"The AOS: building the evidence base for evaluating policy impacts on human outcomes and socio-economic mobility"

For the APPAM Big Data and Public Policy Workshop, November 11th, 2015

Second Draft, November 8, 2015

David B. Grusky, Stanford University

Timothy M. Smeeding, University of Wisconsin—Madison

C. Matthew Snipp, Stanford University

Abstract

Access to administrative data alone does not create good science, but it does allow better research designs as well as the ability to answer more interesting questions than those allowed by survey and experimental data alone. The question has now become how to institutionalize and regularize researcher access to public administrative records for social policy purposes? This is the question we hope to answer by building the American Opportunity Study, a vehicle for using public administrative "big data" to study policy and program impacts on human outcomes and socio-economic mobility. We describe some of the recent contributions to policy analysis and program outcomes that have used scattered bodies of public administrative data. We then describe the American Opportunity Study (AOS) which will provide a comprehensive link between survey and evaluation studies and administrative data to examine long term outcomes .The promise of AOS are described along with the current status of the project. In closing we suggest the problems and promise of the AOS for the study of policy and program impacts on human outcomes and socio-economic mobility.

I Introduction

The revolution in the use of public administrative data for research is almost in full blossom. Just a year ago, Decker (2014) argued that the big (public) 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. Federal statistical agencies are increasingly aware of the economies and usefulness of linked and shared data for research purposes (Smith, 2015). President Obama's budget and CEA both cry out for better access to public administrative data to establish evidence on program effectiveness (see White House, 2015). And congress is active on the topic as well (Murray-Ryan bill discussed below) . Increasingly most of the important and policy relevant findings in social and behavioral science research on the effects of poverty and inequality on human outcomes and social mobility could not be estimated without the availability of and access to public administrative data as the basis for this research. In the era of evidence based policy (Haskins, 2014), the next step is to both broaden and generalize access to public administrative data while continuing to protect the privacy and confidentiality of those being studied.

And nowhere is this more important than in the social policy arena which focuses on studying policy and program impacts on human outcomes, poverty and socio-economic mobility. The results of recent analyses show clearly that public efforts to support low-income families arguably produce better longer term health, learning and economic outcomes far in excess of their costs. Public investments in pre-schools, income, housing, health care and nutrition, mainly those aimed at poor families, show substantial long-term gains for the children who benefit from these programs and policies. At the same time, researchers have documented the negative effects of long term exposure to poor neighborhoods for middle class kids as well as poor ones and the long term lasting effects of fetal origins on human outcomes, health and lifespan.

Access to administrative data alone does not create good science, but it does allow better research designs as well as the ability to answer more interesting questions than those allowed by survey and experimental data alone. The question has become how to institutionalize and regularize researcher access to public administrative records for social policy purposes? This is the question we hope to answer by building the American Opportunity Study.

The next section of the paper begins by summarizing some of the import evidence based research and policy lessons on human outcomes learned from access to both state and federal administrative data on social policy programs .These include studies based on public administrative data, but more importantly those that begin with a treatment or survey which is then linked to administrative data to deepen and broaden the value of the research undertaken. We mainly address questions related to social policy and human outcomes below[1]. Economists (Einav and Levin, 2014); medical care researchers and psychologists (Walkup Yanos. 2005); and other social and behavioral science projects such as those building multi-purpose neighborhood observatories (Morana , et al, 2014), have already accomplished much from use of administrative data and will be able to do even more if the AOS becomes a reality.

In the third section of the paper, we argue we can both improve and enhance survey research and evaluation research, by systematically linking and making more effective research use of the data we already have collected. A new research project underway now at the National Research Council termed the American Opportunity Study (AOS) would create the vehicle and infrastructure upon which all qualified researchers can have regular and systematic access to public record data in a safe protected environment. The AOS infrastructure would revolutionize not only research on economic and social mobility across and within generations, but it would also provide a framework for all types of social and behavioral science researchers to enhance and improve their survey usefulness and

[1] There is also a vast literature on the effects of public administrative data on program management which we do not address

estimate the longer term effects of policy treatments and experiments. In other words, AOS would democratize access for every researcher to the types of administrative data now available only to a handful of researchers who have one way or another been granted limited access to public administrative data (Mervis, 2014).

In the last section of the paper we outline the steps needed to make the AOS a reality and the benefits for policy analysis and practitioners that are indeed the backbone of the evidence based policy revolution. Whether one thinks that better data will help demonstrate program and policy effectiveness and lead to expansions in these programs, or if one remains skeptical, and expects that data and analysis will help identify and ultimately shut down inefficient programs, the use of public big data in research can only help build the evidence base that should always support policy and practice.

II .The evidences base: what we have learned so far from access to big public data?

Public administrative data has formed the basis for social science research for decades in Scandinavia due to the linkages of family records to data registers on economic (income, earnings, wealth) and social outcomes (educational achievement, family structure, health) for the entire population of a given country . The economics profession's drive for causal analysis increasingly relies on administrative data collected over a number of years as the basis for findings (Einav and Levin, 2014). Chetty (2012) surmises that the share of microdata-based articles in the "top four" economics journals that rely on survey data declined from about 60 to 20 percent between 1980 and 2010, while the share of articles relying on administrative data rose from about 20 to 60 percent.

While non-USA based national and cross-national analyses are of use and usually contain high internal national validity for the countries studied, American scholars and policy analysts pause to consider the external validity for the US when told of a significant effect of a particular program or event on children in a different social and cultural context, say in Norway or Denmark. But increasingly we realize that we indeed have the same data

available here in the United States, if only we would make better use of it for our own research purposes. Other nations, for instance the UK, have been realizing the same reality and working toward the same types of institutional changes in their research infrastructure (Yiu, 2012).

The USA revolution in the use of administrative data for policy purposes began in the states not in the federal government. For example in the state of Wisconsin, access to child support records has allowed researchers to answer questions of importance to state policy makers and researchers for over two decades under conditions of privacy that protect the confidentiality of the persons being studied (e.g., see Cancian, Heinrich and Chung, 2013) . This project, based on mutual learning, trust and understanding has mushroomed into a larger linked database, the multi-program system file, which directly links administrative records on 25 programs to individual and family beneficiaries back to the mid-1980s. These administrative data are used annually in the Wisconsin Poverty Model, combined with the American Community Suirvey (ACS) to fully assess the effects of programs and policies on Wisconsin poverty using a National Academy of Sciences/Supplemental Poverty measure approach (Smeeding, et al, 2015) . Moreover these same techniques can be used to assess multiple program participation and anti-poverty effects in other states as well. Currently California, New York City and a number of other states have similar poverty models built in part on administrative data.

The availability and assessment of state education, health and birth records has allowed researchers in Florida, Tennessee, North Carolina and California to learn about the long-term impacts of birth conditions (Figlio, et. al, 214), government preschool programs (Chetty et al., 2011) , earnings gains from specific community college programs (Stevens and Kurlander, 2015) and the effects of economic downturns on school achievement and teen births (Ananat, et .al ., 2011, 2013). Moreover linking state data on individuals to federal data on long term income and earnings gains has allowed researchers to see how program participants have fared over several decades. And a growing body of evidence based on administrative data at the state and local levels shows that a few model social programs — home visits to vulnerable families, K-12 education programs like “Success for All “, Long Acting Reversible Contraception (LARC) pregnancy prevention programs , as well as some

community college and employment training, programs produce measurable positive impacts that can last for many years (Haskins, 2014) .

Recent work on the analysis of health and education programs which invest in children suggests that the long term development of human capital is closely tied to conditions in early childhood where it might be most beneficial for progressive policy inputs to take place (Cunha and Heckman, 2008; Heckman and Mosso, 2014; White House, 2015). Indeed in determining long-term outcomes, even pre-natal conditions have long-term consequences for children's health and other socio-economic outcomes (Barker 1995; Almond and Currie, 2011; Aizer and Currie, 2014). Earlier studies like the Head Start Evaluation suggested that learning gains from that program faded by third grade (Duncan and Magnuson, 2013). But now a body of research on the longer-term effects of high-quality preschool programs, and even moderately successful pre-school programs like Head Start, and other early-childhood interventions in nutrition and parenting consistently find that they improve a range of adult outcomes, from higher graduation rates and higher earnings to reduced crime rates (Chetty, et al, 2011; White House, 2015). These findings could not have been possible without access to public administrative data.

Studies based on administrative data also show that the earned-income tax credit, one of the government's largest tools to reduce child poverty, also reduces the incidence of low birth weight, raises math and reading scores and boost college enrollment rates for the children who benefited (Dahl and Lochner,2012; Evans and Garthwait 2012) . The Supplemental Nutrition Assistance Program, formerly known as food stamps, has been shown to have similar benefits for child recipients that can last decades with positive effects on pregnancy outcomes (Almond, Hoynes and Schanzenbach 2011) and adult obesity (Hoynes, Schanzenbach and Almond, 2012). Positive effects of many types of refundable cash transfers through the tax code in the US and Canada have increased child cognitive achievement and health (Dahl and Lochner 2012; Milligan and Stabile, 2008; Evans and Garthwaite, 2012). And finally, receiving Medicaid in childhood makes it substantially more likely that a child will graduate from high school and complete college,

less likely that an African-American child will die in his late teens or be hospitalized by age 25 (Wherry, et al. 2015).[2]

A recent historical paper that has helped solve a long term issue in income support policy is instructive. While we know that increasingly evidence based research on linked long term administrative and survey data shows that non cash and refundable tax credits improve the lives of the poor in multiple ways, much less is known about the effects of direct cash income supports for those who can and cannot work. While much ink has been spilled over the effect of cash income support programs, "negative income taxes " on shorter term outcomes like work effort and childbearing over the past 50 years , the long term effects of a pure income support program have only recently been unearthed with the use of historic program and linked records.[3] Using administrative records from the precursor to the ADC program—the Mothers' Pension program (1911-1935), researchers have been to assess the impact of cash transfers across the entire life course by matching program participants to WWII enlistment records and 1940 census records (Aizer, et al, 2014). Social Security data were then used to follow program beneficiaries until as late as 2012, allowing researchers to show that the benefits of receiving even a few years of assistance as a child could persist for 80 years or more. Receiving cash transfers increased longevity by about one year. These effects were driven by the poorest families in the sample; for them, the longevity increases were even larger (about 1.5 years of life).The results suggest 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 important strand of social policy research could not be accomplished without public administrative data. From these efforts we have learned that much of the long term benefit from cash (direct cash payments and refundable tax credits) and near cash (food, housing, nutrition) programs appears to come directly from helping low-income families

[2] Further recent studies using administrative data from Medicaid and tax records suggest that the public will recoup their Medicaid investment thorough additional taxes (Brown, et. al. 2015).

[3] Indeed the data collected by the various negative income tax studies of the 1960s and 1970s which examined mainly labor supply, has not been preserved, negating the study of long term outcomes form these experiments.

pay for basic needs like food, housing or health care, reducing the intense economic pressure they often face and improving longer term investments in child well-being over and above the effects of direct service programs like pre-school education and health care provision(Duncan, Magnuson and Votruba-Drzal, 2014 ; Furman, 2015).

In summary, it has become increasingly obvious that frontier level social, behavioral and economic research (surveys, experiments, evaluations) can make efficient and cost saving linkages to Census and administrative data to save on survey costs, improve data accuracy, exponentially increase our ability to understand the long term consequences of economic and social change, and ultimately to better the evidence base for policy. So far we have concentrated mainly on *single purpose* studies which use a specific set of administrative data to answer a relatively narrow question. We have not mentioned the use of census, administrative and survey linked data *systems* to directly investigate social and economic mobility which are at the heart of our efforts and policy concerns, but which will also enable more and better analyses of program and policy effectiveness on mobility, inequality and human outcomes more generally.

III. The AOS: from economic and social mobility to a broad scaffold for policy analysis and program evaluation on human outcomes

The belief in social and economic mobility has long been the cornerstone of American economic democracy. The belief that hard work and ingenuity will be justly rewarded with material success is central to what many, if not most Americans consider "The American Dream". It is an idea deeply embedded within American culture, and it has been an idea that has attracted generations of immigrants seeking a better life in this nation.

But what is the evidence to support or refute this claim? Has social mobility indeed declined in America? There are a handful of studies to suggest that this possibly has happened but other studies suggest otherwise (see Mazumder, 2015; Hout 2015, Chetty et al. 2014, Mitnik et al. 2013, Krueger 2012, Lee and Solon 2009). Moreover, most of this

research is limited to economic (earnings or income) mobility alone. This is an exceptionally narrow view of social mobility insofar as it gauges the income of children against the income received by their parents. It does not speak directly to the nature of the underlying economic opportunities that gave rise to those incomes. Economic mobility is in fact endogenous to the underlying structure of opportunities from which income is derived; notably the work performed by individuals—investment bankers and janitors alike within this structure. Occupational structure also represents the organization of opportunities in a society, and by the same token, how those opportunities are either facilitated by or limited by various mechanisms associated with social closure ranging from nebulous class boundaries to the rules established by labor unions.

Given the centrality of occupations, and occupational mobility for understanding the structure of opportunities in American society, assertions that Americans today have fewer opportunities than previous generations begs the question: is there less mobility today than in the past? It is impossible to overstate the importance of this question. At the same time, it is a question that is currently impossible to answer because there is no evidence whatsoever that can be used to address this matter. More than forty years have passed since there has not been a single comprehensive assessment of social mobility including income, occupation, earnings and other types of mobility (see Blau and Duncan, 1967; Hauser and Featherman, 1978).

To return to the previous question, has social mobility declined in America? The best answer to this question is that we are not really sure.. This is a striking lapse considering the profound changes that have taken place in American society over the past four decades. To mention only a few relevant changes, we can list the massive increase in women's labor force participation, the decline in manufacturing jobs and the rise of service employment, immigration and the ethnic diversification of the labor force, the decline in White male labor force participation, changes in family and household structure, and the striking increase in economic inequality, to name but a few. What has been the impact of these changes on opportunities within American society?

Project initiation and history

Amid the growing public concerns about economic inequality and the state of the American Dream, and the utter dearth of data to speak to these issues, two of the authors of this paper traveled to Washington, D.C. to meet with officials at the Census Bureau, the Office of Management and Budget, and the National Science Foundation in 2012. These conversations led to one inescapable conclusion: mounting another study of social mobility comparable to previous studies was going to be a massive, complex and expensive undertaking. The 1973 study, based largely on a monthly supplement to the Current Population Survey (CPS), cost approximately $2.0 million (personal communication, Robert M. Hauser) or more than $11.0 million in 2015 dollars. Moreover, a new study would cost many times this amount and likely exceed the entire annual budget that NSF allocates for sociological research.

With support from the National Science Foundation and the National Research Council, work commenced on planning the American Opportunity Study in 2013 to develop a plan for launching a new study of social mobility, with the idea being that such a project ought to be comparable to previous studies but also one intended to advance current theory and methods[4]. In particular, this group had to resolve two especially important issues. One was to identify a survey vehicle for the study; namely whether to use the CPS as the previous studies had done or to take advantage of the Survey of Income and Program Participation (SIPP) or the American Community Survey (ACS). The latter two surveys did not exist in 1973. The second task was to identify the most important content domains to be included in the new study.

To accomplish these tasks, this group identified a larger group of social scientists, mainly sociologists and economists who were experts in area such as the measurement of education, immigration research, and family and household composition. These individuals were invited to prepare papers in their subject matter areas for presentation at a workshop

[4] The initial meeting and founding group included the three authors of this study and a small number of additional social and behavioral scientists.

held in June 2013 at the National Academy of Science's Keck Center in Washington, D.C. These papers were subsequently published as a volume in *The Annals*(Grusky, Smeeding and Snipp, 2015) . This workshop was followed by a series of meetings of a smaller executive committee, with the final meeting being held in August 2014. In addition to the original group, representatives of the Census Bureau and others were active and a invited guests.

The issue of which survey vehicle should be used for a new mobility study became clouded by a number of external considerations. One of the workshop papers expertly reviewed the pros and cons of the surveys available and suggested that the ACS might be the best option (Warren 2015). The Survey of Income and Program Participation (SIPP) is rich in content but was rejected for having too small a sample size to capture less common immigrant groups and certain types of family structures. The CPS is less rich in content, larger than the SIPP, but still too small for certain purposes such as state and local samples. The ACS contains less content than the other two surveys but undoubtedly delivers the most statistical power by virtue of its sample size. Ultimately, other considerations led this group to decide that neither the CPS nor the ACS would be suitable, but the smaller SIPP panel with its "gold standard" linkages to administrative data might become the prototype for the larger study.

The decision to reject these surveys rested on considerations connected with the Census Bureau. First, the idea of adding questions to the ACS was flatly rejected by Census Bureau management. The ACS was annually the target for Congressional objections. It is a large costly survey and many Americans object to disclosing the information it requests. Law mandates completing the survey and a sizable number of members in the House and Senate would gladly vote to terminate the ACS. In response, the Census Bureau was in the midst of a review aimed at eliminating questions from the survey. In this environment, the Bureau management was not inclined to entertain the possibility of adding yet more, potentially controversial questions.

Using the CPS remained an option and it was not mired in politics in the same way as the ACS. However, the CPS is the federal government's survey workhorse. Each month, a variety of federal agencies fund special supplements to collect information on topics ranging from Internet access to health insurance coverage. While in time, administrative data might substitute for the CPS, to add questions pertaining to social mobility meant developing a CPS supplement and somehow getting it into the queue. Regrettably, it might have taken five years or more until it was possible to add these questions in a supplement. Furthermore, federal agencies have a priority in this queue. It was conceivable that a social mobility supplement might be delayed indefinitely as the needs of federal programs would take precedence.

Given these constraints, the committee learned about the Census Bureau's research program in the Center for Administrative Records Research and Analysis (CARRA). The staff in this center had been involved in an ambitious research program for just a short time. Work had been done linking the recent 2010 census with other data sources such as the preceding 2000 census and data from the Social Security Administration and the Internal Revenue Service, and linking them all to the 2004 and 2008 SIPP panels to form the SIPP 'gold standard' file (Johnson et al. 2015). As an alternative to the aforementioned surveys, the committee concluded that a much more robust project could be developed by linking post-war decennial census data and augmenting it with administrative data from sources such as the Social Security Administration and other agencies. This plan was dubbed the *American Opportunity Study* (AOS).

The American Opportunity Study

A little-known fact, unappreciated by social scientists and policymakers alike is that the U.S. has the basic elements of a large-scale panel, with a comprehensive array of intergenerational items. This panel has, however, gone unnoticed because it is in an unassembled form and has never been used in the extensive and ambitious way that we

envision.[5] This panel, dubbed the *American Opportunity Study* (AOS) can be assembled by the following steps:

- Assigning identification keys to the individual records in the 1990 long form census.
- Using these identifiers to then track the same individuals into the 2000-2010 decennial censuses, the 2008-2012 American Community Surveys (ACS), and ultimately future decennial censuses and American Community Surveys.
- Extending the resulting panel by using the same identifiers to link to data from administrative sources.
- Effecting intergenerational links between parents and children within the AOS by drawing on existing databases that match the Social Security numbers (SSNs) of parents to those of their children.[6]
- Once a tool has been developed for processing the 1990 census, repeating this operation until every decennial census from 1940 to the present are linked into a single panel.
- Developing algorithms for the imputation of personal characteristics from administrative records and statistical methods.

This is an ambitious plan that can only be realized by overcoming many practical and administrative hurdles. Even if the hurdles prove surmountable, we also appreciate that we would likely have to create two versions of the AOS: one that omits a great deal of information to prevent deductive disclosure and a "secure version" that could only be analyzed in Federal Statistical System (formerly Census Bureau) Research Data Centers (RDCs). The latter would be a "highly-controlled" version that includes administrative data and accessible only under stringent restrictions and protocols[7]. However, it is important to clarify the structure of the proposed AOS. (Johnson et al. 2015; Warren,2015). A schematic diagram for the AOS appears in Figure 1.

[5] See Grusky, Smeeding and Snipp. 2015a for a more complete description
[6] Since 1987 all children have been required to have SSNs at birth
[7] Data security issues are discussed below.

Figure 1 about here

The AOS construction is outlined in the following steps:

Assigning PIKs: The first step in assembling the AOS is to assign a protected identification key (PIK) to each individual in the 1990 long form census. This step is carried out by using a set of variables (e.g., first name, last name, year of birth, address, sex) that, when taken together, allow us to reliably find an individual's SSN in the Social Security Administration (SSA) Numident file. Warren (2015) and Johnson et al. (2015) discuss in detail the technical challenges presented by this procedure.

Because the 2000 census, 2010 census, and 2008-12 American Community Surveys are already PIKed, this first step is costly mainly because the 1990 long form census is not yet PIKed. Once the 1990 PIKs are assigned and the post-1990 linkages completed, we will have a panel for all individuals appearing in the 1990 long form census, with post-1990 information (e.g., education, occupation, income) available for each year in which the 1990 census respondents show up in later censuses or American Community Surveys. The same design may of course be applied to earlier decennial censuses (for instance the 1950-1980 Censures) as well as to subsequent ones. And finally the linkages can be extended to the entire Censuses, allowing for basic panel data on all individuals who participated in the Censuses, with an observation every 10 years.

Administrative linkages: The AOS panel can be supplemented by acquiring administrative records for the individuals within it. For instance, if approval to link to IRS 1040 and SSA earnings records were secured, additional high-quality reports of income, earnings, and other variables would become available on an annual basis. These data are especially important for imputing personal characteristics for persons who appear only in the short form censuses (or as we argue below, those who have been identified by SSN in any study, survey or evaluation where there is permission to link). Although IRS 1040 and SSA earnings reports are perhaps the most valuable linkages for the purposes of mobility research, other administrative records could of course, be usefully incorporated (e.g.,

program participation records, incarceration records, veterans records). The practical and legal obstacles to linking to administrative data are discussed by Johnson et al. (2015).

Intergenerational matches: The AOS panel, as described so far, would provide repeated observations on individual income, education, occupation, and other demographic variables for individuals appearing in the 1990 long form census, any linked administrative data for persons in the short form census, and subsequent censuses or American Community Surveys. In the next step, links between parents and children are established, thereby converting this simple panel into an intergenerational one. Using the existing "Kidlink" files (these identify, for each parent's SSN, the corresponding SSNs for his or her children) makes intergenerational matches possible. These files, which are currently used by the IRS to determine whether tax filers are making legitimate claims to dependent children, could in principle be used for our matching purposes as well. [8] Additionally, IRS 1040 forms can be directly 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 claimed children. Finally, the ACS and decennial censuses also identify children of the household head, thus providing a further source of parent-child matches. Although more research on these and other approaches is required, the initial evidence on intergenerational matching rates is promising (see Johnson et al. 2015).

"Sliding in" surveys: The AOS, if assembled as laid out above, would provide a high-quality scaffold or infrastructure for monitoring mobility without the cost of mounting a new mobility survey and without further burdening existing surveys with (possibly low-quality) intergenerational modules. Indeed social and economic surveys are being criticized mainly because they have low response rates combined with non-sampling errors , particularly on socio-economic reports of earnings and incomes (Meyer, and Sullivan, 2015) But this is *not* to suggest that mobility or other surveys would no longer be needed in a post-AOS world. To the contrary, the AOS would allow surveys to become more efficient vehicles, because they could be used exclusively for the purpose of

[8] See Johnson et al. (2015) for details and limitations

ascertaining variables that were not already available in the AOS. Given the AOS's architecture, any sufficiently large survey with individual identifiers could be linked to it, thus making it possible to supplement the AOS with any of the additional variables collected as part of that linked survey.[9] Although an analysis based on the AOS alone would suffice for a wide range of descriptive analyses, a survey supplement to the AOS might be useful for studies of the causes, consequences, and social correlates of mobility and other program and policy effects as outlined below.

Moreover, as Bollinger, et. al. (2015) and Hoyakem, et.al. (2015) have demonstrated , using administrative data alone, such as the Detailed Earnings Records (DER) of the Social Security Administration to accurately measure earnings is not enough. Respondents at the bottom ranges of the income and earnings scale often "over-report" DER based earnings, largely because of wages earned outside the Social Security system. For higher income earners both the DER and IRS tax data help more accurately fill in non-reported incomes. But in this case at least, one must be careful how to combine administrative and survey data when there is no one source alone that suffices.

Obstacles and Benefits

The obstacles to assembling the AOS, particularly access to administrative data cannot be overstated. However large such obstacles may be, including costs of perhaps $70 million for full linkage of all short and long form Censuses, it is important to stress that the dividends to the AOS are also sizable.

These dividends come in many forms:

- substantial cost savings and efficiencies that arise from exploiting information that has already been collected for other purposes (rather than mounting a new and replicative data collection effort);
- the capacity to characterize intergenerational parameters on the basis of contemporaneous reports (and hence obviate the need for retrospection);

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

- the capacity to exploit high-quality administrative data and high-quality Census products rather than field new and almost inevitably lower-quality surveys to gather the same information (given cost constraints);
- the spinoffs and cost savings to various Census products that accrue to 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 (e.g., the PSID or the NLS);
- the opportunity to gradually grow the AOS and extend its research uses by adding new administrative records (e.g., health data, program use data);
- the opportunity for the evaluation community to ascertain the longer term effects of any past pre-school, schooling, health, training or any other treatment that can be linked to AOS (see below) ; and
- the capacity to field leaner and more efficient surveys by relying on the AOS for core economic and demographic items.

This is obviously not to suggest that the AOS, even when supplemented with add-on surveys, satisfies all the requirements that the *Annals* volume contributors have laid out. While a great many problems remain, one of the more obvious ones is that intergenerational matches cannot be made for children with parents living abroad (at least insofar as such parents do not have Social Security numbers and are not co-residing with their children at the time of the ACS or decennial census). But still the AOS would lead to a revolution in the data we have on immigrant mobility, where the current knowledge is based on older panels started in the 1960s or 1970s and which have excluded 40 million new US immigrants (see Duncan and Trejo 2015). A mobility study which started with a sample of current US residents could use the AOS to go back in time and locate all of the adults, children and parents who participated in the 2010 Census, with questions for year of emigration for those whose parents were not born in the USA.

Other uses

Another key question is how would this architecture benefit other numerous very valuable program and policy evaluations, surveys, experiments, and treatments ? Figure 2 is similar to Figure 1, but lays out more clearly the three tiers of our larger AOS plan for

other users. The top layer is Census data links, the middle layer are surveys, studies or evaluations of social programs and the bottom layer are linked administrative datasets.

Figure 2 about here

Some of the possibilities include using the surveys like SIPP to link to respondents, children and parents to investigate intergenerational and intra-generational issues: mobility, long run outcomes of early life circumstances, outcomes, intergenerational effects on occupation and employment (see Stinson and Wignall, 2014). This is the one linkage which we have thought most about, as it is in the mobility context of our original charge.

But there are many more linkages that apply to other surveys, policy analyses and program evaluations at state and local as well as the national level:

- with permission to link one can 'look back' or "look forward' at parents/ grandparents/children as well as the current generation with such surveys as NHANES, AdHealth, NES, GSS, Fragile Families, HRS, PSID, and NLS. Given identifiers for the original sample, one could use the architecture to further examine non-response and attrition on any panel survey, thus improving their information.
- One could also use an older sample of any past 'treatment' or life cycle spell : education, military service, incarceration, or location, experimental or otherwise and then 'look forward' to examine their LR effects. One example is linkage of data to Bureau of Justice Statistics on incarceration could be used in an event history format to learn about the effects of incarceration on longer term outcomes such as earnings post release and recidivism. Linking Veterans Administration data on military service involvement would allow us to assess the net effects of service attachment to longer-term economic, health and social well-being of those who served in the military.[10] Another use would be state data on educational experiences that could be linked to national sources of earnings and income data (like that in the IRS sample) for both parents and later children as they mature into adulthood. Such a system which would be far superior than the current links to state UI records.

[10] For instance, see Autor, et al. 2015, who have combined 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.

- One could take any job training program/any child care dataset/and school experiment/ any early life health status treatment or program and do the same. One could examine outcomes for any control group as well as any treatment group where they could be "found" in a census or administrative data source.
- One could also link to any state or local survey or administrative data where one can skip 'economic' reporting to get better , but as above not perfect, income and earnings data and to examine the history of family structure issues using the AOS architecture.

The AOS will be available to be deployed for other practical applications such as extensions of current topics in policy analysis, many of which were mentioned above in section II. For example, it might be possible to examine the long-term influence of the Earned Income Tax Credit (EITC) on the outcomes of low-income families. Similarly, if records from the Federal Emergency Management Agency could be 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.

The recent emergence of tax-return analyses of economic mobility is also complementary to the proposed AOS (see Chetty et al. 2014; Chetty and Hendren, 2015; Mitnik et al. 2015). Although tax-return analyses have proven immensely important, they unfortunately do not provide the full and complete monitoring framework that the United States needs. It is not merely that there are stringent limits on the types of uses to which tax-return data may legitimately be put. The tax-return framework is additionally limited because tax returns (a) provide no information about race, ethnicity, or generational status; (b) fail to cover non-filers (and supplementing with earnings reports and other administrative sources does not fully solve the missing data problem); (c) can only be used to measure recent trends (because identifiers for children were first secured in 1987); (d) provide only low-quality occupation reports (which are only available in machine-readable form since 1996); and (e) provide educational reports, via the 1098-T form, that only pertain to college attendance. While we have learned much on social and economic

mobility and the importance of neighborhood exposure from this data, we can learn much more with the AOS (Chetty and Hendren, 2015;Chetty, Hendren and Katz, 2015) .

Although tax-return analyses will hopefully continue to be a key national resource for monitoring mobility, the AOS initiative will build a more comprehensive infrastructure by allowing for simultaneous monitoring of different types of mobility (e.g., economic, education, occupation), incorporating key demographic variables (e.g., race-ethnicity), reducing missing data problems, and constructing long trend series of trends in mobility. Even if the AOS is not linked to the Form 1040, it will allow for comprehensive analyses of this sort. It goes without saying that, insofar as permission to link to the Form 1040 is also secured, the AOS will combine all the advantages of conventional tax-return analysis with the comprehensiveness of an AOS approach, a long term scaffolding infrastructure and architecture to benefit all the social and behavioral sciences.

Next Steps

The backdrop to the AOS initiative is an ongoing effort at the Census Bureau to develop new capacities for strategically reusing administrative data from federal, state, and commercial providers. This initiative, dubbed CLIP (Census Longitudinal Infrastructure Project, 2015), is founded on a commitment to reduce the country's reliance on high-cost surveys for economic research and program evaluation. The current focus of CLIP links the 2000 Census to later ACS and administrative data . Another ongoing CLIP project is PIKing the 1940 census records and then matching them to existing PIKed products in the present day (e.g., 2000-2010 censuses) or earlier in history (i.e., pre-1950 censuses).And another links the LEHD administrative data to the 2000-2010 Censuses to examine the dynamics of earnings and income mobility as tied to employment and relocation.

Although this is an immensely useful effort for analyzing a host of substantive questions, it relies exclusively on *existing* PIKed products for linking with 1940 and possibly earlier censuses and thus omits the 1950-1990 censuses, which are critical for studies of contemporary social mobility and for developing a permanent infrastructure for

monitoring trend in mobility. The 1950-1990 censuses are an important key for our purposes because they provide income, education, and occupation information on parents (with whom contemporary workers will have been co-residing). The very youngest workers (e.g., 25 year-olds in 2015) would, of course, typically be living with their parents in 2000 and hence are available in the already-PIKed 2000 Census or 2005 ACS if they fall into that sample. However, for the bulk of workers currently in the labor force, we need to reach back earlier to the 1950-1990 censuses, which are precisely those that CLIP does not include. By PIKing those censuses, we can (a) secure parental information on contemporary workers and thus examine contemporary patterns of mobility, and (b) construct high-quality trend data that speak to ongoing concerns that rates of social mobility may be declining.

The AOS initiative is therefore an important complement to CLIP that allows us to address issues of contemporary and recent mobility. And each project will learn from the other. For instance, because CLIPP is carrying out very relevant research on PIKing technology with the 1940 census, early evidence from CLIP suggests that state birth certificates can assist in locating parents, a result that may hold when we PIK the later censuses as well. We will of course collaborate closely with the CLIP team throughout the AOS initiative.

Development of the AOS, including the necessary linkage technology and the infrastructure for authorizing and delivering data for research, needs to be carried out in phases. The phased implementation will ensure that each part of the developmental effort is soundly carried out before proceeding to the next part. We have ascertained partial funding for the first phase of the project which addresses the full complement of technical, statistical, and software needs that will ultimately allow the AOS to go into production from the Carnegie Corporation. And we have applied for additional funds from NSF and others going forward.

The first phase of AOS is underway and is described in the technical appendix and only summarized here. The first step of Phase 1 entails testing and adapting existing

scanning and handwriting-recognition technology to digitize a subset of 1990 census long-form records. This step is a prerequisite for the next two and hence will be undertaken and completed first. The key problem is that the household member names were never captured electronically from the 1990 census, and the census forms in that year were not designed to facilitate optical character recognition. It follows that scanning and "reading" this information will be a critical first step before any linkage is possible. The second and third steps, which can commence once the 1990 data are successfully read in, entail developing algorithms for assigning PIKs and for linking parents to children. The remaining three steps can be carried out concurrently with the first three.

Although the first phase addresses the software, technical, and statistical issues associated with developing the AOS, it does not entail actual production of the AOS. This will be undertaken in a separate production operation that will entail: (1) digitizing the full 1990 census; (2) PIKing the full 1990 census; (3) testing links of large samples of the1990 census with other data sets; (4) developing methodology to fill in gaps for nonresponse (e.g., using administrative data to fill in missing earnings reports in census or ACS records) and for differences in questionnaires (e.g., the 2010 census has short-form information only); and (5) establishing protocols for permitting research access to extracts of linked data.

Finally, and most recently, the National Research Council has appointed a standing committee, working with the Census Bureau, to guide the AOS effort throughout. The Committee first met on October 15-16 , 2015. The committee is headed by Mike Hout of NYU and staffed by the NRC by Connie Citro and Carol House of the Committee on National Statistics. Amongst its initial charges, the AOS committee formed three subcommittees to work on specific topics to move AOS forward. It was decided that census staff should be active in all three subcommittees.

The subcommittees include:

1. *Workshop Planning* - led by Tim Smeeding, this subcommittee will plan a workshop that will focus on "moving beyond mobility" to other uses of the AOS, many of which are suggested above. Researchers from several different areas would be invited to talk about the potential of the AOS for their work. We hope to team with CLIP in this event,

thus tying CLIP research projects and the AOS together. Committee subcommittee members are:

- Tim Smeeding, head , U of Wisconsin
- Florencia Torche, NYU
- Rucker Johnson, UC Berkeley
- Mike Hout, NYU
- Sara McLanahan, Princeton

2. *Matching and Record linkage methodology* - led by Steve Fienberg. This subcommittee will focus on understanding the implications of record linkage for AOS merged/linked files. This requires a comparative study of record linkage methods for the files at issue, and will involve getting some outside assistance form graduate students and others to do some focused analysis at the Census Bureau. Subcommittee members include

- Steve Fienberg, Carnegie-Mellon
- Jerry Reiter, Duke
- Kick Wolter, NORC
- David Grusky, Stanford

3. *Governance* - led by Matt Snipp. This subcommittee will look long term at the AOS, how it should be structured, maintained, and how its priorities should be assessed. Members include:

- Matt Snipp, Stanford
- Alan Karr, RTI International
- Vida Maralani, Stanford
- David Johnson, Department of Commerce and University of Michigan
- Yu Xie, Princeton

To accord with laws and regulations protecting the confidentiality of federal census, survey, and administrative records data, the AOS will be housed within the Federal Statistical System Research Data Center network, and access by researchers would be through existing sworn status procedures. As an added protection and to forestall the

appearance of an all-encompassing database, linkages will be performed on demand for subsets of variables needed for specific research projects.

IV. Summary and Conclusion

The need to prioritize public spending and investment is clear. Evidence based policy analysis and management ought to drive resource allocation. The greater availability of administrative data will continue to transform the analysis and management of public programs at all levels of government as well as the policy analyses listed here. And the time for action is now. President Obama's latest budget, released last month, dedicates an entire chapter to proposals to expand access to administrative data (Office of Management and Budget, 2015, chapter 7). Senator Patty Murray, a Democrat, and Wisconsin Representative Paul Ryan, a Republican, have joined forces to sponsor a bill to create a commission to recommend ways to expand access to and use of government data in policymaking, thus providing joined support for ventures like the AOS.

Of course we realize that efforts to change the way the government collects statistics face legal, bureaucratic and practical hurdles and in some cases could run afoul of privacy advocates worried about how the government tracks its citizens. A host of issues, from privacy (such as the Privacy Act of 1974 which puts limits on how the government can use administrative records) to cost are before us. But access to government administrative data that can be linked to surveys may hold the key to unlocking the causal factors behind social mobility, and the longer term effects of public programs, two important but poorly understood phenomenon

Once it is assembled, the AOS will provide social science researchers a sweeping and unparalleled view of changes over time in American society. An even grander ambition might extend the current scope of this project to even earlier censuses and administrative data. It remains to be seen whether youths coming of age in the 1950s and 1960s enjoyed a vastly larger range of opportunities than those who entered the workforce 20 years later when economic inequality increased substantially. However, this issue is at the heart of the so-called "American Dream" that stipulates that hard work and ingenuity will be rewarded

with material success. Are the rewards offered today smaller and more limited than those offered to previous generations of Americans? If we achieve the goal of creating and AOS we will have the answer.

In closing, if the AOS can be successfully developed, it will prove to be a transformative tool for the social sciences. It would be a resource of unparalleled statistical power; an opportunity for causal research on an exceptional scale; and a source of data for a wide variety of problems unprecedented in the social and behavioral sciences. We are now actively cultivating others in the evaluation and research community to see how the AOS can provide a better evidence base for policy evaluation in each of their areas of expertise, and to building the sustained user community momentum needed to create the AOS. There is much work left to be done to make the AOS a reality but the initial stages are underway, and the promise of this project makes the effort well worth it.

References

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Mune. 2014. "The Long Term Impact of Cash Transfers to Poor Families "NBER Working Paper 20103, May, at http://www.nber.org/papers/w20103

Aizer, Anna, and Janet V. Currie. 2014. The Intergenerational Transmission of Inequality: Maternal Disadvantage and Health at Birth." *Science* 344 (May): 856

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, Hilary W. Hoynes, and Diane W. Schanzenbach. 2011. "Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes." *The 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 Volume 50, Issue 6, pp 2151-2171

Ananat, E. 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 17104, June at 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. 2114 May, at http://www.nber.org/papers/w21144

Barker, David. 1995. "Fetal Origins of Coronary Heart Disease." British Medical Journal 311:171. doi: http://dx.doi.org/10.1136/bmj.311.6998.171 .

Blau, Peter M. and Otis Dudley Duncan. 1967. *The American Occup ational Structure*. New York, NY: John Wiley and Sons.

Bollinger, Christopher, Barry T. Hirsch, Charles Hokayem,and James P. Ziliak, 2015."Measuring Levels and Trends in Earnings Inequality with Nonresponse, Imputations, and Topcoding " June, *Journal of Labor Economics*, in press.

Brown, David W. , Amanda E. Kowalski, Ithai Z. Lurie. 2015. "Medicaid as an Investment in Children: What is the Long-Term Impact on Tax Receipts? "NBER Working Paper No. 20835, at http://www.nber.org/papers/w20835

Cancian, Maria, Carolyn J. Heinrich & Yiyoon Chung, 2013. "Discouraging Disadvantaged Fathers' Employment: An Unintended Consequence of Policies Designed to Support Families," *Journal of Policy Analysis and Management*, vol. 32(4) : 758-784,

Casselman, Ben. 2015. "Big Government Is Getting In The Way Of Big Data" , FiveThirtyEight , March 9, at http://fivethirtyeight.com/features/big-government-is-getting-in-the-way-of-big-data/

Census Longitudinal Infrastructure Project(CLIP),.2015. *Policies and Procedures*, October, US Census Bureau.

Chetty, Raj. 2012. "Time Trends in the Use of Administrative Data for Empirical Research." NBER Summer Institute presentation available on Chetty website .

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): 1593-1660.

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: 141-47.

Chetty, Raj and Nathan Hendren. 2015. "The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates" Harvard Univ. mimeo.

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, May, at http://www.nber.org/papers/w21156

Cunha, Flavio and James J. Heckman .2008.. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation" . *Journal of Human Resources*, 43, 738–782.

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: 1927–56.

Decker, Paul, T. 2014. "False Choices, Policy Framing, and the Promise of "Big Data" *Journal of Policy Analysis and Management, Vol. 33, No. 2, 252–262*

Duncan, Brian and Stephen J. Trejo. 2015. "Assessing the Socioeconomic Mobility and Integration of U.S. Immigrants and Their Descendants." *The Annals: Monitoring Social Mobility in the Twenty-First Century,* 657: 108-135. Thousand Oaks, CA: SAGE Publications.

Duncan, Greg J., Katherine Magnuson, and Elizabeth Votruba-Drzal. 2014. "Boosting Family Income to Promote Child Development." *Future of Children* 24 (1): 99–120

Duncan, Greg J., and Katherine Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives* 27 (2): 109-31

Einav, Liran and Jonathan Levin. 2014. "Economics in the age of big data" *Science* 346, 6210, November : 715-721, 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: Economic Policy* , 6 (2): 258-290.

Featherman, David L. and Robert M. Hauser. 1978. *Opportunity and Change*. New York, NY: 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.

Furman, Jason. 2015. "Smart Social Programs" *New York Times*, May 11 at http://www.nytimes.com/2015/05/11/opinion/smart-social-programs.html?ref=opinion&_r=0

Grusky, David B., Timothy M. Smeeding and C. Matthew Snipp. 2015 (eds.). *The Annals: Monitoring Social Mobility in the Twenty-First Century,* 657: 63-82. Thousand Oaks, CA: SAGE Publications.

Grusky, David B., Timothy M. Smeeding and C. Matthew Snipp .2015a."The American Opportunity Study: A Link to the Past and a Bridge to the Future" presented to the annual meeting of the Population Association of America, April 30-May 2, 2015, San Diego, CA.

Haskins, Ron. 2014*. Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy.* Brookings Institution Press, Washington DC., December.

Heckman, James and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, Annual Reviews, vol. 6(1), pages 689-733, 08

Hout, Michael. 2015. "A Summary of What We Know About Social Mobility." in *The Annals: Monitoring Social Mobility in the Twenty-First Century,* 657: 27-36. Thousand Oaks, CA: SAGE Publications.

Hokayem, Charles Christopher Bollinger and James P. Ziliak .2015. " The Role of CPS Nonresponse in the of Poverty" , *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2015.1029576

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2012. "Long Run Impacts of Childhood Access to the Safety Net", NBER Working Paper 18535. Cambridge, MA: National Bureau of Economic Research, November. http://www.nber.org/papers/w18535

Johnson, David S., Catherine Massey, and Amy O'Hara. 2015. "The Opportunities and Challenges of Using Administrative Data Linkages to Evaluate Mobility." In *The Annals: Monitoring Social Mobility in the Twenty-First Century,* 657: 247-264. Thousand Oaks, CA: SAGE Publications.

Krueger, Alan B. 2012. "The Rise and Consequences of Inequality in the United States." Speech delivered to the Center for American Progress, January 12. Available from: http://www.whitehouse.gov/sites/default/files/krueger_cap_speech_final_remarks.pdf .

Lee, Chul-In and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *Review of Economics and Statistics* 91: 766-772.

Meyer, Bruce D., Wallace K.C. Mok and James X. Sullivan. 2015. "Household Surveys in Crisis", *Journal of Economic Perspectives* Volume 29, Number 4, Fall: 199–226

Morana, Emilio F., Sandra L. Hofferth, Catherine C. Eckel, Darrick Hamilton, Barbara Entwisle, J. Lawrence Aber, Henry E. Brady,Dalton Conley, Susan L. Cutter, Klaus Hubacek, and John T. Scholz. 2014. "Opinion: Building a 21st-century infrastructure for the social sciences" PNAS , November 11, 111 (45) : 15855–15856, at www.pnas.org/cgi/doi/10.1073/pnas.1416561111

Mazumder, Bhashkar.2015. "Estimating the Intergenerational Elasticity and Rank Association in the US: Overcoming the Current Limitations of Tax Data",Federal Reserve Bank of Chicago ,June, at http://ssrn.com/abstract=2620727

Mervis, Jeffrey. 2014. "How Two Economists Got Direct Access to IRS Tax Records" , *Science*, May 22, at http://news.sciencemag.org/2014/05/how-two-economists-got-direct-access-irs-tax-records

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. 2013. "Social Mobility in a High Inequality Regime." *Stanford Center for Poverty and Inequality Working Paper*. Stanford, CA: Stanford University.

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 D.C.: Statistics of Income Division, Internal Revenue Service.

Office of Management and Budget .2015. "Budget of the United States Government, Fiscal Year 2016" at https://www.whitehouse.gov/omb/budget/Overview

Smeeding, Timothy M., J. Isaacs, and K. Thornton. 2015. "Wisconsin Poverty Report: Poverty Rises in 2013 Despite Growth in Jobs" Seventh Annual Report of the Wisconsin Poverty Project. Institute for Research on Poverty, University of Wisconsin–Madison: (April).

Smith Katherine R. 2015. as cited in " Major administrative datasets of the U.S. government — all in one place " , Shorenstein Center, Kennedy School, Harvard University ,January, at http://journalistsresource.org/tip-sheets/research/websites-u-s-federal-government-administrative-datasets#sthash.ynD5hHzk.dpuf

Stevens, Ann, Michal Kurlaender, and Michel Grosz. 2015. "Career Technical Education and Labor Market Outcomes: Evidence from California Community Colleges", NBER Working Paper No. 211137, April , at http://www.nber.org/papers/w21137

Stinson, Martha and Christopher Wignall .2014. "Fathers, Children, and the Intergenerational Transmission of Employers ". SIPP papers No. 265 .U.S. Census Bureau, at https://www.census.gov/content/dam/Census/library/working-papers/2014/demo/SIPP-WP-265.pdf

Wagner, Deborah and Mary Layne, "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software," CARRA Working Paper Series, #2014-01, U.S. Census Bureau, July 1, 2014 , at https://www.census.gov/srd/carra/CARRA_PVS_Record_Linkage.pdf.

Walkup, James T. and Philip T. Yanos. 2005. "Psychological Research With Administrative Data Sets: An Underutilized Strategy for Mental Health Services Research" *Professional Psychology, Research and Practice* 36(5): 551-557

Warren, John Robert. 2015. "Potential Data Sources for a New Study of Social Mobility in the United States." in *The Annals: Monitoring Social Mobility in the Twenty-First Century,* 657: 208-246. Thousand Oaks, CA: SAGE Publications.

White House. 2015 "THE ECONOMICS OF EARLY CHILDHOOD INVESTMENTS " , January at https://www.whitehouse.gov/sites/default/files/docs/early_childhood_report_update_final_non-embargo.pdf

Wherry, Laura R. , Sarah Miller, Robert Kaestner, Bruce D. Meyer. 2015. "Childhood Medicaid Coverage and Later Life Health Care Utilization", NBER Working Paper No. 20929, at http://www.nber.org/papers/w20929

Yiu, Chris.2012. "The Big Data Opportunity: Making government faster, smarter and more personal." *The Policy Exchange* at http://www.policyexchange.org.uk/publications/category/item/the-big-data-opportunity-making-government-faster-smarter-and-more-personal

Figure 1. Schematic design of the American Opportunity Study

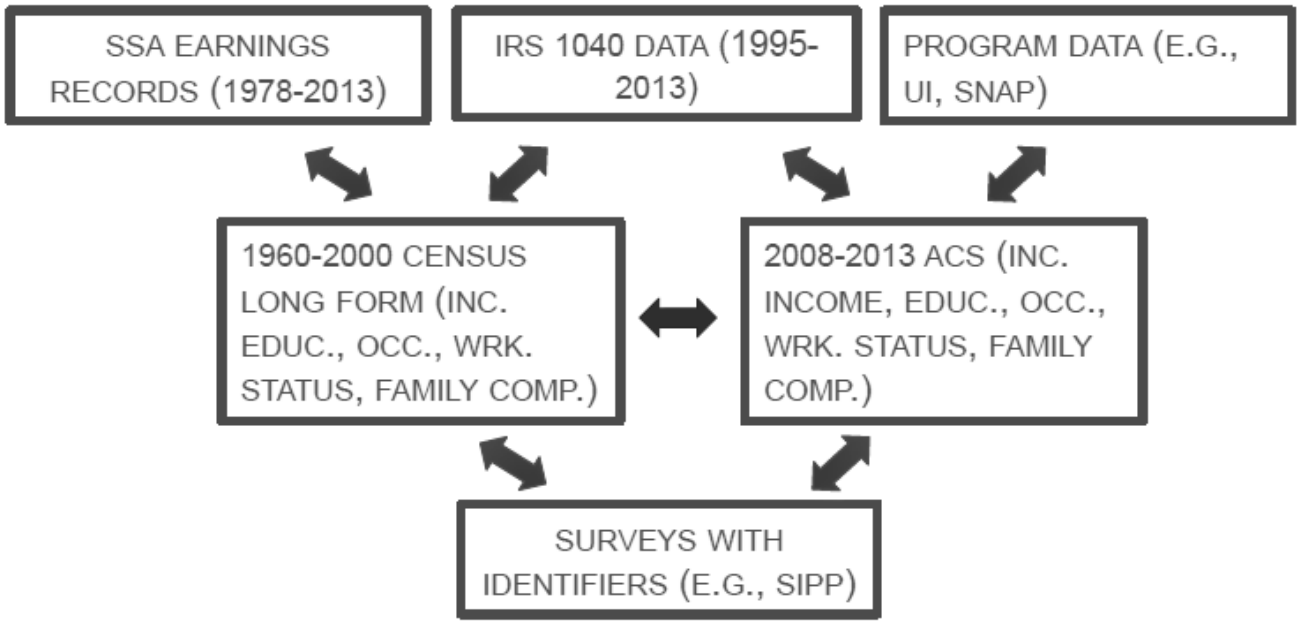
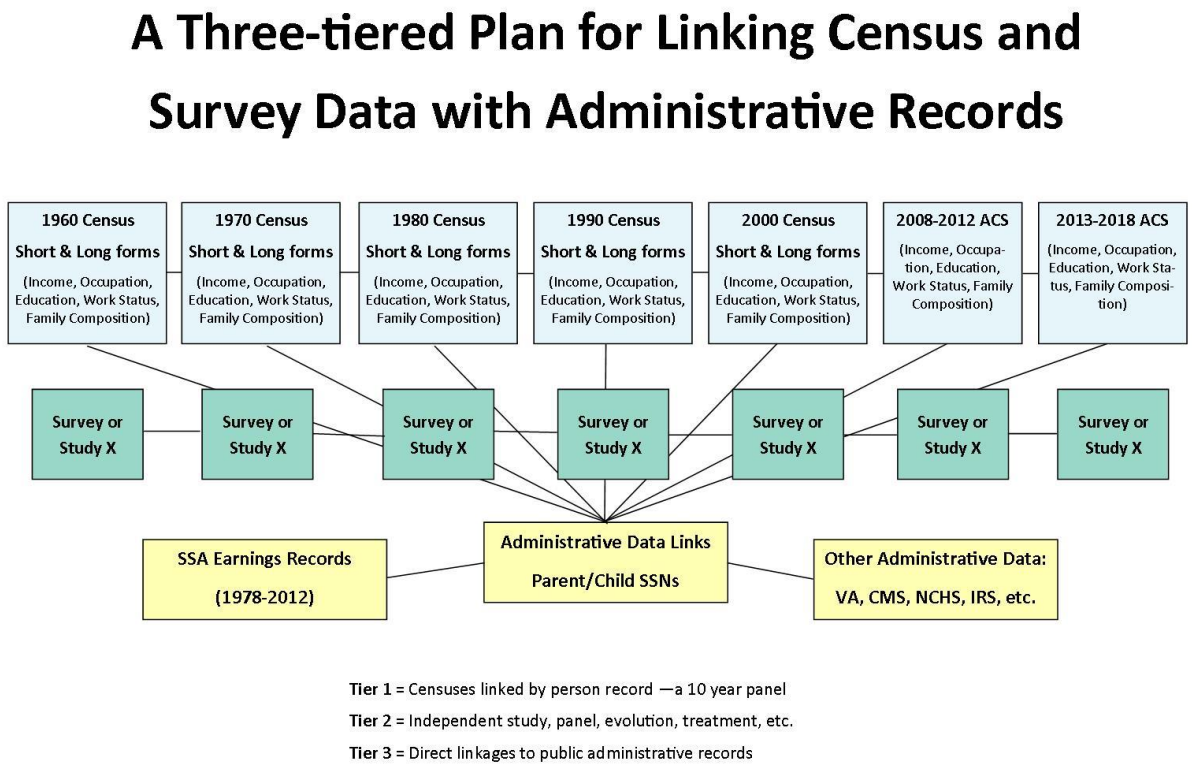


Figure 2:



Tier 1 = Censuses linked by person record —a 10 year panel

Tier 2 = Independent study, panel, evolution, treatment, etc.

Tier 3 = Direct linkages to public administrative records

Technical Appendix:

The first phase of the AOS project includes the following four steps :

Step A: Testing and selecting digitizing software and hardware to accurately capture names and addresses from a "truth deck" of 1990 census microfilm long-form records—this information, along with birth date (age and year of birth are already contained in the 1990 census electronic records), is required for Steps B and C. In the 1990 census, names and addresses were not electronically captured for tabulation purposes; they are currently available only in handwritten form on 62,000 reels of microfilm. Before any digitizing software products can be tested and a vendor selected, a necessary step is to develop a "truth deck," a small sample of images of page 2 of the 1990 questionnaire, which lists all household members and provides the household address, scanned from the microfilm. This truth deck will be used to evaluate optical character software systems in terms of accuracy of reading the scanned handwritten responses and converting them to electronic form. The chosen system will be used to digitize a larger sample of scanned images and link them to the full (electronic) census records for the sample to feed into Step B.

Step B: Developing the statistical methodology to assign identifiers (PIKs) to the augmented 1990 census electronic records and to link individual records across the 1990-2010 decennial censuses and the 2008-2013 American Community Surveys (ACS), building on and refining existing Census Bureau PIKing and matching software. Johnson et al. (2015) describe the probabilistic procedure to match identifying information (principally name and exact date of birth) from the file of interest (census, survey, administrative records dataset) to restricted Social Security Administration data to assign unique protected identification keys (PIKs). The PIKs then make possible matches among any two or more files that have gone through the process to assign PIKs The output from Step A will be used to refine the Census Bureau's PIKing methodology for assigning identifiers to the census records. The PIKing process entails assigning protected identification keys to person records using the Numident database (which contains all Social Security Numbers (SSNs) ever assigned). The Numident contains the full name, full birth date, sex, race, state or country of birth, and parent's first and last names (with name changes trackable because each name associated with an SSN is recorded). Under current Census Bureau protocols, a probabilistic matching algorithm named PVS is deployed , an algorithm that takes into account the address, full name, sex, and full date of birth. To compare names, PVS measures Jaro-Winkler distances between names in the input and reference files, thus allowing it to accommodate variability in spelling. Once PIKed, records can be linked with other sets of PIKed records.

One complication in the PIKing work for this project is that the 1990 census contains only the birth year for household members and not the full birth date. The potential problems from PIKing with less than optimal information may, however, be less than might be anticipated. The Census Bureau has already PIKed all records in the 1990 Content Reinterview Survey (CRS), which includes 20,832 adults and 7,146 children. Using a modified version of the PVS on the CRS, a match was found for 87.3 percent of the adult records and 72.8 percent of the children records, results that are roughly comparable to those secured for the 2010 census (Johnson et al., 2015). These results are very encouraging, and they can very likely be improved on by developing better algorithms (e.g.,

fuzzy matching) and by deploying additional data for the purposes of effecting the match. Input to improvements to the PVS will come from: (a) applying the CLIPP program's ongoing research on PIKing to the case of the 1990 census; and (b) drawing on the statistical recommendations of the commissioned papers on improving PIKing technology.

Step C: Effecting intergenerational matches between parents and children within the resulting AOS by exploiting relationship pointers in the 1990 census and by drawing on databases that link the Social Security numbers (SSNs) of parents to those of their children. There is a wide range of sources that may be used to match parents to children within the AOS infrastructure. The census or ACS may be used, for example, to identify the adults with whom a child is living, with such co-resident adults presumably indicating a child's "social parents." It is also possible to use the SS-5 Form for the period prior to Enumeration at Birth, which serves as the application for an SSN, to identify the likely biological or adoptive parents. The SS-5 Form not only includes the mother's and father's SSN, but also indicates whether the mother and father are legal guardians or the natural or adoptive parents. Finally, the Form 1040 indicates which adults have claimed the child as a dependent, thus indicating "financial parenthood." It follows that the AOS may in principle be used to distinguish between social, biological, and financial parenthood. Moreover, the AOS can (imperfectly) detect changes in family situations during both childhood and adulthood, thus making it possible to capture some of the complex family histories and arrangements that may affect mobility. If new types of family complexity (e.g., single parenthood, blended families) are indeed working to reduce mobility, the AOS may give us the capacity to detect just that. The purpose of Step C, then, is to explore these types of relationship pointers and determine which should be included in the production version of the AOS.

Step D: Developing and testing a tool that enables on-demand links and extracts of 1990 census data with more recent censuses and other administrative and survey data. The Census Bureau has a data governance structure extending from project proposal and review to data provisioning and monitoring. A data management system automates data requests and microdata access. This system would be fortified and expanded for the AOS including extranet access to the standing committee for proposal submission and comment, inclusion of metadata and information on analytic universes available through linking the decennial census files. Data stewardship is provided through the data management tool during research on the AOS, including required reporting for administrative data providers and disclosure avoidance review to remain compliant with contractual and statutory obligations.