From Big Data to Useful Research Evidence: Forging a Path Toward Better Youth Outcomes

Kim DuMont
WILLIAM T. GRANT FOUNDATION

Timothy M. Smeeding
UNIVERSITY OF WISCONSIN-MADISON

JUNE 2016
Using Data to Produce Useful Research Evidence

Access to data doesn't guarantee the production of useful research evidence, nor does it guarantee that evidence will be used. But linking large data sets and facilitating researcher access creates opportunities to answer different sets of questions than those allowed by survey, experimental, or qualitative data alone. Data linkages set the stage for scientific discoveries that can inform smarter policies and programs.

Both researchers and policymakers have been calling for access to big data—large volumes of digital information about individuals, their activities, and their interactions with different systems, including government. Others have gone so far as to argue that the “big data revolution,” which facilitates analysis of large-scale data sets drawn from administrative records or linked records from multiple sources, will form the future of evidence based policymaking (Decker, 2014).

Thanks to efforts underway at the federal, state, and local levels to integrate and allow access to a variety of data sets, some researchers are already tracking individuals over long periods of time and across multiple systems and institutions, yielding findings with the potential to challenge assumptions and generate new insights. For example, researchers have accessed, linked, and analyzed state education and federal labor statistics to learn about the effects of state-wide economic downturns on school achievement (Ananat, Gassman-Pines, Francis, and Gibson-Davis, 2011), and others have used state-level accountability and test score data to better understand shifts and correlates of racial and ethnic achievement gaps (Reardon, Valenzino, Kalogrides, Shores, and Greenberg, 2013).

Others have used big data for research to advance an understanding of how and when to respond to problems (Duncan, Magnuson, and Votruba-Drzal, 2014). At the William T. Grant Foundation, we are convinced that social science research, including that which taps into the potential of big data, can play an important role in addressing the challenge of inequality. We think that both the extent of inequality and its effects on youth outcomes are not inevitable, but amenable to social policy. And we believe that high-quality research can help identify and build understanding of approaches that help reduce short and long term inequalities.

Access to administrative data provides researchers with a unique opportunity to generate research that informs these efforts. Some have argued that this access will also facilitate faster and lower-cost evaluations of federal, state, and local programs and provide better estimates of program costs. But to fully maximize the value of data, we need to ensure that the research evidence produced from analyses of big data is useful and used. We think this requires:

- linking administrative data to other sources of information,
- involving decision makers in the process from the outset, and
- establishing strategies and structures to initiate and sustain such efforts.

These ideas are consistent with the Foundation’s focus on improving the use of research evidence. We are interested in investigations that identify and test how to connect decision makers with research evidence—including research resulting from thoughtful uses of administrative data, as well as studies that identify and test the incentive structures that encourage the production of research that answers decision makers’ most pressing questions. Ultimately, we hope such efforts will foster a culture of evidence wherein what is learned about programs and policies to reduce inequality is moved into action.

There is growing enthusiasm among social scientists about using “big data” to conduct research and develop bodies of knowledge. Many think this work has the potential to inform policies and programs to expand opportunities and reduce inequalities among young people. But what steps can we take to ensure that access to big data leads to the production of high-quality, useful research evidence? And what else do we need to know to ensure that this evidence is ultimately used by decision makers in ways that benefit youth? These questions are central to our foundation’s interest in improving the use of research evidence.
Linking different data sources

The potential of big data would be stronger if the data were integrated with other sources of information. Many also argue that more researchers need access to these combined data. These are central aims of a new project underway at the National Research Council known as the American Opportunity Study (AOS). AOS would create a structure that allows all qualified researchers—instead of only a select group—regular access to public record data in a safe, protected environment (Mervis, 2014).

Data that comprise the AOS are extensive, and would offer new opportunities to study social mobility (see . The project aims to develop a comprehensive link between survey and evaluation studies and administrative data from the Census and agencies such as the Social Security Administration. The AOS panel would provide repeated observations on individual income, education, occupation, and other demographic variables for individuals (Grusky, Smeeding, and Snipp, 2015). Parent–child or intergenerational matches are also possible with access to data from the IRS, the American Community Survey, and the decennial census.

The structure of AOS is also being designed to inform program and policy evaluations aimed at contributing knowledge about upward mobility for disadvantaged youth. Figure 2 presents the three tiers of the broad plan (Grusky, Smeeding, and Snipp, 2016). The top layer shows data links across decennial Censuses, a long-term, ten year panel of everyone who completes a Census form; the middle layer includes surveys, studies, or evaluations of social programs; and the bottom layer represents linked administrative datasets.

If successfully developed, AOS would provide a resource of unparalleled statistical power and an opportunity for large-scale causal research (Grusky et al., 2015; 2016). One can easily imagine how survey or evaluation supplements to AOS might reveal the contours and correlates of social mobility as well as potential responses. One might investigate intergenerational issues and look at long-run outcomes of early life circumstances (Mitnik, Bryant, Grusky, and Weber, 2015; Stinson and Wignal, 2014). Researchers could also link state data on educational experiences to national sources of earnings and income data (like that in the IRS sample) for both parents and children, and look forward and backward to see how they fare on a range of outcomes (Grusky et al., 2015; 2016).

“If an AOS of this sort were assembled, it would open up new fields of social science inquiry; increase opportunities for evidence-based policy on poverty, mobility, child development, and labor markets; and otherwise constitute a new social science resource with much reach and impact.” (Grusky et al., 2015, p. 63).

AOS might also be leveraged to answer questions about long term outcomes of policies. This would build on prior work that used administrative data to evaluate the educational and economic payoffs of federal social programs for cash transfers and refundable tax credits (Aizer, Eli, Ferrie, and Lleras-Muney, 2014; Chetty, Friedman, and Rockoff, 2011; Dahl, Gordon, and Lochner, 2012) and examined the unintended consequences of policies designed to support families (Cancian, Heinrich, and Chung, 2013). With AOS, a research team might expand on these findings and use new sources of administrative data or survey work to examine areas of research that are currently lacking, such as the effects of policies on behavior and mental health (Duncan et al., 2014).
For example, a team might use linkages to access IRS data and incarceration statistics from the Bureau of Justice Statistics to learn about the effects of youth incarceration on longer term outcomes, such as earnings post release and recidivism. Another team might examine Veteran’s Administrative data to assess the effects of military service on health and economic outcomes later in life. Alternatively a team might use linked survey data from the Panel Study of Income Dynamics or the National Longitudinal Study of Adolescent to Adult Health to probe potential mechanisms or competing hypotheses about why the observed relationships exist. This range of findings could encourage further study on implementation and context, and, in turn, inform smarter policies and programs.

Ultimately, looking across data sets may build comprehensive evidence about how young people fare in the systems through which they live and grow, and provide new understanding about the intersections of these systems. Andrews, Imberman, and Lovenheim (2016), for example, have linked administrative records about K-12 education, postsecondary education, and earnings to examine the impact of two programs in one state for students from low-income high schools. The programs provided additional financial aid and enhanced supports for students once enrolled at one of two highly regarded public universities in Texas. The team’s study demonstrated the potential of such programs to yield long-term earning benefits, and highlighted how differences in the design of these programs may have long-term implications (Andrews et al., 2016).

While these lessons are compelling, questions remain, however, about what it takes to move from data to “a new social science resource with much reach and impact” (Grusky et al., 2015; 2016). Unless we use this data to produce research evidence that addresses the needs of decision makers, the promise of the moment may go unrealized. We suspect this requires knowing how to foreground decision makers’ information needs at the front end of the process and employing strategies and structures to improve the likelihood that the research evidence produced is ultimately used.
Moving from Data to Research to Policy: What Does it Take?

The evidence base for policymaking needs to offer the best data it can to make headway in reducing inequality and expanding opportunity for youth. One effort to help institutionalize and regularize researcher access to public administrative records is the newly established Commission on Evidence-based Policymaking. Federal legislation broadly charges the Commission to recommend ways to expand access to and use of government data and address “how data and results can be used to inform program administrators and policymakers” (H.R. 1831, 2016). This will likely involve discussions about technical challenges in linking disparate data, concerns about privacy, rules for accessing the data, and managing and documenting the content of data sets (Johnson, Massey, and O’Hara, 2015; Warren, 2015; Yiu, 2012).

These considerations are essential, but we know from Foundation-supported work that there is also a social side to research use. This includes consideration of how decision makers will use the resulting research and the relationships, incentives, and supportive structures it takes to bring these considerations into the research process. Thus, while promising efforts are underway to facilitate access for researchers, we need to do more to ensure that what is learned from these data is useful—that research is not just produced and pushed out for consumption without careful consideration of how and by whom the resulting research will be used.

But connecting decision makers’ interests with researchers’ pursuits is a significant challenge. Work from those studying the use of research evidence offers some promising paths for establishing important safeguards.

**Working with decision makers**

The move from data to use-ready research must consider how the evidence generated might be used. Input from decision makers on the types of questions they want answered could inform the types of data sets that need to be connected, the level of analysis (e.g., city, district, school, or student), and the identification of salient variables. For example, state education leaders may want to know how specific investments at a school level—say, technology, textbooks, and physical improvements—affect achievement. This list may reflect district leaders’ working hypotheses about competing expenditures and what matters for student success. Not all of these variables may be on the minds of researchers, however. The construction of data sets, then, begins the path to generating research evidence—and this beginning delimits the range of questions researchers are able to answer with big data.

Conversely, decision makers may be experts on their systems, but they may not have the capacity to anticipate the analytic structure of data sets, design research studies, or critically evaluate, prioritize, and interpret research evidence. But these skills can be nurtured through partnerships or other opportunities that allow users to engage with research evidence and make sense of how it might apply to their organizations (Honig, Venkateswaran, Twicthell, 2014; McDonald and Weatherford, 2014; Palinkas et al., 2014).

Research-practice partnerships and similar collaborations offer a sound strategy for involving decision makers (Coburn, Peneul, and Geil, 2012; Palinkas, Short, and Wong, 2015). Reciprocal participation in the social process of building and using evidence strengthens the joint effort between researchers and decision makers. In partnerships, decision makers are provided with opportunities to articulate their needs, and this can improve the likelihood that the resulting research evidence will be useful to them (Coburn, et al., 2012; Palinkas et al., 2015). At the same time, frequent exchanges help researchers build familiarity with the needs and motivations of decision makers. This, in turn, can help researchers construct a process and communicate findings in ways that are accessible and oriented toward use (AcademyHealth, 2014; Nisbet, 2015).
The literature has strong examples of the value of partnerships when working with administrative data, especially at state and local levels. The University of Chicago Consortium on School Research, a long-standing partnership between researchers at the University of Chicago and within the Chicago Public School district, for instance, has used administrative data to identify indicators of success, chart improvement, and conduct theory-driven evaluations of within-district programs and policies (Roderick, Easton, and Sebring, 2009). The Child and Adolescent Services Research Center at Rady Children’s Hospital in San Diego has also demonstrated the value of marrying research expertise and experience with national datasets with local knowledge of system operations and information needs. By embedding researchers within the local organization, they have linked local administrative systems, data from nationally representative longitudinal survey samples, and Medicaid data to show that court referrals were associated with racial disparities in the use of mental health services (Landsverk, Garland, Reutz, and Davis, 2010).

Thus, partnerships between decision makers and researchers, either within their organizations or at research institutions, may produce work that, from the outset, is likely to penetrate the realms of policy and practice.

**Incentives to facilitate collaboration**

Incentivizing collaboration and knowledge integration is fundamental to fostering broader uses of research in ways that benefit youth. For example, the recent reauthorization of the federal Every Student Succeeds Act includes provisions that call for local decisions about programs and activities to be “evidence-based.” Further, elements of the law link funding to these levels of evidence. This is just one example of how policy can encourage integration of research evidence in decision making. Creating these and other kinds of incentives may strengthen collaborations by structuring expectations and routines that promote use from beginning to end—from data set construction to evidence production to research use.

Beyond legislation, though, what structures might incentivize links between research institutions or trusted intermediaries and decision makers? To some extent, demand for program evaluations may encourage researchers and decision makers to work together. Still, more direct incentives may be needed to upend existing tendencies to focus on the demands and rewards of one’s own institution (Casselman, 2015; Ferber, SSIR, 2014). This might include providing direct support for sustained research collaborations. On the research side, incentives might include course releases to allow time for partnership activity and recognition that partnering constitutes valuable service to the community. For decision makers, partnering may be more desirable if state and local agencies had expedient mechanisms to receive funds and award contracts, or if they could reallocate resources from programs deemed ineffective to those with a stronger theoretical or empirical evidence base.

**Other strategies and supports for improving the use of research evidence**

Another promising avenue to improve use includes interventions that are designed to develop skills related to accessing and appraising research (Langer, Tripney, and Gough, 2016). Since partnerships and other one-on-one collaborations are often not possible, technical assistance is an important alternative for cultivating such skills. Technical assistance providers often have deep knowledge of research and the decision makers’ system, as well as sustained contact with decision makers. It is important to imagine how this supportive infrastructure might be refined to develop both leaders’ capacity to access and process research evidence and opportunities to discuss how research might be integrated with existing evidence. In turn, technical assistance providers may also emerge as important conduits for sharing their knowledge with researchers or intermediaries engaged in analyzing data sets. This type of give and take can lead to new ways of seeing old problems and help inform action, thereby contributing to a culture in which it is normative for leaders to use research evidence in decision making processes (Farrell and Coburn, 2016).

Lastly, because not all research will be interpreted with a researcher on hand, there need to be structures in place to make sense of the abundance of information that amasses from big data projects like the American Opportunity Study. Currently, numerous clearinghouses provide access to research evidence, but their structures are far from uniform. These outlets differ in terms of the level
of rigor required for inclusion, elements reported about each study, supporting documentation, and synthesis across studies. What’s more, the kinds of information decision makers need—like user reviews from peers, implementation context, and infrastructure, training, and cost requirements—are often omitted. This variation, unfortunately, makes it difficult for decision makers to identify research evidence that they can use.

But there are promising examples of approaches that respond to these challenges. In the United Kingdom, for instance, systems (and funding) are in place to routinely conduct systematic reviews of emerging evidence bases, and centralized and developed frameworks ease the organization and sharing of research evidence. The London-based Education Endowment Foundation, for instance, rigorously evaluates strategies to improve the use of research evidence and is offering roadmaps for developing supportive structures for using research.

**Moving between data, research evidence, and use: What does it look like?**

Beyond generalities about building bridges and translating research to practice or policy, how do we create the real-world conditions for research use? What will the infrastructure actually look like? What strategies and incentives might it comprise?

One example of how administrative data can be leveraged to generate research evidence that informs decision making is CalYouth, a collaboration between researchers at the University of Chicago and members of the California Child-Welfare Co-Investment Partnership, including leaders from the California Department of Social Services, Child Welfare Directors Association of California, the Judicial Council of California and a group of philanthropic funders (Courtney, Charles, Okpych, Napolitano, and Halsted, 2014). Together, the collaboration is responding to a mandate from the state legislature to evaluate the implementation of extended foster care.

Federal legislation gives states the option to extend services and receive reimbursement for youth in care until age 21, but policymakers in California and elsewhere want to know how and under what conditions the extended care is benefiting youth, and at what cost. Financial resources from a number of private funders have allowed CalYouth to link administrative data on youth foster care histories, employment insurance wage claims, use of public assistance programs, Medicaid, college engagement, and arrests. A sampling frame was generated from the administrative data, and led to extensive interviewing and web-based surveys. State and county administrators and supervisors, youth in foster care, and others provided input on the tools. And, over time, collaborators have used the assembled data to generate research evidence that describes youth and the services they receive (Courtney et al., 2014), tackle concrete problems related to case-load sizes (Courtney et al., 2016), and respond in real time to policy debates about the value of extended care (Courtney and Okpych, 2015). Thus, both the mandate to evaluate the implementation of extended services for youth and an influx of financial resources have helped to incentivize the production of research.

This example shows data being leveraged with an orientation toward use, but it also gets to a larger point about the use of research evidence. As we’ve said, linking and analyzing data sets can contribute to strong, useful bodies of evidence, as long as we’re looking to the data with purpose and remaining mindful that big data is, by itself, only a means to an end. But, as illustrated by the example, in order to increase the likelihood that the evidence we produce is used to inform policies, practices, and programs that benefit youth, we need an infrastructure that encourages opportunities for stakeholders and researchers to engage one another and benefit from each other’s perspectives. We need a framework that cultivates the use of research evidence by incentivizing collaboration and establishing structures that focus attention and resources on integrating the perspectives of decision makers into the research production process.

Work by Foundation-supported researchers and others suggest that merging research evidence in these processes requires having structured discussion about research (Honig et al., 2014; McDonnell and Weatherford, 2014). In the CalYouth example, research evidence—cost-benefit studies in particular—was included in deliberations about whether to extend care (Mosley and Courtney, 2012).
And negotiations about what administrative data to link and what items were needed in the survey included input from decision makers and other stakeholders (Courtney et al., 2014). In this way, the data extraction and subsequent data collection efforts were designed for use from the outset (Bogenschneider, Little, and Johnson, 2013); the team solicited the stakeholders’ feedback, down to specific survey items and data fields.

Lastly, the CalYouth example also highlights some outstanding questions. Providing incentives for cross-sector work does not guarantee a productive collaboration. Although there was some evidence that the CalYouth cross-sector partners collaborated in critical deliberations, it is unclear whether these practices were happenstance or intentional and routine. We need to know more about the extent to which practices were codified. For example, were there formal agreements or memorandums of understanding that established guidelines for working with one another? Was there a coordinating body that instructed the process? Examining the extent and value of such formal structures may inform strategies for facilitating research use.

Conclusion

Linking big data sets is a promising first step to producing research evidence that is used by decision makers. But it is not enough by itself. Efforts to link large-scale data sets from diverse sources have the potential to rapidly enhance what can be learned from surveys, experiments, evaluations, and qualitative data about the lives of young people, as well as ways to improve their outcomes. Yet we must be aware that “many factors shape what research is sought, how it is shared, and the ways in which it is evaluated, used, contorted, or dismissed” (DuMont, 2015).

This knowledge demands intentional efforts to orchestrate and incentivize the move from big data to research evidence to use (Langer et al., 2016). And, we have accumulated promising lessons about what is required to design and promote research use (DuMont, 2015). For instance, research is more likely to inform decision making when it comes from a trusted source and is deeply understood. Participation in the research process and sustained relationships may enhance trust. Structured opportunities to discuss and push back on research appear to deepen knowledge. And these processes are hypothesized to increase appropriate uses of research evidence. When it comes to harnessing the potential of big data, these lessons should not sit idle.

At the same time, we need to know how to create these conditions; here, the knowledge base is relatively weak. Partnerships between researchers or intermediary organizations and decision makers present one potentially promising vehicle. But we need to create and test these and other mechanisms for improving the use of research evidence. Toward that end, our Foundation welcomes studies that identify, build, and test new strategies for improving the use of research evidence. These proposals should describe the body of evidence that is ripe for use; how use is being conceptualized, operationalized, and measured; and why use is expected to improve decision making and, ultimately, youth outcomes. These questions are a key line of inquiry for our Foundation. We suspect these pursuits will lead to new insights about the infrastructure needed to produce and use research in ways that benefit youth.

Throughout the essay, we have offered hypotheses and questions of interests related to our research interest in improving the use of research evidence. Here are few of the ones mentioned:

What structures incentivize links between research institutions or trusted intermediaries and decision makers?

What relationships and supports are needed for researchers to incorporate decision makers’ needs into the research process? Do these strategies and conditions produce research that is used?

Under what conditions do interventions that are designed to develop skills to access and appraise research improve the use of research evidence?

When practices for conducting joint work are codified and formal agreements or memorandums of understanding exist, do these guidelines for working with one another improve the usefulness of the research produced?

Under what conditions do partnerships between researchers or intermediary organizations and decision makers improve the use of research evidence?
References


REPRINT FROM BIG DATA TO USEFUL RESEARCH EVIDENCE: FORGING A PATH TOWARD BETTER YOUTH OUTCOMES