

The American Opportunity Study: A New Infrastructure for Monitoring Outcomes, Evaluating Policy, and Advancing Basic Science



DAVID B. GRUSKY, MICHAEL HOUT, TIMOTHY M. SMEEDING,
AND C. MATTHEW SNIPP

The American Opportunity Study is an ongoing initiative to build the country's capacity to access and analyze linked administrative data. It is best viewed as a population-level scaffolding on which other administrative data can then be hung. This scaffolding, if used as a stand-alone resource, will allow for long-run analyses of fundamental population and labor market processes. If combined with data from other sources, it will allow for long-run program evaluation and other experimental and quasi-experimental analyses. We discuss the current status of the American Opportunity Study, its potential to advance the field, remaining obstacles that must be overcome to build it, and how it can work within the guidelines suggested by the Commission on Evidence-Based Policymaking.

Keywords: administrative data, data linkage, social mobility, program evaluation

The administrative data revolution is in full blossom (Reamer and Lane 2018). Until recently, this revolution was pitched largely in terms of its promise and represented as our future (see, for example, Decker 2014). But that future has clearly arrived.

A broad consensus about the value of admin-

istrative data has been reached among policymakers, elected officials, data administrators, and researchers at the federal and state levels. This consensus arose when federal statistical agencies began sharing special-purpose linked administrative files with researchers in universities, think tanks, and program-evaluation

David B. Grusky is Barbara Kimball Browning Professor in the School of Humanities and Sciences at Stanford University. **Michael Hout** is professor of sociology at New York University. **Timothy M. Smeeding** is Lee Rainwater Distinguished Professor of Public Affairs and Economics at the La Follette School of Public Affairs, University of Wisconsin–Madison. **C. Matthew Snipp** is Burnet C. and Mildred Finley Wohlford Professor of Humanities and Sciences in the Department of Sociology at Stanford University.

© 2019 Russell Sage Foundation. 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. The authors thank all of those who have inspired and advanced the ideas, concepts, and moving parts of the American Opportunity Study. We also thank the reviewers for their careful reading and excellent comments, David Chancellor for graphing assistance, and the Carnegie Corporation, the National Research Council, and Stanford University for their support of the AOS. The authors alone assume responsibility for all errors of omission and commission. Direct correspondence to: Timothy M. Smeeding at smeeding@wisc.edu, Department of Economics, 3464 Social Science Building, 1180 Observatory Dr., University of Wisconsin, Madison, WI 53706.

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.

companies (Kille 2015). These partnerships demonstrated the research value of data that governments collected for administrative purposes. Likewise, former President Obama and his Council of Economic Advisers stressed the need for better access to public administrative data to examine program effectiveness (see Council of Economic Advisers 2015). Meanwhile, and relatedly, Congress passed the Evidence-Based Policymaking Commission Act of 2016 (P.L. 114–140) that created the Commission on Evidence-Based Policymaking (CEP). The commission’s final report advocated for a “national secure data program” that would facilitate access to administrative data (CEP 2017). Within the policy world, many of the most important findings on the effects of poverty, inequality, education policy, and social programs now rest on public administrative data.

It is useful to consider the next steps that should be taken in this “new era of administrative data and evidence-based policy” (Haskins and Margolis 2014, 238). In any era of innovation, the early days in retrospect look chaotic. That applies here as well. Access to administrative data has frequently depended on personal networks and relationships with personnel in federal or state agencies. Datasets were built for one-time analyses. Documentation served agency insiders but was inadequate for new users. Given this state of affairs, some routinization is essential but, to be productive, we need to take a long view. The nation needs to develop—as swiftly as possible—an administrative data infrastructure that guarantees high-quality linkages, common standards for documentation, security and confidentiality, and fair access to researchers with good ideas and the requisite skills.

How might these goals be realized? The first step entails building an on-demand administrative database of the sort proposed by the Commission on Evidence-Based Policymaking in the form of a National Secure Data Service (NSDS) (CEP 2017). The second step is to set up institutions for ensuring both full access for

qualified users and the confidentiality of the data for the public whose data are encoded.

The American Opportunity Study (AOS), which is being developed in collaboration with the Census Bureau, can assist in realizing these goals and thus achieving better access with confidentiality. The AOS is an ongoing effort to link the censuses of 1960 through 2010 and the American Community Surveys (ACS) and thereby convert cross-sectional decennial census data into a bona fide panel that will represent the full U.S. population over the last seventy years. Because this panel will be continuously refreshed as additional census and ACS data become available, it can serve as a population-level scaffolding on which other administrative data (such as tax records, earnings reports, program data) are then hung. The National Research Council (NRC) established the Standing Committee on Creating the American Opportunity Study and charged it with facilitating the digitizing of census data, examining matching and record linkage methodology, building a user committee for the AOS (NRC 2016), and examining governance options based on the 2017 CEP report.¹

The AOS capitalizes on the importance of long-run analyses for both evaluating programs and monitoring trends. It promises not only to upgrade the country’s capacity to study economic and social mobility but also to assist with a broader range of analyses oriented to monitoring long-term trends in poverty, inequality, and labor market outcomes and assessing the long-term effects of policy treatments and experiments. The AOS initiative is thus part of the broader goal to democratize access to administrative data, make that access safe and secure, and thereby realize the potential of linked administrative data (Mervis 2014). In other countries that have linked data, such as Wales and New Zealand, a well-developed infrastructure allows access to carefully vetted scholars, with the result that high-quality evidence is more frequently brought to bear on policy decisions. The payoff to developing this capacity is clear. In collaboration with the Cen-

1. See “Standing Committee on Creating the American Opportunity Study: First Phase,” National Academies of Sciences, Engineering, and Medicine, http://sites.nationalacademies.org/DBASSE/CNSTAT/DBASSE_172151 (accessed October 5, 2018). The authors are members of the committee.

sus Bureau, we have thus begun to build the AOS and close the gap between the United States and other countries in this capacity for authentic evidence-informed policy. The cost of failing to do so is substantial and puts the United States at a competitive disadvantage relative to other countries.

Although many other countries are well ahead of the United States in developing administrative data resources, evidence of its payoff within the United States is already ample. This payoff takes the form of administrative data research on policy and program impacts on health, longevity, well-being, poverty, and socioeconomic mobility (Chetty et al. 2017; Figlio et al. 2014; Almond, Hoynes, and Schanzenbach 2011; Almond, Currie, and Duque 2018; Chetty, Hendren, and Katz 2015; see also CEP 2017). By virtue of these analyses, we now know that absolute economic mobility has fallen steadily since the 1940 birth cohort; that government efforts to support low-income and at-risk families yield health, educational, and economic benefits far in excess of their costs; that public investments in preschool, income support, housing, health care, and nutrition bring about substantial gains for children; and that long-term exposure to poor neighborhoods and polluted areas have lasting effects on social outcomes, health, and life expectancy. We review these and other studies to make it clear that the country's capacity to deliver authentic evidence-based policy will depend in no small part on successfully institutionalizing the administrative data revolution. Throughout this review, we focus on those types of analyses that, by virtue of requiring long-run assessments, reveal the payoff to building the AOS.

This payoff has been substantial because linking administrative data solves a host of problems that have long plagued conventional survey-based analyses of long-run processes. If the AOS is built, it can deliver enough cases to examine long-run effects on smaller populations; provide the power needed to carry out nonparametric analyses; enable intergenerational linkages that allow us to examine social mobility, intergenerational transfers, and sib-

ling effects; provide new opportunities for quasi-experimental analysis over the long run (via, for example, state and local variability in the timing of program delivery); combine sources that allow for more comprehensive studies of program use and labor market processes; and both reduce and better understand the attrition that has long been the bane of survey-based panels.

In this article, we review these potential payoffs in more detail, focusing on what needs to happen to ensure that they are fully realized. The theme throughout is that the AOS should attract widespread support. It should appeal at once to those who believe that better data will demonstrate that existing programs and policies are typically effective and those who are skeptical and expect that many existing programs will be shown to be inefficient. The AOS is, in short, a critical vehicle for shifting the terms of debate about social programs into a straightforward discussion of the evidence. Although we are not so naive as to believe that this transformation will be absolute or uncomplicated, the AOS will bring us closer to realizing the vision of authentic evidence-based policymaking.

THE PAYOFF TO ANALYZING PROGRAMS AND POLICIES WITH BIG PUBLIC DATA

The administrative data revolution is not a recent development. Most notably, public administrative data has been the basis of social science research for many decades in Scandinavia, where data registers provide family records for economic and social outcomes for the entire population of a country. These data, along with more recently available administrative data in the United States, have improved causal analyses within economics and, to a lesser extent, other social sciences (Einav and Levin 2014). The turn to administrative data has accelerated over the last three decades: only 20 percent of microdata-based articles in the “top four” economics journals used administrative data in 1980, whereas 60 percent did so by 2010 (Chetty 2012).² This explosion in administrative data analysis is also clearly in play in the major social

2. The ACS, an extension of the decennial census, counts as administrative in this reckoning, despite including the word *survey* in its title.

science policy journals, such as the *American Economic Journal*, *Economic Policy*, and *Journal of Policy Analysis and Management*.

Although U.S. administrative data are becoming increasingly available, this revolution has relied disproportionately on non-U.S. data, especially from Scandinavia. The research based on Scandinavian population registries has straightforwardly informed policy in the countries from which the data are drawn. It is less clear that we can safely generalize these results to other countries. This reluctance to base policy or social science around results from countries with very different institutions, populations, and cultures is one of the driving forces behind the move to develop Scandinavian-inspired data resources in the United Kingdom (Yiu 2012), Australia (Tam and Clarke 2015), Canada (Trépanier, Pignal, and Royce 2014), and many other countries.

Because the United States is arguably more decentralized than any of these countries, it faces special challenges in harmonizing data across jurisdictions, but also has the potential advantage of leveraging policy variations across those jurisdictions to develop quasi-experimental evidence. It also already has a relatively long and distinguished history of administrative data analysis within many states. For example, Wisconsin state agencies have partnered with researchers to study child support records to evaluate the effectiveness of programs, an arrangement that has become an important model for many other states (see Cancian, Heinrich, and Chung 2013). This partnership involved mutual learning, trust, and understanding, and ultimately led to the development of a larger linked database, the multi-sample person file. The Wisconsin multisample person file links administrative records on more than twenty programs to individual and family beneficiaries (Noyes 2015). Because it is one of the earlier state efforts, it now allows for relatively long-run analyses, though it does not of course have the reach that the AOS promises.

This is but one example of state-level partnerships. In many other states, researchers have effectively exploited administrative data, most notably in Florida, Tennessee, North Carolina, Washington, and California. These data have allowed researchers to examine the long-

term effects of birth conditions (Figlio et al. 2014), government preschool programs (Chetty et al. 2011), community college programs (Stevens, Kurlaender, and Grosz 2015), mandatory college preparation curricula in high schools (Jacob et al. 2016), prison release programs (Harding, Siegel, and Morenoff 2017; Lee, Harding, and Morenoff 2017), and even economic downturns (Ananat et al. 2011; Ananat, Gassman-Pines, and Gibson-Davis 2013). But state-level administrative data have limits because of interstate geographic mobility and because some of the key variables of interest are unavailable with state data. The AOS, by contrast, will allow us to overcome the problem of geographic mobility, provide additional variables of interest, and allow for longer-run analyses of social and economic program effects.

In the following section, we review three signal achievements coming out of these analyses of state and federal administrative data. In doing so, our intent is not just to illustrate the potential of linked administrative data, but also to focus on how the AOS, in particular, could advance the administrative data revolution.

Poverty and the Safety Net

The payoff to administrative data has been especially obvious in the fields of poverty measurement and antipoverty program evaluation. For example, Wisconsin has combined state-level administrative data with the ACS to build a Supplemental Poverty Measure (SPM), thus improving on traditional poverty measures that do not take program income and benefits into account (NRC 2005). With the SPM, Wisconsin policymakers can monitor trends in poverty, assess the effects of social programs on poverty without underreporting key program receipt and benefit levels, and gauge how possible changes in policy will affect poverty (Smeeding and Thornton 2017). The California Poverty Measure (CPM), which is likewise based on administrative data, has been used in similar ways (Wimer et al. 2013, 2014). The California government used the CPM, for example, to understand the likely costs and benefits of a state earned income tax credit supplement before it was enacted (Wimer et al. 2016). This measure will soon be upgraded by combining California Franchise Tax Board data and program data.

Similar administrative data initiatives are under way in New York City (Office of the Mayor 2014) and several other states, such as Oregon. In a related effort, Bruce Meyer, Wallace Mok, and James Sullivan have also assessed antipoverty program effects across many states using program data, leading to the key finding that income support benefits are underreported in survey data (2015).

The AOS, when it becomes available, will make it possible for other states to more easily calculate SPM-style measures. By providing a population-level panel, it will allow states to calculate poverty rates at more detailed geographic levels (by supplementing the ACS with population earnings and tax data), at more frequent intervals (by exploiting tax data, earnings reports, and other frequently released data), and with direct evidence on program use (via administrative program data).

Early Childhood Interventions

A decade ago, James Heckman argued that programs and policies that entail intervening prenatally or in early childhood show an especially attractive return on investment (Cunha and Heckman 2008; Heckman and Mosso 2014; see also Barker 1995; Council of Economic Advisers 2015). Although some of the evidence on behalf of early intervention is survey based, much of it has also turned on administrative analysis (Almond and Currie 2011; Cascio and Schanzenbach 2013; Aizer and Currie 2014).

Administrative data have been important, for example, in reassessing the claim that early childhood education programs may initially stimulate learning but that, over time, the benefit of this early participation tends to diminish. Although the Head Start Evaluation suggested, for example, that learning gains from that program faded by the third grade, analyses of administrative data revealed other compensating long-run benefits (Duncan and Magnuson 2013). When high-quality administrative data were used to reexamine the long-term effects of nutrition interventions, parenting programs, and various high-quality and “moderate-quality” preschool programs (Head Start, for

example), they revealed persisting later-life effects on graduation rates, earnings, and crime (Hoynes, Page, and Stevens 2011; Chetty et al. 2011; Council of Economic Advisers 2015). The AOS, when available, will allow us to build an even richer evidence base on the long-term effects of home visiting programs, childcare and preschool, and early childhood education.

Administrative data have also been important in establishing the long-run effects of tax credits, cash transfers, and near-cash programs (for recent important reviews of this literature, see Shaefer et al. 2018; Almond, Currie, and Duque 2018). We now know, for example, that the earned income tax credit, one of the government’s most important child-poverty programs, reduces the incidence of low birth weight, raises math and reading scores, and boosts college enrollment rates (Dahl and Lochner 2012; Evans and Garthwaite 2014). The Supplemental Nutrition Assistance Program (food stamps) has similarly long-lasting benefits for child recipients as well as positive effects on pregnancy outcomes (Almond, Hoynes, and Schanzenbach 2011) and adult obesity (Hoynes, Schanzenbach, and Almond 2016). Likewise, evidence from the United States and Canada indicates that many types of tax-based refundable cash transfers, such as the Canadian Child Tax Benefit, increase child cognitive achievement and health (Dahl and Lochner 2012; Milligan and Stabile 2009; Evans and Garthwaite 2014). Children who receive Medicaid are more likely to graduate from high school, more likely to complete college, and less likely (at least if they are African American) to die in their late teens or be hospitalized by age twenty-five (Wherry et al. 2015).³

Long-Run Effects of Cash Support

Our third illustrative example pertains to administrative data analyses of direct cash support, for both those who can and cannot work. Much research has been completed on the effects of cash income support and “negative income taxes,” often called basic income, on such short-term outcomes as work effort or child-bearing. But until recently we knew less about

3. Research using administrative data from Medicaid and tax records reveals that the public recoup their Medicaid investment via increased tax revenues (Brown, Kowalski, and Lurie 2015).

their long-term effects.⁴ Using administrative records from the Mothers' Pension program (1911–1935), a precursor to the AFDC program, researchers have now assessed the impact of cash transfers across the entire life course by matching program participants to World War II enlistment records and 1940 census records (Aizer et al. 2014). By using Social Security data to follow program beneficiaries, it was shown that children who receive benefits, even for just a few years, are affected for as long as eighty years or more. Most notably, the poorest children in this sample experienced a 1.5-year increase in longevity by virtue of receiving cash transfers; better-off children saw smaller increases. It was further shown that cash transfers reduced the probability of being underweight by half, increased educational attainment by 0.4 years, and increased income by 14 percent during adulthood (Aizer et al. 2014; Furman 2015). This benefit, which comes mainly from helping low-income families pay for basic needs (such as food, housing, health care), has been shown to have effects on child well-being over and above those of direct service programs, like preschool education and health care (Shaefer et al. 2018; Duncan, Magnuson, and Votruba-Drzal 2014; Furman 2015).

This short review, which is more illustrative than exhaustive, shows that the frontier of research on the effectiveness of social policy has relied on—and will likely continue to rely on—linkages to census and administrative data. This approach has improved the accuracy of our data, reduced the need to field costly surveys, allowed for better monitoring of labor market outcomes, and provided high-quality evidence on the long-term consequences of policies, interventions, and economic and social change.

For all the successes to date, administrative data have yet to be fully exploited because access has been granted idiosyncratically to a few well-connected researchers, and because studies have relied on a small number of administrative data sources and thus been able to address only a limited subset of questions. These

problems can be overcome with the AOS. It will serve as a standing resource that regularizes and expands access to administrative data, that links a more comprehensive constellation of census and administrative data, and that makes a wider range of long-run analyses possible.

A SHORT HISTORY OF THE AMERICAN OPPORTUNITY STUDY

The payoff to building this more comprehensive resource is wide ranging, but in the early history of the AOS most of the protagonists were motivated by a rather narrow interest in monitoring recent trends in social mobility. We briefly review this impetus because of its relevance to how the larger AOS initiative developed.

The initial animating interest in the AOS rested on assessing the long-standing American Dream that hard work and ingenuity will be rewarded with material success even for children born into poor families. The American Dream narrative is deeply embedded within American culture, has attracted generations of immigrants seeking a better life in this nation, and continues to be widely embraced and celebrated among Americans (Manza and Brooks 2016; Mitnik et al. 2015).

It is nonetheless striking that the empirical evidence on recent trends in social mobility is relatively scarce. Although many commentators have openly worried that both relative and absolute mobility are declining across generations, the evidence bearing on these worries is limited. Because the necessary data are unavailable, our evidence on long-term trends in absolute mobility has been pieced together from cross-sectional data and strong assumptions about the trend in relative mobility (see Chetty et al. 2017; also see Hout 2018 for a study of more recent trends in absolute mobility based on the General Social Survey). A handful of studies further suggest a possible decline in relative mobility (Aaronson and Mazumder 2008; Mitnik, Cumberworth, and Grusky 2016). But other studies suggest otherwise (Chetty et al. 2014; Lee and Solon 2009; Hout 2015).

4. The data collected by the various negative income tax studies of the 1960s and 1970s, which mainly examined effects on labor supply, have not been preserved and thus cannot be used to address long-term effects (but see Price and Song 2018).

It is also problematic that most of the current research on mobility is limited to economic (earnings or income) mobility alone. Income offers an important but incomplete view of social mobility. The case for a more complete assessment rests on evidence that occupations and related measures of social class can affect behaviors, attitudes, and political participation in ways not understandable in wholly economic terms (Weeden and Grusky 2005, 2012). It also rests on the understanding that economic mobility is endogenous to the underlying structure of occupational opportunities that give rise to earnings and income. The occupational structure represents this larger organization of opportunities that may be facilitated or limited by various types of social closure that operate partly at the occupation level (such as unions or occupational licenses). Even more important, it is likely that some individuals trade off earnings for other occupational rewards (autonomy, prestige), the implication being that analyses based on economic standing alone may misrepresent the true amount of opportunity. In short, data on occupational mobility are fundamental in themselves and would be important to collect even if it were possible to fully describe income and earnings mobility.

Nearly a half-century has passed since the last multidimensional assessment of social mobility in the United States (Featherman and Hauser 1978; see also Blau and Duncan 1967), a state of affairs that contrasts sharply with practices in other countries (Breen, Mood, and Jonsen 2016). This is a striking lapse considering the profound changes that have taken place in U.S. society over the past four decades. These changes include, for example, the historic increase in women's labor force participation, the decline in manufacturing jobs, the rise of service employment, rising immigration and the associated ethnic diversification of the population, the decline in white men's labor force participation, the ongoing changes in family and household structure, and the increase in economic inequality (Fischer and Hout 2006). How

have these changes affected opportunities within American society? We cannot know until a full multidimensional assessment of mobility is undertaken.

This state of affairs led David Grusky and Matthew Snipp to meet with officials at the Census Bureau, the Office of Management and Budget (OMB), and the National Science Foundation in 2012 to begin a conversation about how to collect the requisite data. These conversations made it clear that a follow-up study of social mobility comparable to the two previous studies (Blau and Duncan 1967; Featherman and Hauser 1978) would be an expensive undertaking. The 1973 study, based largely on a monthly supplement to the Current Population Survey (CPS), cost approximately \$2.0 million, which is nearly \$12.0 million in 2018 dollars.⁵ Worse yet, a new study that does not take the form of a CPS supplement likely would cost many times this amount, possibly exceeding the entire annual budget that the National Science Foundation allocates for sociological research.

With support from the National Science Foundation and the National Research Council, work commenced in 2013 to develop a plan for launching a new study of social mobility.⁶ The core task at that point was to identify a survey vehicle for the study, with the main possibilities being the CPS, the Survey of Income and Program Participation (SIPP), and the ACS. The second task was to identify the most important content domains to be included in this new study. To accomplish these tasks, the group sought the assistance of a wide range of social scientists, mainly sociologists, political scientists, and economists who were experts in social mobility, education, immigration, and demography. These individuals were invited to prepare papers on possible content domains for presentation at a workshop held in June 2013 at the National Academy of Science's Keck Center in Washington, D.C. The resulting papers were subsequently published as a volume in the *Annals of the American Academy of Political and Social Science* (Grusky, Smeeding, and Snipp 2015).

5. Robert M. Hauser, personal communication with the authors.

6. The initial meeting and founding group included the authors and a few additional social and behavioral scientists.

This workshop was followed by a series of meetings of a smaller executive committee. The final meeting was held in August 2014. In the interim, Grusky, Timothy Smeeding, and Snipp met with the Census Bureau, OMB, the Department of Commerce, and other agencies to discuss the development of plans for the AOS. In addition to the original group, representatives of the Census Bureau, the Internal Revenue Service (IRS), and other federal government organizations were included in the meetings and deliberations.

Throughout these deliberations, the choice of the survey vehicle became clouded by various external considerations, especially the viability of securing space on the CPS or ACS instruments. Holding such external considerations aside, one of the workshop papers expertly reviewed the costs and benefits of different survey vehicles, with the conclusion that the ACS might be the best option (Warren 2015). The SIPP, although rich in content, was rejected because its sample was not large enough to capture small immigrant groups and less common family structures. The CPS, although larger than the SIPP, was less rich in content and still too small to analyze certain immigrant groups and areas smaller than states. The ACS was even more circumscribed in content but delivered the most statistical power by virtue of its sample size. Ultimately, the group concluded that neither the CPS nor the ACS would be suitable, whereas the smaller SIPP panel with its “gold standard” linkages to administrative data contained the all-important hint that administrative data might be a way forward.⁷

Given these constraints, the committee explored the possibility of a “linkage solution” in which parent-child linkages were identified (via co-residency) in, for example, the 1990 census, and the subsequent occupation of the child was secured by linking to the 2000 census and the ACS. This approach was congruent with Census Bureau’s research program in the Center for Administrative Records Research and Analysis (CARRA). The CARRA staff had successfully linked records in the 2000 and 2010 censuses

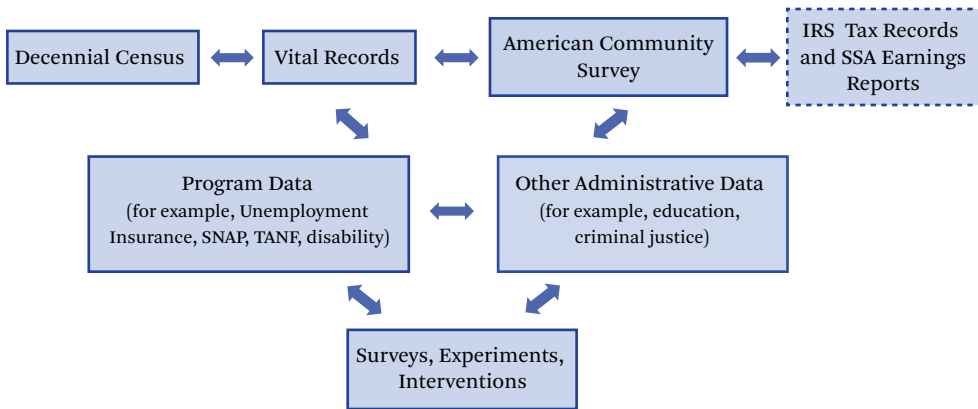
to data from the Social Security Administration and the Internal Revenue Service. These data were further linked to the 2004 and 2008 SIPP panels to form the SIPP gold standard file (Johnson, Massey, and O’Hara 2015). Based on these and other linkage projects, the committee developed a more robust and lasting project that entailed first adding links to the 1990 census, then turning to those from 1950 to 1980. The committee christened it the American Opportunity Study.

THE STRUCTURE OF THE AMERICAN OPPORTUNITY STUDY

Because the data making up the AOS already exist, the initiative adds value solely by finding low-cost ways to digitize existing data, link them, and deliver them widely and safely (see Warren 2015; Johnson, Massey, and O’Hara 2015). The AOS will rest on two types of links: intergenerational links to parents and other ancestors and intragenerational links across all censuses. The panel that results from linking censuses will be very useful in and of itself, but the research value will be even greater if it becomes possible to link them to other administrative databases and surveys. The resulting full panel is represented in figure 1. Although this figure represents the AOS as a single massive panel tracing many generations of American families from their arrival in the United States to the present, in practice it will be a *potential* dataset, in which only parts are assembled for any given research project. It is highly unlikely that any researcher would be given access to the AOS in its entirety.

The payoff to building out a full AOS, as represented in figure 1, would be substantial. If, for instance, approval to link to IRS 1040 and Social Security Administration (SSA) earnings records were secured, additional high-quality reports of income, earnings, and other variables would become available on an annual basis. Although IRS 1040 and SSA earnings reports are perhaps the most valuable linkages for the purposes of mobility research, other administrative records could be usefully incorporated

7. Indeed, a new version of the SIPP might follow the path of omitting detailed questions on income and earnings, instead asking respondents for permission to access IRS, SSA, and other administrative data to measure income, earnings, and other variables more accurately assessed by administrative data.

Figure 1. Schematic of the American Opportunity Study

Source: Authors' schematics.

as well (program participation records, incarceration records, veterans' records). The practical and legal obstacles to including additional types of administrative data are not trivial (for discussion, see Johnson, Massey, and O'Hara 2015; CEP 2017).

How will the links be made? Researchers at the Census Bureau have developed a crosswalk from the universe of all social security numbers (SSNs) ever assigned to a new set of identifiers known as the protected identification keys (PIKs). Using the names, addresses, and birthdates in the 2000 and 2010 censuses and then comparing them with the names, birthplaces, and birthdates associated with SSNs, the researchers have assigned PIKs to nearly every person record in the 2000 and 2010 censuses and the 2008 through 2016 ACS. The 1990 and earlier censuses, by contrast, have not yet been PIKed. Assembly of the AOS thus requires four more steps:

Assign PIKs to the person records in the 1990 census (and ultimately all censuses from 1950 to 1980 as well).

Use these identifiers to then link to the same individuals in the 2000 through 2010 decennial censuses, the 2008 through 2016 ACS, and ultimately future decennial censuses and ACS.

Add variables by using the same identifiers to link to data from other administrative sources (such as Social Security, Medicare

and Medicaid, Veterans Administration, Bureau of Justice Statistics, Department of Education, and Internal Revenue Service data).

Link parents to co-resident children within censuses. Verify and extend these intergenerational links by drawing on existing databases that match the SSNs of parents to those of their children.

In all likelihood, two versions of the AOS will have to be created, one that omits sensitive administrative data and other information to prevent deductive disclosure, and another highly "secure version" that could only be analyzed in federal statistical research data centers (FSRDCs). The latter highly controlled version would include administrative data and would be accessible only under stringent restrictions and protocols, such as those outlined by the Commission on Evidence-Based Policymaking. In the balance of this section, we elaborate on each of the four steps, paying special attention to the various obstacles likely to be encountered in the course of doing so.

It is useful to begin by discussing how PIKs can be assigned to each individual in the 1990 census. This procedure is carried out by using a set of variables (first name, last name, date of birth, address, sex) that, when taken together, make it possible to find an individual's SSN in the Social Security Administration's Numident file. The current PIKing procedures can likely be improved (see Warren 2015; Johnson, Mas-

sey, and O'Hara 2015). It is possible, for example, that direct census-to-census linkages will yield results that are superior to approaches that involve "going up" to the Numident and assigning PIKs. When an optimal procedure is settled upon, it can be used to redo the existing PIKs for the 2000 through 2010 censuses and the 2008 through 2016 ACS, and then to PIK, for the first time, the 1950 through 1990 censuses. The latter step will allow us to go back further in time to monitor long-term trends or carry out long-run analyses of programs.

The AOS panel will thus provide observations on individual income, education, or occupation for individuals appearing in the 1990 and earlier censuses, the relevant administrative sources (IRS 1040, SSA earnings reports), and the ACS. The final step is to match parents and children by exploiting relationship pointers in the 1990 census and by drawing on databases that link parents' SSNs to those of their children. The "Kidlink" files currently used by the IRS to determine whether tax filers are making legitimate claims to dependent children could be used, for example, for intergenerational matching in the AOS (for details and limitations, see Johnson, Massey, and O'Hara 2015). Additionally, IRS 1040 forms can be used to improve the quality and scope of parent-child matches, given that parents claiming children as dependents have been required, since 1987, to list the SSNs of the children they claim, whether the children live with the parent or not. Finally, the ACS and decennial censuses also identify children of the household head, thus providing a further source of parent-child matches (see Johnson, Massey, and O'Hara 2015).

The AOS, as designed, will provide a high-quality scaffold for monitoring mobility without the cost of mounting a new mobility survey, without further burdening existing surveys with intergenerational modules, and without troubling people by repeating questions they have

already answered. It thus provides a partial solution to the problems arising from low response rates and measurement errors for many survey-based reports, particularly earnings and income (Meyer, Mok, and Sullivan 2015). It is unlikely, however, that surveys will disappear altogether from this post-AOS world (NASEM 2017). To the contrary, the AOS would allow surveys to become more efficient because they could be used exclusively to ascertain variables that are not available in the AOS. Given the AOS's architecture, any sufficiently large survey with individual identifiers could be linked to it, which means that additional variables collected as part of that linked survey could be appended to the AOS variables.⁸ Although an analysis based on the AOS alone would suffice for a wide range of descriptive studies, a survey supplement to the AOS might be useful for studies of the causes, consequences, and social correlates of mobility and of other program and policy effects.

The AOS will also provide multiple reports on many outcomes. We know, for example, that the detailed earnings records (DER) do not accurately measure earned income, given that respondents at the bottom ranges of the income and earnings distribution often overreport DER-based earnings because of wages earned outside the Social Security system (Bollinger et al. 2015; Hokayem, Bollinger, and Ziliak 2015). Even for higher-income earners, both the DER and IRS tax data help fill in unreported and underreported earned incomes, which again speaks to the value of repeated measures of the sort that the AOS will provide.⁹

BENEFITS OF THE AMERICAN OPPORTUNITY STUDY

Will the benefits of building the AOS outweigh the costs? We address this question by describing how the AOS will assist with census operations and how it will support basic, applied, and policy-relevant research. We suspect that, re-

8. For voluntary surveys, respondent consent is required before any links can be made to administrative records, to the ACS, or to decennial censuses.

9. The AOS panel will be compromised insofar as many individuals are incorrectly linked (either intergenerationally or intragenerationally), or many individuals cannot be linked at all. The available evidence suggests that these problems will be relatively minor and can be successfully remediated with sample weights and other approaches (only some of which assume that the data are missing at random).

ardless of whether the AOS is built out in the near term, the country will eventually turn to an infrastructure of the AOS sort. This outcome is likely given the dividends to fully exploiting the country's capacity to assemble a high-quality panel. These dividends include the following:

the substantial cost savings and efficiencies that arise from reusing information that has already been collected for other purposes (rather than mounting a new and replicative data collection effort);

the capacity to base multigenerational comparisons on contemporary reports rather than recollections;

the relatively high quality of administrative data (relative to survey-based measures);

the spinoffs and cost savings to various census products that become possible by advancing methods for PIKing and intergenerational matching (see Johnson, Massey, and O'Hara 2015);

the development of a monitoring infrastructure that, by virtue of being automatically "refreshing," sidesteps the problems with unrepresentativeness that plague other long-running panels, such as the Panel Study of Income Dynamics (PSID) and the National Longitudinal Surveys (NLS);

the capacity to examine patterns of nonresponse and attrition in surveys, such as the PSID and NLS, when they are linked to the AOS;¹⁰

the opportunity to gradually grow the AOS and extend its research uses by adding new administrative records (health data, program use data);

the capacity to field leaner and more efficient surveys by using the AOS as a sampling frame and filling in core economic and demographic items before contacting survey respondents; and

the spinoff of an automatically refreshing sampling frame that, by virtue of combining

census, ACS, tax, earnings, and other sources, may be superior to any competing frames.

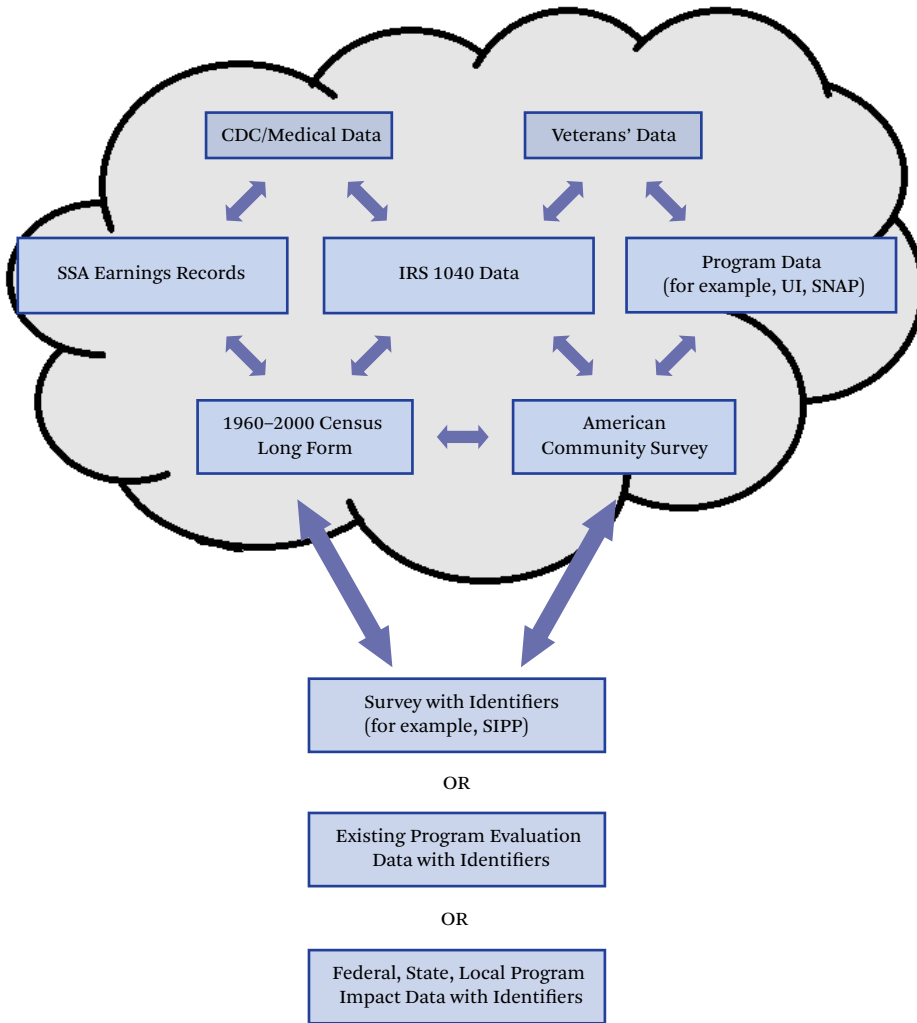
This formidable list of infrastructural benefits justifies in itself a move to an AOS-style panel. The policy and research benefits add further weight to the case for an AOS. The analysis of social mobility, which was the main impetus for developing the AOS, will of course benefit. It will be possible to carry out trend analyses of mobility, sibling analyses of shared family effects, multidimensional analyses of mobility (combining income, education, occupation, and other dimensions), and even twin analyses of mobility (given the large sample size and hence large twin population). It will also be possible to exploit the replicate measurements embedded in the AOS to complete better analyses of economic, socioeconomic, and labor market outcomes.

When additional administrative or survey data are linked to the infrastructure, even more research payoffs open up (see figure 2). It will be possible, for example, to examine the long-run effects of earlier life circumstances on any of the additional dependent variables that then become available (health, political attitudes, social attitudes, retirement behavior). There would likely be substantial payoff, for example, to linking to the National Health and Nutrition Examining Survey, Add Health, the National Election Survey, the General Social Survey, Fragile Families, the Health and Retirement Study, the Panel Study of Income Dynamics, and the National Longitudinal Surveys.

The AOS could also be used to examine the long-run effects of key independent variables that take the form of experimental treatments, nonexperimental exposure, or other types of life-cycle spells. The main payoff to the AOS is precisely this new capacity to examine the long-run effects of social programs or policies (such as the GI Bill), individual-level institutional participation (such as military service), tax policy (such as the earned income tax credit), or various types of cohort or period effects (such as the Vietnam War). This capacity is represented at the bottom of figure 2.

10. Existing surveys, such as the PSID, can be used to test the accuracy of intergenerational linkages in the AOS (by PIKing the PSID and linking it to the AOS).

Figure 2. Expanded Schematic of the American Opportunity Study



Source: Authors' schematics.

The effects of many different types of institutional experiences could also be examined. We could, for example, link to Bureau of Justice data (or state and local data), making it possible to assess the long-term effects of incarceration on earnings, recidivism, and much more (see Looney and Turner 2018). We could likewise assess the effects of training programs during and after incarceration as well as various post-incarceration conditions (such as reception

programs) on the life course of those experiencing incarceration. If we linked to Veterans Administration data on military service, we could examine the long-term effects of service on various economic and non-economic outcomes (see Autor et al. 2015).¹¹ By linking to state data on schooling, we could better estimate the economic and non-economic payoffs to various types of education. The same approach could be used to examine the effects of nurse home

11. This study combines administrative data from the U.S. Army, Department of Veterans Affairs (VA), and the Social Security Administration to analyze the effect of the VA's Disability Compensation (DC) program on veterans' labor force participation and earnings.

visiting programs, job training programs, child-care and early childhood education, school experiments, tax credits, and much more (Berlin 2016).

The independent variable of interest can also take the form of historical events that affect all or some birth cohorts. If, for example, records from the Federal Emergency Management Agency were linked to the AOS, it would be possible to study the long-term impact of Hurricane Katrina and the role of federal assistance in mitigating the disruption the storm caused to those who were exposed to it. In this case, a standard one-off assessment would be very expensive because those who were affected were relocated to sites throughout the country. The AOS would allow for an inexpensive assessment that exploited powerful quasi-experimental designs. We could likewise use the AOS to assess the long-run effects of various wars, the effects of economic recessions and crises on those living in hard-hit areas, and the effects of school shootings, terrorist attacks, and other traumatic events (such as the attacks of September 11, 2001) on those living within the affected areas (even if they moved thereafter).

For all of these examples, the analysis becomes possible insofar as those exposed to a treatment can be identified via the census or administrative data already in the AOS, new surveys that are linked to the AOS, or new administrative data that are linked to the AOS. In the limiting case, a simple list of those participating in the treatment would also suffice, assuming that approval to use that list is secured.

How would treatment effects then be assessed? If a bona fide experiment has been carried out (such as a basic income experiment), we could observe the long-run outcomes of the treatment and control groups within the AOS architecture without incurring the usual high costs of tracking participants and repeatedly administering survey protocols.¹² Although the country's main social science experiments have of course already been assessed in some fashion, these assessments frequently have not been able to fully examine long-run effects. The AOS would

also provide for high-quality assessments in the absence of an explicit control group. In this case, the AOS would have the reach and sample size to allow for matched within-community comparisons, various types of regression discontinuities, fixed effects on individuals, and all manner of related nonexperimental or quasi-experimental approaches.

This suggests that policy evaluation in the context of an existing AOS would be revolutionized. Although the analysis of past experiments and treatments will sometimes be complicated by this need to find a list of participants, the same requirement can be met more easily for future evaluations. It follows that experiments and treatments could be assessed at very low cost. Whenever a new experiment or program was established, it could automatically be evaluated—with only relatively small investments in planning—within the context of the AOS architecture.

Could research of this sort be completed without the AOS? Could the current state of affairs—relying as it does on one-off analyses of tax and other administrative data—get the same work done? The existing approach is problematic because access to administrative data is limited and meted out idiosyncratically; the analyses typically have to be completed within the context of a single administrative dataset or a relatively small number of linked datasets; the total cost is high because each analysis is completed as a one-off study; and the capacity to carry out long-run analyses is compromised (given that the pre-2000 censuses have not been PIKed and converted into a panel). If the AOS were built, all of these problems would be solved at once.

THE AMERICAN OPPORTUNITY STUDY AND THE COMMISSION ON EVIDENCE- BASED POLICYMAKING

Two critical questions affect prospects for further developing the AOS. First, how does the AOS fit within the recommendations recently issued by the Commission on Evidence-Based Policymaking? Second, how does the National

12. The basic income experiment conducted by Y Combinator Research will rely, for example, on administrative data to unobtrusively monitor key economic and non-economic outcomes.

Secure Data Service, as laid out by the CEP, relate to the AOS?

The CEP provides a principled approach to ensuring privacy and confidentiality, addressing key research questions, providing channels for public comment, and ensuring public availability of evidence. The NSDS further ensures that the capacity to generate, access, and use data and evidence can be integrated within government institutions that are adequately funded and staffed. The NSDS may be understood in this context as the “cloud” in figure 2 surrounding the top two levels of the AOS in figure 1.

The NSDS sub-agency charged with data access will receive projects, surveys, and evaluations from qualified and approved researchers. These newly received data will then be linked to existing data within the cloud (when permission to link has been granted). Without directly assembling a master file in the cloud, the NSDS will make the required linkages, prepare a secure dataset for analysis, and provide the qualified research team with a report on the quality of the matched data and data edits. The NSDS will eventually be charged with producing reports on data sensitivity and risk assessments for public release of de-identified confidential data. At the same time, each agency producing administrative data will have an office charged with producing cleaned administrative data, with these data then made available to the NSDS to link with other data for qualified and approved research.

This vision will require streamlining and modernizing standards for accessing data for research purposes. The NSDS will be guided by data-sharing agreements, data use agreements, and memorandums of understanding. As it stands, a standard template for these legal documents does not exist; instead, a patchwork quilt of laws governs data access and security. These must be updated with a uniform standard that works for the wide range of data that might be made available under the NSDS. This new standard might, for example, reconcile United States Code (USC) Title XIII (covering access to data collected by the Census Bureau), USC Title XXVI (governing data produced by the Internal Revenue Service), and the Confidential Information Protection and Statistical Effi-

ciency Act (governing data produced by the National Center for Health Statistics and other federal statistical agencies).

These CEP recommendations are in the process of being implemented. In late 2017, the U.S. House of Representatives passed the Foundations for Evidence-Based Policymaking Act of 2017 (H.R. 4174), which implements ten of the recommendations in the CEP report. It lays out the fundamental responsibilities of federal statistical agencies and the proactive duty of parent departments and agencies to support their statistical agency or agencies; it stipulates that each cabinet department and independent agency should designate a chief evaluation officer, a chief data officer, and a statistical official; it declares that agencies are to make their data assets publicly available except where the data are restricted (such as for confidentiality considerations); and it declares that agencies are to make their data assets available to statistical agencies for purposes of developing evidence (unless data-sharing is clearly prohibited by law).

However, twelve of the CEP’s recommendations were not included in H.R. 4174, and the creation of the NSDS was conspicuously among those that were absent. Because the legislation does call for the creation of an advisory board to set policies for the NSDS, it suggests that the NSDS might be established in a subsequent piece of legislation. The Bipartisan Policy Center, a Washington, D.C., think tank and a leading advocate behind H.R. 4174, indicated that this was indeed their strategy. Responding to a query about why all twenty-two recommendations made by the CEP do not appear in H.R. 4174, Bipartisan Policy Center staff reply that “the ten CEP recommendations included in the bill reflect those prioritized to build basic capacity while prioritizing important privacy protections at the outset. Future authorizing and appropriations legislation will incorporate additional CEP recommendations” (Hart and Davis 2017).

Nonetheless, because the organizational foundation for an NSDS has been in place for more than twenty years, it is hardly a radical step to formalize it. In 1994, the Census Bureau opened its first research data center (RDC) at the National Bureau of Economic Research in

Cambridge, Massachusetts. The objective was to create a secure enclave where researchers could access confidential census data in a highly controlled environment. Before any analyses could be carried out in an RDC, all researchers were carefully vetted, and their projects were likewise carefully reviewed. The output from the resulting data analyses were also reviewed by a Census Bureau employee trained in disclosure review before it was allowed to leave the facility.

Some twenty years later, twenty-nine RDCs operate around the country, and plans are to open more in the coming years. Although the early RDCs contained only data produced by the Census Bureau, the National Center for Health Statistics subsequently decided to make its confidential data available through the RDC network. Encouraged by the White House Office of Management and Budget, other agencies also began to make their data available through the RDC network. Currently, data from eleven federal agencies can be accessed through the RDC network, and more agencies are expected to become part of this system. To reflect this growth, the name of the Census Bureau RDC network was changed to the Federal Statistical Research Data Center system in 2014. It is, then, only a short step from the FSRDC system to the more ambitious NSDS. If the NSDS were to be formed, it would entail building up the capacity of the FSRDC system, broadening access to new data, and adding staff to support the new work.

It follows that the AOS aligns well with the CEP's vision for the future. When census, ACS, and tax data are PIKed and thus linkable from 1950 to the present day, the country will have an exhaustive panel that represents the country's population at any point in time, builds links across generations, and provides the rudimentary scaffolding on which a vast array of administrative data might be added. The NSDS, as outlined by the CEP, thus becomes the organizational mechanism through which administrative data are accessed and successfully linked to this scaffolding.

CONCLUSION

The growing availability of administrative data will continue to transform how public policy is

evaluated at all levels of government. The CEP, along with earlier initiatives undertaken at the OMB by the Obama administration, should be understood as important steps in expanding access to government data in policymaking. The AOS initiative should be understood, in turn, as providing the population-level scaffolding for this administrative data revolution.

These efforts to link and deploy administrative data still face many legal, bureaucratic, and political hurdles. The most commonly voiced concerns pertain, of course, to issues of confidentiality. It is important in this regard to distinguish between first-order concerns about actual compromises to identifiable data and second-order concerns about the fallout from unwarranted public worries about such compromises.

The first-order concerns are arguably the less formidable ones. This is because, as legitimate as first-order concerns are in other contexts, no special or troubling confidentiality issues arise with the AOS. In assembling the AOS, the NSDS would indeed rely on various identifiers, but these are only interim "production tools" that will ultimately be stripped from the released product and that will exist only in the cloud of figure 2. This type of procedure is already standard practice for a variety of census products and raises no new or special concerns. Likewise, the AOS would be made available only to carefully vetted researchers and research projects through federal statistical research data centers, again a long-standing and very successful delivery mechanism that raises no new or special concerns.

When the AOS is made available, the demand will likely be so high that existing FSRDCs will have to grow in size, new FSRDCs will have to be opened, and processes for accessing AOS data within the FSRDCs will have to be streamlined within the CEP guidelines.

We obviously cannot rule out the possibility of legitimate first-order concerns. Rather, our point is simply that we are currently unaware of any troubling confidentiality issues that the preceding practices, all of which are standard and ongoing, might raise in the foreseeable future, assuming that the AOS is indeed embedded in the NSDS. Because we could be overlooking legitimate concerns, a crucial part of the

debate about the AOS and the NSDS should be an open and wide-ranging discussion of the types of security breaches that might occur and how they might be prevented. The Census Bureau is, for example, currently developing new standards for the disclosure review process in direct response to concerns raised by the data security literature.

Although a full discussion of first-order concerns should be an important part of any deliberations about the AOS, our strong suspicion is that the most pressing worries will prove to be of the second-order variety, especially in light of recent revelations of data misuse by private social media companies. These revelations create a problematic climate for discussing data security even though they do not bear directly on AOS security. That is, most of us very reasonably worry about the public's perception of the AOS, not about any actual compromises to privacy that the AOS might imply. The standard prescription for such misinformation problems, and one to which we default here, is simply a call for a full and frank discussion of the facts of the matter.

We are hopeful that the public will conclude that the benefits are profound and that the risks are minimal and can be contained. If the AOS is successfully developed, it will prove to be a transformative tool that upgrades the country's capacity to evaluate programs and policies, allows for evidence-based debate about our programs and policies, and improves the science of poverty, unemployment, and other social and economic problems.

REFERENCES

- Aaronson, Daniel, and Bhashkar Mazumder. 2008. "Intergenerational Economic Mobility in the United States, 1940 to 2000." *Journal of Human Resources* 43(1): 139–72. DOI: 10.3368/jhr.43.1.139.
- Aizer, Anna, and Janet V. Currie. 2014. "The Intergenerational Transmission of Inequality: Maternal Disadvantage and Health at Birth." *Science* 344 (May): 856.
- Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muné. 2014. "The Long Term Impact of Cash Transfers to Poor Families." NBER working paper 20103. Cambridge, Mass.: National Bureau of Economic Research. Accessed October 5, 2018. <http://www.nber.org/papers/w20103>.
- Almond, Douglas, and Janet Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25(3): 153–172. DOI:10.1257/jep.25.3.153.
- Almond, Douglas, Janet Currie, and Valerie Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature* 56(4): 1360–446.
- Almond, Douglas, Hilary W. Hoynes, and Diane W. Schanzenbach. 2011. "Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes." *Review of Economics and Statistics* 93(2): 387–403.
- Ananat, Elizabeth O., Anna Gassman-Pines, and Christina Gibson-Davis. 2013. "The Effect of Local Economic Downturns on Teen Births: Evidence from North Carolina." *Demography* December 50(6): 2151–71.
- Ananat, Elizabeth O., Anna Gassman-Pines, Dania V. Francis, and Christina M. Gibson-Davis. 2011. "Children Left Behind: The Effects of Statewide Job Loss on Student Achievement." NBER working paper no. 17104. Cambridge, Mass.: National Bureau of Economic Research. Accessed October 5, 2018. <http://www.nber.org/papers/w17104>.
- Autor, David H., Mark Duggan, Kyle Greenberg, and David S. Lyle. 2015. "The Impact of Disability Benefits on Labor Supply: Evidence from the VA's Disability Compensation Program." NBER working paper no. 21144. Cambridge, Mass.: National Bureau of Economic Research. Accessed October 5, 2018. <http://www.nber.org/papers/w21144>.
- Barker, David. 1995. "Fetal Origins of Coronary Heart Disease." *British Medical Journal* 311(6998): 171–74. DOI: 10.1136/bmj.311.6998.171.
- Berlin, Gordon. 2016. "Using Evidence as the Driver of Policy Change: The Next Steps in Supporting Innovation, Continuous Improvement, and Accountability." Testimony before the Senate Finance Committee. New York: MDRC. Accessed October 5, 2018. https://www.mdrc.org/sites/default/files/Berlin_testimony_051016.pdf.
- Blau, Peter M., and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: John Wiley & Sons.
- Bollinger, Christopher, Barry T. Hirsch, Charles Hokey, and James P. Ziliak. 2015. "Measuring Levels and Trends in Earnings Inequality with

- Nonresponse, Imputations, and Topcoding." Paper presented at the Joint Meetings of the American Statistical Association. Chicago (August 1, 2016).
- Breen, Richard, Carina Mood, and Jan O. Jonsson. 2016. "How Much Scope for a Mobility Paradox?" *Sociological Science* 3(3): 39–60. DOI: 10.15195/v3.a3.
- Brown, David W., Amanda E. Kowalski, and Ithai Z. Lurie. 2015. "Medicaid as an Investment in Children: What Is the Long-Term Impact on Tax Receipts?" NBER working paper no. 20835. Cambridge, Mass.: National Bureau of Economic Research. Accessed October 5, 2018. <http://www.nber.org/papers/w20835>.
- Cancian, Maria, Carolyn J. Heinrich, and Yiyoon Chung. 2013. "Discouraging Disadvantaged Fathers' Employment: An Unintended Consequence of Policies Designed to Support Families." *Journal of Policy Analysis and Management* 32(4): 758–84.
- Cascio, Elizabeth U., and Diane Whitmore Schanzenbach. 2013. "The Impacts of Expanding Access to High-Quality Preschool Education." *Brookings Papers on Economic Activity* 47 (Fall): 127–92.
- Chetty, Raj. 2012. "Time Trends in the Use of Administrative Data for Empirical Research." Paper presented at the 34th Annual NBER Summer Institute. Cambridge, Mass. (July 9–27, 2012).
- 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, David B. Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The Fading American Dream: Trends in Absolute Income Mobility Since 1940." *Science*, April 24, eaal4617. DOI: 10.1126/science.aal4617.
- Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2015. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." NBER working paper no. 21156. Cambridge, Mass.: National Bureau of Economic Research. Accessed November 28, 2018. <http://www.nber.org/papers/w21156>.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review* 104(5): 141–47. DOI: 10.1257/aer.104.5.141.
- Commission on Evidence-Based Policymaking (CEP). 2017. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*. Washington: U.S. Department of Health and Human Services. Accessed November 28, 2018. <https://www.cep.gov/cep-final-report.html>.
- Council of Economic Advisers. 2015. "The Economics of Early Childhood Investments." January. Washington: Government Printing Office.
- Cunha, Flavio, and James J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources* 43(4): 738–82.
- Dahl, Gordon B., and Lance Lochner. 2012. "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit." *American Economic Review* 102(5): 1927–56.
- Decker, Paul T. 2014. "False Choices, Policy Framing, and the Promise of 'Big Data.'" *Journal of Policy Analysis and Management* 33(2): 252–62.
- Duncan, Greg J., and Katherine Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives* 27(2): 109–31.
- Duncan, Greg J., Katherine Magnuson, and Elizabeth Votruba-Drzal. 2014. "Boosting Family Income to Promote Child Development." *Future of Children* 24(1): 99–120.
- Einav, Liran, and Jonathan Levin. 2014. "Economics in the Age of Big Data." *Science* 346(6210): 715–21. DOI: 10.1126/science.1243089.
- Evans, William N., and Craig L. Garthwaite. 2014. "Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health." *American Economic Journal* 6(2): 258–90.
- Featherman, David L., and Robert M. Hauser. 1978. *Opportunity and Change*. New York: Academic Press.
- Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth. 2014. "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review* 104 (December): 3921–55.

- Fischer, Claude S., and Michael Hout. 2006. *Century of Difference: How America Changed in the Last One Hundred Years*. New York: Russell Sage Foundation.
- Furman, Jason. 2015. "Smart Social Programs" *New York Times*, May 11. Accessed October 5, 2018. <http://www.nytimes.com/2015/05/11/opinion/smart-social-programs.html>.
- Grusky, David B., Timothy M. Smeeding, and C. Matthew Snipp. 2015. "A New Infrastructure for Monitoring Social Mobility in the United States." *Annals of the American Academy of Political and Social Science* 657(1): 63–82.
- Harding, David J., Jonah A. Siegel, and Jeffrey D. Morenoff. 2017. "Custodial Parole Sanctions and Earnings After Release from Prison." *Social Forces* 96(2): 909–34. DOI: 10.1093/sf/sox047.
- Hart, Nick, and Sandy Davis. 2017. "Fact Sheet: Foundations for Evidence-Based Policymaking Act." *Evidence* (blog), Bipartisan Policy Center, November 30. Accessed October 5, 2018. <https://bipartisanpolicy.org/blog/fact-sheet-foundations-for-evidence-based-policymaking-act/>.
- Haskins, Ron, and Greg Margolis. 2014. *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*. Washington, D.C.: Brookings Institution Press.
- Heckman, James, and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6(1): 689–733.
- Hokayem, Charles, Christopher Bollinger, and James P. Ziliak. 2015. "The Role of CPS Nonresponse in the Measurement of Poverty." *Journal of the American Statistical Association* 110(511): 935–45. DOI: 10.1080/01621459.2015.1029576.
- Hout, Michael. 2015. "A Summary of What We Know About Social Mobility." *Annals of the American Academy of Political and Social Science* 657(1): 27–36.
- . 2018. "Americans' Occupational Status Reflects the Status of Both of Their Parents." *PNAS* 115(38): 9527–32.
- Hoynes, Hilary, Marianne E. Page, and Ann Huff Stevens. 2011. "Can Targeted Transfers Improve Birth Outcomes? Evidence from the Introduction of the WIC Program." *Journal of Public Economics* 95(7–8): 813–27.
- Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. "Long Run Economic and Health Impacts of Participation in the Food Stamp Program." *American Economic Review* 106(4): 903–34.
- Jacob, Brian, Susan Dynarski, Kenneth Frank, and Barbara Schneider. 2016. "Are Expectations Alone Enough? Estimating the Effect of a Mandatory College-Prep Curriculum in Michigan." NBER working paper no. 22013. Cambridge, Mass.: National Bureau of Economic Research.
- Johnson, David S., Catherine Massey, and Amy O'Hara. 2015. "The Opportunities and Challenges of Using Administrative Data Linkages to Evaluate Mobility." *Annals of the American Academy of Political and Social Science* 657(1): 247–64.
- Kille, Leighton Walter. 2015. "Major Administrative Datasets of the U.S. Government — All in One Place." *Journalist's Resource*, January 15. Shorenstein Center, Kennedy School, Harvard University. Accessed October 5, 2018. <https://journalistsresource.org/tip-sheets/research/websites-u-s-federal-government-administrative-datasets>.
- Lee, Chul-In, and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *Review of Economics and Statistics* 91(4): 766–72.
- Lee, Keun Bok, David J. Harding, and Jeffrey D. Morenoff. 2017. "Trajectories of Neighborhood Attainment After Prison." *Social Science Research* 66 (August): 211–33.
- Looney, Adam, and Nicholas Turner. 2018. "Work and Opportunity Before and After Incarceration." Washington, D.C.: Brookings Institution. Accessed October 5, 2018. https://www.brookings.edu/wp-content/uploads/2018/03/es_20180314_looneyincarceration_final.pdf.
- Manza, Jeff, and Clem Brooks. 2016. "Why Aren't Americans Angrier About Rising Inequality?" *Pathways Magazine*, 2016, 22–26. Accessed November 28, 2018. https://inequality.stanford.edu/sites/default/files/Pathways_Presidential_Anger.pdf.
- Massey, Catherine G., and Amy O'Hara. 2014. "Person Matching in Historical Files Using the Census Bureau's Person Validation System." CARRA working paper no. 2014–11. Washington: U.S. Census Bureau. Accessed October 5, 2018. <https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-11.html>.
- Mazumder, Bhashkar. 2015. "Estimating the Intergenerational Elasticity and Rank Association in the US: Overcoming the Current Limitations of

- Tax Data." FRB of Chicago working paper no. 2015-04. Chicago: Federal Reserve Bank. Accessed October 5, 2018. <http://ssrn.com/abstract=2620727>.
- Mervis, Jeffrey. 2014. "How Two Economists Got Direct Access to IRS Tax Records." *Science*, May 22. Accessed October 5, 2018. <http://news.sciencemag.org/2014/05/how-two-economists-got-direct-access-irs-tax-records>.
- Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29(4): 199–226.
- Milligan, Kevin, and Mark Stabile. 2009. "Child Benefits, Maternal Employment, and Children's Health: Evidence from Canadian Child Benefit Expansions," *American Economic Review* 99(2): 128–32.
- Mitnik, Pablo A., Erin Cumberworth, and David B. Grusky. 2016. "Social Mobility in a High Inequality Regime." *Annals of the American Academy of Political and Social Science* 663(1): 140–84. DOI: 10.1177/0002716215596971.
- Mitnik, Pablo A., Victoria Bryant, David B. Grusky, and Michael Weber. 2015. "New Estimates of Intergenerational Mobility Using Administrative Data." Statistics of Income working paper. Washington: Internal Revenue Service. Accessed October 5, 2018. <https://www.irs.gov/pub/irs-soi/15rpintergenmobility.pdf>.
- National Academies of Sciences, Engineering, and Medicine (NASEM). 2017. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. Washington, D.C.: National Academies Press.
- National Research Council (NRC). 2005. *Experimental Poverty Measures: Summary of a Workshop*. Washington, D.C.: National Academies Press.
- . 2016. "Using Linked Census, Survey, and Administrative Data to Assess Longer-Term Effects of Policy: Proceedings of a Workshop—in Brief." Washington, D.C.: National Academies Press. Accessed October 5, 2018. <https://www.nap.edu/catalog/23583/using-linked-census-survey-and-administrative-data-to-assess-longer-term-effects-of-policy>.
- Noyes, Jennifer. 2015. "Using Linked Administrative Data to Improve Child Welfare Programs: The Wisconsin Experience." Paper presented at the Mathematica Policy Research CIRE Forum Connecting the Dots. Washington, D.C. (November 5, 2015). Accessed October 5, 2018. <https://www.mathematica-mpr.com/events/cire-forum-child-welfare-programs-nov-2015>.
- Office of the Mayor. 2014. *The CEO Poverty Measure, 2005–2012*. New York: The City of New York. Accessed October 5, 2018. https://www1.nyc.gov/assets/opportunity/pdf/14_poverty_measure_report.pdf.
- Price, David J., and Jae Song. 2018. "The Long-Term Effects of Cash Assistance." Industrial Relations Section working paper 621. Princeton, N.J.: Princeton University. Accessed January 28, 2019. <https://dataspace.princeton.edu/jspui/handle/88435/dsp01ng451m210>.
- Reamer, Andrew, and Julia Lane. 2018. "A Roadmap to Nationwide Data Infrastructure for Evidence-Based Policymaking." *Annals of the American Academy of Political and Social Science* 675(1): 28–35.
- Shaefer, H. Luke, Sophie Collier, Greg Duncan, Kathryn Edin, Irwin Garfinkel, David Harris, Timothy M. Smeeding, Jane Waldfogel, Christopher Wimer, and Hiro Yoshikawa. 2018. "A Universal Child Allowance: A Plan to Reduce Poverty and Income Instability Among Children in the United States." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 4(2): 22–42.
- Smeeding, Timothy M., and Katherine A. Thornton. 2017. "Wisconsin Poverty Report: The Recovery from the Great Recession Lowers Poverty Rates in 2015." Ninth Annual Report of the Wisconsin Poverty Project. Madison: Institute for Research on Poverty, University of Wisconsin–Madison. Accessed October 5, 2018. <https://www.irp.wisc.edu/resource/wisconsin-poverty-report-the-recovery-from-the-great-recession-lowers-poverty-rates-in-2015/>.
- Stevens, Ann, Michal Kurlaender, and Michel Grosz. 2015. "Career Technical Education and Labor Market Outcomes: Evidence from California Community Colleges." NBER working paper no. 21137. Cambridge, Mass.: National Bureau of Economic Research. Accessed October 5, 2018. <http://www.nber.org/papers/w21137>.
- Tam, Siu-Ming, and Frederic Clarke. 2015. "Big Data, Official Statistics and Some Initiatives of the Australian Bureau of Statistics." *International Statistical Review* 83(3): 436–48.
- Trépanier, Julie, Jean Pignal, and Don Royce. 2014. "Administrative Data Initiatives at Statistics Canada." Paper presented at the 2013 Federal Committee on Statistical Methodology Research

- Conference. Washington, D.C. (November 5, 2013).
- Warren, John Robert. 2015. "Potential Data Sources for a New Study of Social Mobility in the United States." *Annals of the American Academy of Political and Social Science* 657(1): 208–46.
- Weeden, Kim A., and David B. Grusky. 2005. "The Case for a New Class Map." *American Journal of Sociology* 111(1): 141–212.
- . 2012. "The Three Worlds of Inequality." *American Journal of Sociology* 117(6): 1723–85.
- Wherry, Laura R., Sarah Miller, Robert Kaestner, and Bruce D. Meyer. 2015. "Childhood Medicaid Coverage and Later Life Health Care Utilization." NBER working paper no. 20929. Cambridge, Mass.: National Bureau of Economic Research. Accessed October 5, 2018. <http://www.nber.org/papers/w20929>.
- Wimer, Christopher, Marybeth Mattingly, Sara Kimberlin, Jonathan Fisher, Caroline Danielson, and Sarah Bohn. 2016. "Using Tax Policy to Address Economic Need: An Assessment of California's New State EITC." Stanford Center on Poverty and Inequality Research Brief (December 2016). Accessed November 28, 2018. <https://inequality.stanford.edu/sites/default/files/EITC-california.pdf>.
- Wimer, Christopher, Marybeth Mattingly, Matt Levin, Caroline Danielson, and Sarah Bohn. 2013. "A Portrait of Poverty Within California Counties and Demographic Groups." Stanford, Calif.: Stanford Center on Poverty and Inequality. Accessed October 5, 2018. http://web.stanford.edu/group/scspi/poverty/cpm/CPMBrief_CPI.pdf.
- . 2014. "Poverty and Deep Poverty in California." Stanford, Calif.: Stanford Center on Poverty and Inequality. Accessed October 5, 2018. http://web.stanford.edu/group/scspi/poverty/cpm/CPM_Brief_Poverty-Deep-Poverty.pdf.
- Yiu, Chris. 2012. "The Big Data Opportunity." London: The Policy Exchange. Accessed October 5, 2018. <http://www.policyexchange.org.uk/the-big-data-opportunity>.