attention in administrative data infrastructure has focused on large-scale population-level data. However, as noted earlier, one of the potential advantages of using administrative data is that they provide information about social processes that are otherwise very difficult to study (such as the hiring pipeline). Research using administrative records to study otherwise inaccessible processes typically does not focus on linking across domains to the same degree as population-level administrative data research, presumably because of the underlying logic of these projects, which focuses primarily on isolating a hard-to-identify set of processes. Further, the unique relationship between data owners (often private companies) and researchers, and the difficulty in linking with public administrative sources (for example, the Census Bureau must avoid doing research that would favor one company over another) make linkage particularly challenging. That said, linking administrative records from these contexts with other administrative records could provide important insights and would appear to be an important frontier for administrative records research. For example, such data could help us understand how graduates from job search and other training programs fare at different stages in the hiring process. To date, we are aware of only one proj-

ect that has linked human resource records with other individual-level data: linking human resource data on the hiring pipeline at a school district with data at the Census Bureau (Brummet and Penner 2017). Among other things, such data linkages provide opportunities for understanding the labor market implications of unsuccessful applications. We suspect that as the importance of evidence-based practices grows—both generally and in the context of securing foundation funding—opportunities for linking data from local organizations with important domain-specific information will continue to increase.

Combining Survey Data with Existing Administrative Records

The narrow specificity of some administrative data files often limits the range of scientific research questions that analyses drawing solely on that file can examine. Although this has the benefit of focusing researchers' attention on the measures salient to practitioners and policymakers, researchers often supplement administrative records with other information. For example, by linking administrative records with surveys measuring constructs of interest, researchers have examined teacher effects on motivation (Ruzek et al. 2015), shown how school climate can mitigate the academic effects of neighborhood violence (Laurito et al. 2019), and demonstrated how a manager's human resource practices moderate the relation between manager gender and gender wage inequality among workers (Abendroth et al. 2017). Future research in this vein linking implicit bias measures with hiring managers' real-world decisions from human resource data would also help us greatly expand our understanding of how organizational context and policies might moderate the effects of these biases. Likewise, researchers who have information on particular individuals often supplement that information with administrative records. A number of studies, for example, have used administrative data to examine the long-term outcomes associated with interventions, linking researchers' information about who received the treatment with administrative records (Chetty et al. 2011; Holbein 2017).

Elsewhere in this double issue, David Grusky,

Michael Hout, Timothy Smeeding, and Matthew Snipp highlight an additional benefit of combining survey and administrative data, noting that a common data infrastructure would allow surveys to be overlaid on top of administrative data and alleviate respondent burden (2019). This would enhance what is possible using either the survey or the administrative data independently.

Qualitative Research with Administrative Data

Although much of the research using administrative data uses quantitative information, administrative records also contain vast amounts of qualitative information. Archival research using administrative records provides a strong indication of the considerable value of qualitative work using administrative records. Although qualitative social science research using contemporary administrative records is also just beginning to realize its potential, several examples evince the promise of such approaches. Recent qualitative research in medicine, for example, highlights gender differences in the feedback that medical school residents receive (Mueller et al. 2017), and research on online dating profiles underscores how racial boundaries are reinforced not just by racial homogeneity, but also by those looking to date across racial lines (Rafalow, Feliciano, and Robnett 2017).

In many administrative contexts, given the scale of textual data, advances in machine coding offer a promising approach to turning rich qualitative data into quantitative data. In this double issue, Emily Penner and her colleagues provide one example of this approach, showcasing how essays submitted as part of teacher applications are correlated with a variety of policy-relevant considerations (2019). The promise of such approaches in researcher-practitioner partnerships is difficult to overstate, because when these organizations begin to leverage their data in the ways that large tech firms do, there would appear to be substantial benefits for both policy and science. With text mining becoming increasingly sophisticated and common, and the growth of software to aid in the transcription, storage, coding and sorting processes in qualitative research, the distinction

between quantitative and qualitative research is one that could quickly fade in administrative data research.

TECHNICAL, LEGAL, ETHICAL, AND PERCEPTUAL CHALLENGES

In this last section, we highlight a few current challenges specific to working with administrative data. Many are extensions of challenges that exist in social science more broadly around balancing the privacy of research participants with making data widely accessible to lower the barriers to conducting research. In this respect, we see parallels between current efforts to democratize access to administrative records (see, for example, Grusky et al. 2019) and the advent of the General Social Survey, which made nationally representative survey data widely available to the scientific community. Prior to the General Social Survey, social scientists collected their own surveys and typically did not provide data access to outside researchers, so that access to survey data was typically restricted to prominent scholars and their students. More recently, calls for greater transparency and reproducibility have underscored the value of open science in experimental fields (see, for example, Ioannidis 2005; Open Science Collaboration 2015). Against this broader backdrop, thinking about what open science looks like in the context of administrative data research is critical.

Aggregation of individual data into group scores provides a partial solution to the challenge of privacy in many contexts. This double issue includes two studies that use aggregated data files. Brittany Murray and colleagues report the positive relation between strong parent-teacher associations and growth in student achievement (2019). Portia Miller, Elizabeth Votruba-Drzal, and Rebekah Coley find that community-level resources explain variation in student achievement (2019).

However, aggregated data cannot answer all questions, and in many cases answering research and policy questions requires individual-level data. To facilitate sharing of individual-level data, it is likely important to establish incentives for administrative data linking efforts so that more scholars contribute to this public good. One challenge here is that, due to

the sensitivity of many administrative datasets, access is highly regulated, and it becomes prohibitively difficult and time consuming to navigate the multiple processes required to obtain access to data across different contexts. There are two broad models for addressing these challenges in international comparative research: the Comparative Organizational Inequality Network, which brings together researchers with access to the relevant data in different contexts around a set of common analyses that each researcher conducts on data from their home country; and the Luxembourg Income Study, which creates a largely harmonized set of data from across countries (currently non-administrative survey data) and allows researchers to submit code to run on datasets from different countries without accessing the original data. Given their different costs and benefits, we suspect that both models have important roles to play in comparative research.

More broadly, the challenges in working with administrative data can be broken down into technical, legal, ethical, and perceptual challenges. We review each in turn.

Technical Challenges

Important technical challenges remain to constructing administrative data infrastructure. For example, address-based matches are difficult to implement in contexts that lack a well-defined address system (see, for example, Wynn, Reyes, and Caldwell 2011). Likewise, for computationally intensive analyses (for example, some social network analyses) it is currently not practical to conduct analyses that make use of the density of information available at the population level. These and other questions notwithstanding, in our estimation the largest challenges to administrative data are not technical per se but instead technical constraints imposed in response to legal or perceptual considerations.

For example, it is not clear that there is a strong rationale for why researchers need to be in Texas to analyze data from the state of Texas (except that it may be easier to arrest a misuser within state), or that data from Georgia should be allowed to be used on projects only with a collaborator from a university located in Georgia. Nevertheless, such arrangements remain

relatively common. Although they are not insurmountable, they do create nontrivial barriers to access and hinder the democratization that researchers generally support in science. That said, given the level of trust required for companies to allow researchers to analyze key intellectual property (Fernandez-Mateo and Fernandez 2016) or for countries to allow outside researchers access to tax data (King et al. 2017), some restrictions to access beyond those governing survey data are warranted. These barriers highlight the point that the most important challenge to successful administrative data scholarship is not the technical nature of data storage or security, but rather, the human and institutional relationships that must be developed and maintained. The relational nature of data access in many cases—such as in long-term researcher-practitioner partnerships—does result in important constraints that are in tension with norms around data-sharing and open science.

One important challenge surrounding administrative data is the lack of consistency regarding which data are collected and how they are collected. Although national surveys typically use standardized measures and best practices for assessing various constructs, information contained in administrative data can be highly variable in terms of coverage and quality. One advantage of working in close partnership with the organizations generating administrative data is that they typically have a deep understanding of how the data are generated and areas where information may be inaccurate or have limited coverage, and can often adjust practices to generate data that are of mutual interest. Working closely with partners on the ground can also help avoid misattributing causal relations. As with most survey data, administrative data sources require that findings be disseminated only as aggregate statistics in order to protect privacy. As far as we are aware, very few cases of researchers infringing on the privacy of individuals using administrative records have been documented. At this point, then, the technical challenges involved in building data infrastructure are largely surmountable, and the larger remaining question is whether political will is strong enough to move forward.

Legal Challenges

Currently, legal constraints affecting administrative data infrastructure focus on balancing the privacy of individuals whose data are contained in the administrative records with the ability of institutions to find answers to their pressing policy questions, which in many cases will enable them to serve better those who are represented in their data. Allowing access to outside researchers working on behalf of the organization can greatly enhance the research capacity of institutions that generate administrative data and provide expertise in areas that might be otherwise difficult to obtain. In this context, analyses of administrative data should address questions of the data owner, presumably in service of either those represented in the data or the broader public. By contrast, scientists argue that science benefits from widespread, democratic access, and that this access can yield new insights that might be broadly beneficial to society, the institutions generating the data, and their stakeholders, even if these benefits might not have been anticipated. Although making administrative data more widely available is likely generally beneficial, it is currently difficult to know how to assess and weigh the benefits from broad access.

Many forms of administrative data are legally protected in ways that limit access. Under the Family Educational Rights and Privacy Act, identifiable educational data in the United States can only be shared with researchers in a limited number of contexts, including cases where the studies will help the schools improve instruction. Similar challenges apply to health information and Health Insurance Portability and Accountability Act regulations. The lack of a well-established administrative data infrastructure means that lawmakers often do not consider the impact of legislation on administrative records. For example, out of concerns regarding administrative records being used for enforcement purposes, California lawmakers sought to enact laws prohibiting data-sharing and initially did not recognize the limitations this would create for researchers and administrative data infrastructure. Presumably a more robust and salient administrative data infrastructure will help in avoiding such issues in the future.

In many ways, legal constraints are a question of political will. On this point, the bipartisan support for administrative data represented by the Murray-Ryan commission and the Foundations for Evidence-Based Policymaking Act is encouraging. One might imagine, for example, that evidence-based policies around education and workforce training programs might benefit from administrative records from schools, even if the resulting study might not help each school improve instruction. Although individual lawmakers may differ on policy priorities, it is encouraging that they agree on the need for better data and analysis to inform them.

Ethical Challenges

Beyond strictly legal questions, there are ethical questions as well. Typically, potential research participants have the choice to opt out of a study. But this is not possible in most research that uses administrative records. Although consideration of informed consent is routine when it comes to whether a participant's data are used in traditional research designs, administrative records research is often considered to be nonhuman subjects research. To be clear, questions around individuals' rights vis-à-vis their data are a feature of administrative data more generally and not particular to research. This is apparent when one considers medical records. In approximately half of the states in the United States, physicians or hospitals own patients' records, and only in New Hampshire do these data belong to patients (in the remaining states data ownership is not clearly defined). It seems unlikely that patients would take issue with research analyzing these records for patterns that might help save their life. Likewise, it seems probable that most people would not object to their records being analyzed for research that might help save the lives of others. Nevertheless, because this research is often not considered to involve human subjects, and these data (outside New Hampshire) do not belong to the individual, it is unclear what rights patients should have to restrict the use of their data in administrative records research.

Historically, the argument has been that the primary potential harm in this research is that

of disclosure, or harm to the individual due to a breach of privacy. Some legal scholars suggest that this individualistic perspective may be problematic. In a high-profile example, the Havasupai sued Arizona State University for using existing blood samples in ways not covered by agreements. In discussing this case, Katherine Drabiak-Syed notes that our current legal system is ill equipped to consider issues beyond an individualistic framework, so that harm to a collective group may not be recognized (2010). These questions are perhaps especially salient in the context of Native Americans, where issues over the right to opt out are laden with colonial legacies of ignoring indigenous perspectives and also raise questions of tribal sovereignty. These concerns are likely heightened where blood (or other physical samples) are involved, where research focuses on historically marginalized populations, or when researchers are partnering with data-collecting organizations. The concerns are perhaps somewhat attenuated when looking at historical data (for example, the Dutch Hunger Winter), but the larger point remains relevant for administrative records research.

More broadly, the issue could be conceptualized as whether individuals should have the right to ensure that their data are not used in systems against their wishes. One might imagine, for example, critics of structural racism not wanting their data to be used by companies that might perpetuate racial differences in homeownership through credit scores. But it is difficult not to be complicit when almost everyone is part of the administrative data ecosystem that creates and reproduces these inequalities. This is a feature of our societal data infrastructure and is not specific to research using this infrastructure. Nonetheless, administrative data researchers should be cognizant of these issues, particularly in contexts like researcher-practitioner partnerships where they might influence the kinds of data collected, and where the research being conducted might be used to justify or rationalize practices that may otherwise be seen as problematic.

Perceptual Challenges

Perceptual challenges relevant to administrative data research can be divided into those

within the academy and those in the public domain. Within the academy, in many social science disciplines there is a bias against work that is viewed as overly applied. The term *evaluation research* for example, is sometimes used pejoratively in contrast with *pure science*, implying that scientific work is somehow contaminated by being useful to society. We argue that whatever the origins of this bias, it is a distinction that has outlived its usefulness, and that supporting human flourishing—both through better understanding the social world in the broadest and most abstract sense, as well as through understanding the implications of the concrete choices that we as a society make—ought to be one of the aims of science. The degree to which these biases are held in any given scientific field varies, suggesting that social science disciplines can learn much from those more engaged in policy. These disciplinary biases are perhaps a space that academics are well positioned to change. Although these norms may be deeply entrenched, they are nonetheless created and maintained by academics, suggesting that we as a community can change them by changing our hiring criteria, tenure and promotion letters, award nominations, and graduate training. We suspect that these perceptual challenges within the academy are decreasing, in part because administrative data allow researchers to address questions that are not only important for real-world applications, but also make fundamental contributions to discipline-specific and transdisciplinary research goals. We believe the proliferation of administrative data research suggests that over the long term, perceptual issues within the academy are likely to become less pronounced.

Beyond the academy, public skepticism about the limits of confidentiality and data protection threaten public support for the use of administrative data. Recent hacking events and misuse of large private data files at Facebook and Cambridge Analytica have shaken public faith in keepers of supposedly private data. The threat goes beyond misuse to include possible political obstruction by groups such as ALEC (American Legislative Exchange Council), which has taken the position that all governmental action should be minimal. The possibilities of misuse by insiders, hacking by out-

siders, and opposition by politics will always be present, but we believe the marginal extra risk imposed by bringing researchers into this circle is very low. Researchers are required to be trained and credentialed in the use of sensitive data files, and universities tend to implement cutting-edge technologies in data security. Because of these threats and the public's vigilance, however, researchers would be wise to understand the treasure that they behold and to be extremely careful in their use of administrative data files.

At the same time that the public is skeptical, bipartisan support is also strong for administrative data science to improve our capacity to maximize the potential of our human resources. Data-sharing can be difficult in contexts marked by suspicion and mistrust, and larger conversations around privacy remain important. Legal protections governing administrative data use thus play central perceptual and scientific roles, as well as being important for ethical reasons (Anderson and Seltzer 2007). We believe that it is incumbent on scientists to help make the case for administrative data research by ensuring that the public benefits from the use of their data. Although in some cases this might mean working closely with policymakers and practitioners generating and using the data, press coverage of novel findings using engaging data visualizations that reach the public more broadly also play an important role in highlighting the utility of these data to the broader public. Wide dissemination of research findings not only helps inform public discourse around important social questions, but also plays an educational role by engaging people's curiosity and helping them understand how the social world works.

CONCLUSION

As a society, we have the data and expertise to address questions that are vital to our communal life, but we currently do not have the infrastructure to bring data from disparate sources together and provide access to researchers with high-impact projects. U.S. administrative data infrastructure has lagged behind that of its peers, leading to policies that are not as well tuned as they might be, and in many cases leading American social scientists to work with bet-

ter data from other countries. Important policy-relevant scientific questions go unanswered, and scholars and policymakers are left to infer how things might work in the United States based on evidence from elsewhere. The lack of data infrastructure has human costs for our students, patients, and their families; has pecuniary costs for taxpayers; and puts American science at a disadvantage. Recent efforts to create administrative data infrastructure have great promise to rectify the situation, making it an exciting time to be an administrative data researcher.

One final word of caution is perhaps in order: in America, the logic of competition drives many of our collective efforts. When building infrastructure, however, coordination is important. To use a metaphor from physical infrastructure, having five sets of highway systems that do not connect with each other is considerably less useful than having a single, well-planned system. We have an opportunity to create world-class data infrastructure that will enable policymakers to make better policies, scientists to understand society better, teachers to instruct students better, and physicians to treat patients better. In moving forward, coordinating efforts to ensure that we build the best data infrastructure possible, and that our data can benefit the public as much as possible, is paramount.

REFERENCES

- Abendroth, Anja-Kristin, Silvia Melzer, Alexandra Kalev, and Donald Tomaskovic-Devey. 2017. "Women at Work: Women's Access to Power and the Gender Earnings Gap." *ILR Review* 70(1): 190-222.
- Abowd, John M., John Haltiwanger, and Julia Lane. 2004. "Integrated Longitudinal Employer-Employee Data for the United States." *American Economic Review* 94(2): 224-29.
- Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. 2016. Do Low Levels of Blood Lead Reduce Children's Future Test Scores? No. w22558. National Bureau of Economic Research.
- Aizer, Anna, and Joseph J. Doyle Jr. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *Quarterly Journal of Economics* 130(2): 759-803.

- Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. "The Long-Run Impact of Cash Transfers to Poor Families." *American Economic Review* 106(4): 935–71.
- Ananat, Elizabeth O., Anna Gassman-Pines, Dania V. Francis, and Christina M. Gibson-Davis. 2017. "Linking Job Loss, Inequality, Mental Health, and Education." *Science* 356(6343): 1127–28.
- Anderson, Margo, and William Seltzer. 2007. "Challenges to the Confidentiality of U.S. Federal Statistics, 1910–1965." *Journal of Official Statistics* 23(1): 1–34.
- Austin, Megan, R. Joseph Waddington, and Mark Berends. 2019. "Voucher Pathways and Student Achievement in Indiana's Choice Scholarship Program." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 20–40. DOI: 10.7758/RSF.2019.5.3.02.
- Bailey, Drew, Greg J. Duncan, Candice L. Odgers, and Winnie Yu. 2017. "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions." *Journal of Research on Educational Effectiveness* 10(1): 7–39.
- Blau, Peter, and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: John Wiley & Sons.
- Brummet, Quentin, and Emily Penner. 2017. "After School: An Examination of the Career Paths and Earnings of Former Teachers." Paper presented at APPAM 2017. Chicago (November 2–4, 2017).
- Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science* 343(6178): 1478–85.
- Candelaria, Christopher A. 2015. "Rethinking Teacher Effects on Student Achievement." Paper presented at the Association for Education Finance and Policy. Washington, D.C. (February 26–28, 2015).
- Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126(4): 1593–660.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104(9): 2593–632.
- . 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104(9): 2633–79.
- Denice, Patrick, and Betheny Gross. 2016. "Choice, Preferences, and Constraints: Evidence from Public School Applications in Denver." *Sociology of Education* 89(4): 300–20.
- Di, Qian, Yan Wang, Antonella Zanobetti, Yun Wang, Petros Koutrakis, Christine Choirat, Francesca Dominici, and Joel D. Schwartz. 2017. "Air Pollution and Mortality in the Medicare Population." *New England Journal of Medicine* 376(26): 2513–22.
- Dodge, Kenneth A. 2009. "Community Intervention and Public Policy in the Prevention of Antisocial Behavior." *Journal of Child Psychology and Psychiatry* 50(1–2): 194–200.
- . 2011. "Context Matters in Child and Family Policy." *Child Development* 82(1): 433–42.
- Dodge, Kenneth A., W. Benjamin Goodman, Robert A. Murphy, Karen O'Donnell, Jeannine Sato, and Susan Guptill. 2014. "Implementation and Randomized Controlled Trial Evaluation of Universal Postnatal Nurse Home Visiting." *American Journal of Public Health* 104(S1): S136–43.
- Domina, Thurston, Andrew McEachin, Paul Hanselman, Priyanka Agarwal, NaYoung Hwang, and Ryan Lewis. 2016. "Beyond Tracking and Detracking: The Dimensions of Organizational Differentiation in Schools." RAND Labor & Population working paper no. WR1155. Santa Monica, Calif.: RAND Corporation.
- Downing, Janelle, and Tim Bruckner. 2019. "Subprime Babies: The Foreclosure Crisis and Initial Health Endowments." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 123–40. DOI: 10.7758/RSF.2019.5.2.07.
- Drahiak-Syed, Katherine. 2010. "Lessons from *Havasupai Tribe v. Arizona State University Board of Regents*: Recognizing Group, Cultural, and Dignity Harms as Legitimate Risks Warranting Integration into Research Practice." *Journal of Health and Biomedical Law* 6(2): 175–226.
- Fabelo, Tony, Michael D. Thompson, Martha Plotkin, Dottie Carmichael, Miner P. Marchbanks, and Eric A. Booth. 2011. "Breaking Schools' Rules: A Statewide Study of How School Discipline Relates to Students' Success and Juvenile Justice Involvement." New York: Council of State Governments Justice Center.

- Fernandez, Roberto M., and Santiago Campero. 2017. "Gender Sorting and the Glass Ceiling in High-Tech Firms." *ILR Review* 70(1): 73–104.
- Fernandez, Roberto M., and Colette Friedrich. 2011. "Gender Sorting at the Application Interface." *Industrial Relations: A Journal of Economy and Society* 50(4): 591–609.
- Fernandez, Roberto M., and Jason Greenberg. 2013. "Race, Network Hiring, and Statistical Discrimination." In *Networks, Work and Inequality*, edited by Steve McDonald. Bingley, UK: Emerald Group Publishing.
- Fernandez, Roberto M., and Brian Rubineau. 2019. "Network Recruitment and the Glass Ceiling: Evidence from Two Firms." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 88–102. DOI: 10.7758/RSF.2019.5.3.05.
- Fernandez-Mateo, Isabel, and Roberto M. Fernandez. 2016. "Bending the Pipeline? Executive Search and Gender Inequality in Hiring for Top Management Jobs." *Management Science* 62(12): 3636–55.
- Finucane, Mariel McKenzie, Ignacio Martinez, and Scott Cody. 2017. "What Works for Whom? A Bayesian Approach to Channeling Big Data Streams for Public Program Evaluation." *American Journal of Evaluation* 39(1): 109–22.
- Ganzeboom, Harry B. G., Donald J. Treiman, and Wout C. Ultee. 1991. "Comparative Intergenerational Stratification Research: Three Generations and Beyond." *Annual Review of Sociology* 17(1): 277–302.
- Godechot, Olivier. 2013. "Financialization and Socio-spatial Divides." *L'Année Sociologique* 63(1): 17–50.
- Goerge, Robert M., and Emily R. Wiegand. 2019. "Understanding Vulnerable Families in Multiple Service Systems." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 86–104. DOI: 10.7758/RSF.2019.5.2.05.
- Grusky, David B., Michael Hout, Timothy M. Smeeding, and C. Matthew Snipp. 2019. "The American Opportunity Study: A New Infrastructure for Monitoring Outcomes, Evaluating Policy, and Advancing Basic Science." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 20–39. DOI: 10.7758/RSF.2019.5.2.02.
- Hällsten, Martin. 2014. "Inequality Across Three and Four Generations in Egalitarian Sweden: 1st and 2nd Cousin Correlations in Socio-Economic Outcomes." *Research in Social Stratification and Mobility* 35(1): 19–33.
- Holbein, John B. 2017. "Childhood Skill Development and Adult Political Participation." *American Political Science Review* 111(3): 572–83.
- Howard, Lanikque, Lisa Klein Vogel, Maria Cancian, and Jennifer L. Noyes. 2019. "Building Connections: Using Integrated Administrative Data to Identify Issues and Solutions Spanning the Child Welfare and Child Support Systems." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 70–85. DOI: 10.7758/RSF.2019.5.2.04.
- Ioannidis, John P. A. 2015. "Why Most Published Research Findings Are False." *PLoS Medicine* 2(8): e124.
- Kessler-Harris, Alice. 1982. *Out to Work: A History of Wage-Earning Women in the United States*. Oxford: Oxford University Press.
- Kim, ChangHwan, and Christopher R. Tamborini. 2019. "Are They Still Worth It? The Long-Run Earnings Benefits of an Associate Degree, Vocational Diploma or Certificate, and Some College." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 64–85. DOI: 10.7758/RSF.2019.5.3.04.
- King, Joseph, Andrew M. Penner, Nina Bandelj, and Aleksandra Kanjuo-Mrčela. 2017. "Market Transformation and the Opportunity Structure for Gender Inequality: A Cohort Analysis Using Linked Employer-Employee Data from Slovenia." *Social Science Research* 67 (September): 14–33.
- Laurito, Agustina, Johanna Lacoë, Amy Ellen Schwartz, Patrick Sharkey, and Ingrid Gould Ellen. 2019. "School Climate and the Impact of Neighborhood Crime on Test Scores." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 141–66. DOI: 10.7758/RSF.2019.5.2.08.
- Li, Danielle. 2012. "Essays on the Organization of Science and Education." Ph.D. diss., Massachusetts Institute of Technology.
- Liebler, Carolyn A., Renuka Bhaskar, and Sonya R. Porter. 2016. "Joining, Leaving, and Staying in the American Indian/Alaska Native Race Category Between 2000 and 2010." *Demography* 53(2): 507–40.
- Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2013. "Long-Term

- Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity." *American Economic Review* 103(3): 226–31.
- Meghir, Costas, Mårten Palme, and Marieke Schnabel. 2012. "The Effect of Education Policy on Crime: An Intergenerational Perspective." NBER working paper no. w18145. Cambridge, Mass.: National Bureau of Economic Research.
- Miller, Portia, Elizabeth Votruba-Drzal, and Rebekah Levine Coley. 2019. "Poverty and Academic Achievement Across the Urban to Rural Landscape: Associations with Community Resources and Stressors." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 106–22. DOI: 10.7758/RSF.2019.5.2.06.
- Mills, Jonathan N., and Patrick J. Wolf. 2017. "Vouchers in the Bayou: The Effects of the Louisiana Scholarship Program on Student Achievement After 2 Years." *Educational Evaluation and Policy Analysis* 39(3): 464–84.
- Mueller, Anna S., Tania M. Jenkins, Melissa Osborne, Arjun Dayal, Daniel M. O'Connor, and Vineet M. Arora. 2017. "Gender Differences in Attending Physicians' Feedback to Residents: A Qualitative Analysis." *Journal of Graduate Medical Education* 9(5): 577–85.
- Murray, Brittany, Thurston Domina, Linda Renzulli, and Rebecca Boylan. 2019. "Civil Society Goes to School: Parent-Teacher Associations and the Equality of Educational Opportunity." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 41–63. DOI: 10.7758/RSF.2019.5.3.03.
- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349(6251): aac4716.
- Park, Hyunjoon. 2014. "The Legacy of Disadvantaged Origins: Blocked Social Mobility of Descendants of Nobis in Late Joseon (Korea)." Ann Arbor: Nam Center for Korean Studies, University of Michigan.
- Penner, Andrew M., Thurston Domina, Emily K. Penner, and AnneMarie Conley. 2015. "Curricular Policy as a Collective Effects Problem: A Distributional Approach." *Social Science Research* 52 (July): 627–41.
- Penner, Emily K., Jane Rochmes, Jing Liu, Sabrina Solanki, and Susanna Loeb. 2019. "Differing Views of Equity: How Prospective Educators Perceive Their Role in Closing Achievement Gaps." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 103–27. DOI: 10.7758/RSF.2019.5.3.06.
- Persico, Claudia, David Figlio, and Jeffrey Roth. 2016. "Inequality Before Birth: The Developmental Consequences of Environmental Toxicants." NBER working paper no. w22263. Cambridge, Mass.: National Bureau of Economic Research.
- Petersen, Trond, and Laurie A. Morgan. 1995. "Separate and Unequal: Occupation-Establishment Sex Segregation and the Gender Wage Gap." *American Journal of Sociology* 101(2): 329–65.
- Petersen, Trond, and Ishak Saporta. 2004. "The Opportunity Structure for Discrimination." *American Journal of Sociology* 109(4): 852–901.
- Poulton, Richie, Avshalom Caspi, Barry J. Milne, W. Murray Thomson, Alan Taylor, Malcolm R. Sears, and Terrie E. Moffitt. 2002. "Association Between Children's Experience of Socioeconomic Disadvantage and Adult Health: A Life-Course Study." *The Lancet* 360(9346): 1640–45.
- Rafalow, Matthew H., Cynthia Feliciano, and Belinda Robnett. 2017. "Racialized Femininity and Masculinity in the Preferences of Online Same-Sex Daters." *Social Currents* 4(4): 306–21.
- Reardon, Sean F. 2019. "Educational Opportunity in Early and Middle Childhood: Using Full Population Administrative Data to Study Variation by Place and Age." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 40–68. DOI: 10.7758/RSF.2019.5.2.03.
- Ruzek, Erik A., Thurston Domina, AnneMarie M. Conley, Greg J. Duncan, and Stuart A. Karabenick. 2015. "Using Value-Added Models to Measure Teacher Effects on Students' Motivation and Achievement." *Journal of Early Adolescence* 35(5–6): 852–82.
- Song, Xi, and Cameron D. Campbell. 2017. "Genealogical Microdata and Their Significance for Social Science." *Annual Review of Sociology* 43(1): 75–99.
- Sun, Min, Emily K. Penner, and Susanna Loeb. 2017. "Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants." *American Educational Research Journal* 54(4): 607–43.
- Temby, Owen F., and Ken R. Smith. 2014. "The Association Between Adult Mortality Risk and Family History of Longevity: The Moderating Effects of Socioeconomic Status." *Journal of Biosocial Science* 46(6): 703–16.
- Tomaskovic-Devey, Donald. 1993. *Gender & Racial*

- Inequality at Work: The Sources and Consequences of Job Segregation.* Ithaca, N.Y.: Cornell University Press.
- U.S. Census Bureau. 1996. "Puerto Rico, Virgin Islands, and the Pacific Island Territories." In *1990 Census of Population and Housing, History*, chapter 13. Washington: Government Printing Office.
- Wightman, Patrick, and Sheldon Danziger. 2014. "Multi-Generational Income Disadvantage and the Educational Attainment of Young Adults." *Research in Social Stratification and Mobility* 35(1): 53–69.
- Wilde, Melissa J. 2004. "How Culture Mattered at Vatican II: Collegiality Trumps Authority in the Council's Social Movement Organizations." *American Sociological Review* 69(4): 576–602.
- Wynn, John D., Daniel A. Reyes, and Willard E. Caldwell. 2011. "United States Census 2010: A Description of the 2010 Census Operations and Data Products of the Island Areas, and How They Compare to the 50 States and the District of Columbia." Washington: U.S. Census Bureau.
- Young, Cristobal, Charles Varner, Ithai Z. Lurie, and Richard Prisinzano. 2016. "Millionaire Migration and Taxation of the Elite: Evidence from Administrative Data." *American Sociological Review* 81(3): 421–46.