I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

# The Big Data Era and an Integrated Mode of Inquiry for Social Policy-Relevant Research

NICHOLAS MADER*

## I. INTRODUCTION

In his 2013 State of the Union address, President Obama signaled a major federal priority shift towards increased investment in early childhood education, saying "Every dollar we invest in high-quality early childhood education can save more than seven dollars later on – by boosting graduation rates, reducing teen pregnancy, even reducing violent crime."[1] The evidence for this remarkable 7-to-1 return on investment is based on an ambitious study of the long-term effects of a high quality preschool program offered to youth, the Perry Preschool Project. Perry was initiated in the early 1960s to serve a disadvantaged, minority population of youth in Ypsilanti, Michigan. With a view towards demonstrating the value of receiving this program over other available alternatives, children were randomized into groups who were or were not allowed to attend the Perry program, and were then tracked over the subsequent 50 years to determine whether the lives of youth who enrolled in the Perry program would follow a different trajectory than those in the study who did not. They did, and significantly so.[2]

Given the significance of this study in national policy conversation, it may be surprising that the total number of youth involved in Perry

* Senior Researcher, Chapin Hall at the University of Chicago.

[1] See "Remarks by the President in the State of the Union Address," The White House Office of the Press Secretary, last modified February 12, 2013, http://www.whitehouse.gov/ the-press-office/ 2013/02/12 /remarks-president-state-union-address.

[2] James J. Heckman et al., "The Rate of Return to the HighScope Perry Preschool Program," *Journal of Public Economics* 94, no. 1-2 (2010a): 115-16.

was 123. It should be understood that the study was, in both implementation and follow up, costly and rigorous, and the academic literature has done well to debate the ability to draw systematic conclusions from the study.[3] But what seems compelling is the possibility that, in the Era of Big Data—where electronic systems can monitor and analyze large scale on-line purchases for fraud, or track all political donations to identify patterns of abuse—we believe we can do better in tracking youth developmental outcomes going forward.

The potential for new models of data collection for informing social policy does not end with larger sample sizes. The limitation with the Perry study is not simply that it is small, but that it is uncertain as to whether the same benefit would be realized for youth of different backgrounds or types of disadvantage, different family or neighborhood contexts, different youth developmental curriculum or staff qualifications, or for other changes in societal context in the fifty years since the Perry program was initiated. What the Perry study answers is the question of whether there is rationale for federal or state governments to increase funding for early childhood programming relative to other priorities. To address the more technocratic questions of what types of programming, in which communities, with what supports expanded programming should be undertaken, different evidence is needed.[4]

---

[3] See e.g. Michael L. Anderson, "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association* 103, no.484 (2008); James Heckman et al., "Analyzing social experiments as implemented: A reanalysis of the HighScope Perry Preschool Program," *Quantitative Economics* 1, no.1 (2010b).

[4] The Head Start Impact Study (HSIS) was undertaken in 1998 by the Administration for Children and Families under a mandate from the federal Congress to provide evidence of the efficacy of Head Start given large federal investments in the program. The HSIS was larger, more up-to-date, and more broadly representative of youth populations than Perry, involving approximately five thousand youth from 23 states, 84 grantee/delegate agencies. Like Perry, the HSIS tracked a range of youth developmental outcomes although it naturally has had less long-term follow-up. While the HSIS is an important update needed to examine the conclusions of the Perry work in the modern policy and programmatic context, it was still intended to inform the very general policy questions of whether and how the average program impacts nationally-representative populations, rather than more local ones. (The HSIS has mixed findings, showing some short-term effects for general populations, and more encouraging benefits for children from more disadvantaged families. For a detailed description of the report and its findings, see Mike Puma, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer. *Third Grade Follow-up to the Head Start Impact Study Final Report* (Washington, D.C.: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2012) available at http://www.acf.hhs.gov/sites/default/files/opre/head_start_report.pdf.

In the current era, some of the "biggest" data available to the social sector is the administrative data of public agencies, which systematically measure key characteristics, activities, and developments of the populations that they serve. Particularly for urban systems, this can be a large and diverse purview where, for example, a city tracks all youth enrolled in Head Start centers, representing both a wide (literally, the full) range of youth populations who are enrolled and types of centers they attend. Through linking administrative datasets, youth completing preschool through Head Start programming can be tracked into enrollment into the school system and followed until they graduate and leave, and tracked in contact with other agency services along the way. While these data are by design intended for management and thus available for improved service through goal-setting, administration, and tracking, they can also be extremely useful for rigorous, academic-level research to determine how to learn from and improve on practices.

This essay describes various modes of inquiry from academic work that can be combined for the purpose of guiding policy, and the properties of administrative data that enable this process. I call this "*In Situ* research"—translatable as either "on site" or "in the original place"—based on the most important feature of this integrated mode of inquiry, namely, that administrative data allow researchers to use the same populations, programs, policies, and scope in the study as are faced by policy makers in constructing the policy.

## II. TRADITIONAL MODES OF INQUIRY TOWARDS INFORMING SOCIAL POLICY

The current surge in predictive analytics represents applications where big data must be boiled down to help inform action, but where the action to be taken is familiar. Production of flu vaccinations are already capable of high capacity, but the data analyzed by Google Flu Trends to inform the likely timing and magnitude of flu epidemics help mobilize prevention measures.[5] Buying air flights is a familiar action, but recommendation engines such as the current price trend engine on Kayak.com can help crunch information on pricing patterns to know when to buy.[6] We (arguably) know how to reach out to and

[5] Or at least this was the great confidence of recent years until the promise of these methods was tempered with uneven predictions, documented by many writers including David Lazer, Brian Kennedy, Gary King, & Alessandro Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis," *Science* 343, no.6176 (2014).

[6] See "Price Trends & Tips Explanation," http://www.kayak.com/price-trend-explanation, accessed June 29, 2014.

support youth who are becoming disengaged from school, but at-risk indicators on all youth in a given school district help organize and triage efforts by identifying youth who are most at risk.[7]

However, the hardest questions for social policy involve not just knowing when or how to take a known action, but require understanding of which among many different actions will yield the best result. Causal questions are framed by counterfactuals: how would the patterns that we observe change in the hypothetical world where one element (or one definite set of elements) was changed? Correlation establishes that cities with larger police forces have more crime. The act of asking a policy question enforces the need for determining causality by framing new policy actions in a world of counterfactuals: "In a world where I increase the size of my police force, would crime be lower?"

J. J. Heckman describes three prototypical policy evaluation questions, each of which requires a greater understanding of causality than the last. [8] To paraphrase:

1. What impact did a historical policy have on the population to which it was targeted in the environment in which it took place?

2. What impact would a policy implemented for one population or one environment have on a new population and/or in a new environment?

3. What impact would a never-before-seen policy have on a new population and/or in a new environment?

For research investigations to speak to the first question, the key ingredient is internal validity. That is: with how much validity does the investigation address the question of causality for the policy based on available evidence on that policy? The key ingredient for investigations to speak to the latter two questions is external validity: with how much success can the evidence available from one setting be applied to another? Table 1 lays out three broad modes of inquiry in

---

[7] Lynh Bui, "Montgomery Schools look for Dropout Indicators Early On," *The Washington Post*, August 11, 2013, http://www.washingtonpost.com/local/education/montgomery-schools-look-for-dropout-indicators-early-on/2013/08/11/6d41f0b0-02b9-11e3-9259-e2aafe5a5f84_story.html.

[8] James J. Heckman, "The Scientific Model of Causality", *Sociological Methodology* 35, no.1 (2005): 3, 7.

the social sciences which encompass both quantitative and qualitative methods that can be used to address the validity of causal claims.

Table 1 - Advantages and Disadvantages of
Traditional Modes of Inquiry in the Social Sciences

| Mode of Inquiry | Common Advantages | Common Disadvantages |
|---|---|---|
| Randomized controlled trials (RCTs) | Internal Validity | Cost,<br>External Validity |
| Non- or "quasi-" experimental | Cost | Detail,<br>Internal Validity |
| Qualitative/primary data collection | Nuanced understanding of causal mechanisms | Actionability at scale,<br>Generalizability |

A. *Internal Validity of Various Modes of Inquiry*

Internal validity can be difficult or impossible to confirm, but the challenges are familiar. For internal validity in policy investigations, the gold standard is typically held to be the randomized controlled trial (RCT), which is designed to isolate the impact of one causal factor—the policy of interest—on one or more outcomes.[9] Within medicine, RCTs are used to evaluate the efficacy of a given treatment regime on health outcomes, for example, a new medication for lowering cholesterol used to decrease the risk of heart attacks. While many other factors may affect the likelihood of a heart attack such as gender, age, baseline cholesterol, family medical history, and diet, trials conventionally randomize subjects into treatment and control groups so that the only characteristic by which the treatment group

[9] For discussion of the validity of experimental methods in contrast to other statistical methods raised in this section, see William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Boston: Houghton-Mifflin, 2002), 3-7.

systematically differs from the control group is the medical treatment in question. This setup is designed to directly address the counterfactual question: "For a given population, if all we change is the administration of a drug, what impact would likely be realized in the outcome of interest?"

However, most social policy questions do not lend themselves to randomized controlled trials. We cannot ethically or feasibly randomize youth into private versus public schools, or municipalities into property tax regimes. The social sciences have long worked towards developing so-called "quasi-experimental" statistical methods to seek causality in these contexts. But whereas RCTs theoretically need only the fact of randomization to isolate and test only one potential systematic solution, determinations of causality within quasi-experimental investigations require the ability to rule out the many potential types and directions of causality operating within a system. The internal validity of these investigations thus comes down to data quality in terms of breadth in representing these many factors. To continue the medical example above, if a drug were taken by more affluent patients on average, it would be difficult to isolate the effect of that drug on heart attack risk from other factors associated with affluence such as access to other high quality medical care, better diet, or lower stress on average. Any association between use of the drug and health outcomes is in question unless these factors can be accounted for.

### B. *External Validity of Various Modes of Inquiry*

The limits to external validity of quasi-experimental inquiry is ultimately with data, which may be weaker or silent in informing more ambitious extensions or modifications of policy in the second and third policy evaluation questions above. But this is where qualitative modes of inquiry—which are typically those regarded as being the least generalizable—play a critical role in complementing the evidence basis of other work towards external validity. Qualitative methods generally involve primary data collection methods, practiced in order to determine causal mechanisms when existing quantitative data are limited. These methods include survey design, ethnography, focus groups, and various structures of interview methods. These methods are generally characterized by intentional pursuit of understanding a phenomenon of interest. In this way, within the social sciences these investigations are the equivalent of "basic science," necessary for guiding conception and application of thoughtful policies. Because they are generally time consuming and labor intensive in data collection and processing of rich data, these methods typically trade

off scale and systematicity for the sake of obtaining a nuanced understanding of the subject matter.

In the view that external validity for RCTs and quasi-experimental modes is ultimately bounded by the data, it is then natural to realize that use of qualitative methods can complement quantitative methods towards an evidence basis that does have external validity.[10] While qualitative methods cannot determine the general impacts of a given policy, they can be used for *ex post* evaluation of a policy to motivate and direct new efforts for adjusting the policy, identification of promising populations to receive benefits, and determination of new elements of data collection for the next iteration of evaluation.

## III. *IN SITU* RESEARCH AS AN INTEGRATED MODE OF INQUIRY

While the complementarity between these methods is not new, the improving quality, scope, and interoperability of human service records creates new opportunities for mixed methods. As introduced above, the fact that *in situ* research is undertaken in the same setting to which policy is applied helps coordinate the agendas and language of researchers with policy makers, which can expedite the process of exploration and direction of improvements in social policy.

### A. *Administrative Data as Big Data in the Social Sector*

Public agencies from school district offices to health and human services, to juvenile and adult criminal courts, to child and family welfare, to employment security, to city public health departments maintain records on human populations from literally birth to death, and a wide range of critical human welfare concerns in between. These domains represent, by design, the most important human affairs for which we consider improvements in social policy.

The administrative data records from a given public agency are primarily meant to track meaningful characteristics of and interactions with human populations served by those agencies. Modern transactional databases make it possible for many administrators to straightforwardly maintain high quality, systematically comparable records on agency activities. In addition,

---

[10] While mixed quantitative/qualitative methods investigations are increasingly common in the social science, there have historically been paradigm wars between researchers in each camp. See David L. Morgan, "Paradigms Lost and Pragmatism Regained. Methodological Implications of Combining Qualitative and Quantitative Methods." *Journal of Mixed Methods Research* 1, no. 1 (2007): 48-49, for a discussion of the history and positions taken in the quantitative/qualitative debate.

many privately-run social service organizations such as public libraries, city park districts, or youth afterschool or other support programs that are for- or non-profit also track data in similar ways. Their service records could prospectively also be linked to other databases.

Other references describe the legal, political, and logistical issues involved with centralizing, linking, and reporting information from administrative data records.[11] Here, I suggest the following features of such a linked administrative data system that make it of high value for research:

- **Common measures** on all members of a population of interest, which allows researchers to measure outcomes and related factors for any given partitioning of individuals in to treatment or comparison groups, permitting many counterfactual question to be asked.

- **Large sample sizes** which imply high "statistical power," i.e. the ability to identify signal through the noise. This makes answers to existing research questions that much more reliable, such as when making a study of Head Start versus no-Head Start for tens of thousands of youth instead of using data sets that are smaller by an order of magnitude or more. Greater statistical power also makes it possible to address more detailed questions with statistical reliability, such as asking how certain types of Head Start centers compare—such as school-based versus community-based centers—or how Head Start differentially impacts youth of different given backgrounds or types of neighborhoods.

- **Longitudinal design**, meaning that individuals are followed over time. This is a feature shared with the more useful research data sets—such as the National Longitudinal Survey of

[11] V. Joseph Hotz et al., "Administrative data for policy-relevant research: Assessment of current utility and recommendations for development," *Report of the Advisory Panel on Research Uses of Administrative Data of the Northwestern University/University of Chicago Joint Center for Poverty Research*, (1998): viii-ix; Paul G. Stiles and John Petrila, "Research and confidentiality: Legal issues and risk management strategies," *Psychology, Public Policy, and Law* 17, no.3 (2011): 333; Kumar Prashant, "An Overview of Architectures and Techniques for Integrated Data Systems Implementation," (2011): 4. Retrieved from http://www.sp2.upenn.edu/aisp_test/wp-content/uploads /2012/12/0033_12_SP2_ Architectures_Techniques_Data_Systems_000.pdf.

Youth[12]—that are intentionally gathered for research purposes, as it allows linking of early causes to later effects. For example, this allows us to see how youth who are or are not engaged with transition coaching in their senior year of high school may follow different trajectories into freshman year in college, through college graduation and into the labor force.

- **Repeated sampling for consecutive cohorts**, meaning that it is possible to compare measures for equivalent populations over time. By contrast to the opportunity in longitudinal sampling, repeated sampling of many cohorts allows us to see how tweaks in transition coaching policies influence the transition to freshman year of college for the class of 2014, the class of 2015, and so on.

- **Tracking of institutions as well as individuals**. Data sets collected for the purpose of social science research often focus on detailed characterizations of individuals and their experiences, but less or not at all on the details of the institutions or individuals that serve them, including schools, welfare offices, caseworkers, job training centers, or afterschool sites.[13] Having longitudinal tracking of service-providing institutions helps policy identify, learn from, and support organizations which are systematically effective. While some current attempts to use administrative data for the purpose of high stakes accountability has evoked negative political consequences, dialog around these problems have promoted conversations around the use of evaluation results towards lower-stakes "developmental evaluation" processes for improving social programming.[14]

---

[12] The National Longitudinal Surveys are maintained by NORC at the University of Chicago (http://www.norc.org/). For more information, see "National Longitudinal Surveys: A Program of the U.S. Bureau of Labor Statistics," https://www.nlsinfo.org/.

[13] See, for example, the descriptions of "Data Sources for Social Scientists," Cornell Institute for Social and Economic Researcher, http://www.ciser.cornell.edu/info/datasource.shtml.

[14] For a discussion and critique of test-based methods of accountability, see Eva L. Baker et al., "Problems with the Use of Student Test Scores to Evaluate Teachers," *Economic Policy Institute* Briefing Paper #278 (2010): 3. For a discussion of developmental evaluation, see , Michael Quinn Patton, "Evaluation for the Way We Work", *Nonprofit Quarterly* 13, no. 1 (2006): 28–30.

- **Breadth of measures** where, as suggested above, the establishment of human service agencies and the measures and processes that they systematically track are intentionally chosen to represent those which are of highest priorities to social policy. Linked administrative data from these agencies therefore represents a breadth of key populations and measures of interest.

- **Passive accumulation of data** means no additional cost of gathering data relative to what costs are already borne within the system. Because the costs in social science research are relatively low, given that the necessary materials are generally standard computing and statistical software licenses, with suitable data permissions it is possible for a significant amount of work to be leveraged on available data.[15] While new primary data collection initiatives still represent new costs, administrative data can help reduce planning costs by guiding that collection effort with a baseline set of measures for devising a sampling strategy.

- **The ability to interact with data sources** by planning new data collection is a final and transcendent feature of administrative data. The fact that administrative data represent information on identified individuals and institutions is what enables qualitative research to dovetail with quantitative research in the same setting.

B. *Implementing In Situ research*

With the foregoing description of traditional modes of social science inquiry and the features of administrative data as a resource for research, the implementation of *in situ* research is straightforward. The integration of each mode of integrations is cyclical.

First, quasi-experimental methods are used to examine the patterns in existing data. The consistent measurement of outcomes for individuals in many treatments, compared to many potential baseline controls allows, for example, examination of how individuals taking up a given program fare in contrast to comparable individuals taking up a separate program, or to individuals taking up no program at all.

---

[15] For example, at the time of writing this essay a search on Google Scholar for publications using the National Longitudinal Survey of Youth, 1979 cohort—which is freely downloadable—using the quoted phrase "National Longitudinal Survey of Youth" and "1979" yielded 10,700 results, including approximately 750 results in the calendar year 2013 alone.

Established statistical methods can be used to establish notions of "comparability" suitable for addressing the policy question, using other available information on individuals that relate to the outcome. For example, to answer how well different types of afterschool enrichment activities improve youth engagement in school, administrative data can establish comparability for youth based on their test scores, records of misconduct, and school attendance in periods before involvement in the afterschool program. In this way, quasi-experimental methods can deliver associations between social programs and desired outcomes where other explanations are accounted for.

The benefit of starting with quasi-experimental research is that the cost for doing this exploratory work is low. However, these methods are hampered by uncertain internal validity since, in considering the associations between program take-up and effectiveness, it is unclear whether the relationships are causal and, if they are, whether the mechanism of causality is well enough understood to expect similar benefits if policy makers adapt the program for use in other settings.

Whereas quantitative social scientists may leave off at this stage by admitting limitations to the investigation, within the *in situ* model, more information can be gathered to further understanding of the policy. Thus, what is otherwise an inductive exercise becomes an *abductive* one: we do not yet aim to inform policy with this evidence, but rather use this evidence to generate hypotheses on which to guide further investigation.

The second stage in the *in situ* model is to use qualitative and primary data collection methods to enrich the initial quantitative work, using the results from the first stage for leads on which policy efforts are either promising for understanding, or which raise unexpected questions which merit further inquiry. These efforts are designed to address what the first stage could not achieve by determining whether the patterns that were identified were truly causal in the way that was expected or were due to some other phenomenon, by pursuing understanding of causal mechanisms, and by suggesting new measures that may be systematically tracked for large populations in the future. In the example of seeking promising afterschool programs, this second qualitative stage may involve use of rubrics or observation protocols to measure and assess the practices of programs, interviews with program staff to understand what youth developmental factors may be missing from administrative data, and the pursuit or piloting of assessments of youth which may improve our understanding of key factors such as emotional wellbeing, social and academic habits, or behavioral skills.

The third stage in the *in situ* model uses randomized controlled trials to obtain authoritative evidence regarding the value of

promising new policies or extensions of old policies. Whereas RCTs are otherwise expensive given the need to recruit and follow-up study populations as well as administer surveys, their costs can be significantly decreased when embedded within an administrative data framework since identification and tracking of populations is already part of existing data systems. With this support, RCTs need only to ensure the implementation of the programming, maintain fidelity of the study (such as deterring contamination of the randomization), and bear the cost of assessments needed for the study that are beyond the administrative measures. And since the combination of quantitative and qualitative methods in the first two stages suggest and help confirm the promise of new policies and extensions, the cost of these RCTs will be borne with the expectation of a higher reward in terms of new policies which can be confidently promoted.

The fourth stage in the *in situ* model is simply to repeat the first three stages. The ability to sustain this process requires the establishment of a working relationship between policy makers and policy researchers. This requires capacity building on both sides, where researchers learn to speak the practical language and understand the agendas of policy makers, and policy makers in turn understand how to offer practical expertise to inform definitions of research questions, and to translate the guidance offered from research. Sustainability also requires trust among partners and funding from either public or private sources, which may be sourced through traditional grant-making processes for the start of a partnership, and ideally transfers to more stable line-item budgeting if this type of policy research and development proves to be effective.

## IV. CHALLENGES TO USE OF ADMINISTRATIVE DATA, AND LOOKING AHEAD

Many elements of the above proposal for integration of modes of inquiry and administrative data are idealized views of what can be possible with tools of the current Big Data era. This section combines a forward-looking perspective with discussion of the challenges to the above since, over time, some of these limitations may well be addressed through advances in technology and data governance which generally improve data usage.

### A. *Challenges Related to Data Quality and Sample Frames*

A first class of challenges to use of administrative data involves data quality, both in execution and in design. The quality of records in administrative systems can sometimes be poor, making it inaccurate,

unreliable, and/or hard to link to other sources. Some quality issues are due to agencies which are still in transition from pencil and paper or hand entry of data into simple storage formats like Excel spreadsheets to better database storage. Other quality issues may be due to the limited incentives for careful maintenance or documentation of data, in cases were data are not used for purposes of administration, accountability or research—cases which would otherwise require more careful review of records and more formal protocols for data entry. Current trends in transitions to more formal and easier-to-use database solutions, and towards more intentional usage of data may lead to significant improvements in data quality in coming years.

A more challenging class of data quality issues involves designs of the sampling frame. Administrative records are generally only reflective of populations served by public or non-profit institutions. Individuals using private services—such as attending fee-based preschool, or attending parochial schools—or who are independent of public services, or who are undocumented immigrants are by definition not included in public agency records. Because this lack of representation is not a technological matter, these issues will not be addressed by coming advancements in data tracking, and instead require either legislation or reconceptualization of models of data tracking.

Finally, there are some matters of logistics and feasibility that limit what data can be systematically represented in data systems. Some nuanced measures that are relevant to effective programming may not be able to be assessed, or at least may not be systematically comparable for all individuals or programs across a given jurisdiction. Examples may include measures of individual temperament or socioemotional disposition, or measures of program efforts such as the interpersonal manner of case workers or application of teaching techniques of teachers. As discussed above, acknowledgment of these limitations is rationale for the use of qualitative methods in the *in situ* model.

### B. *Limitations to Access and Usage of Administrative Records*

Because of the matter of ensuring privacy and well-being of individuals, the advantages of working with identifiable and sensitive data require equally strong legal protections against misuse or even frivolous research which does not provide adequate benefit to the studied populations. Because these considerations are not new, legal statutes and Institutional Review Boards have been established to provide necessary protections to limit and review data use for research.

However, a better balance arguably needs to be identified to fit different use cases for administrative data. In cases where only exploratory, quasi-experimental data work are called for, guidelines for suitable de-identification of administrative data could be established to help administrators benefit from the potential boon of research by making data available to researchers in a way similar to the National Longitudinal Surveys. One such effort—the University of Chicago Pathways to Adulthood Data Enclave—is currently underway as a collaboration between researchers at the University of Chicago and Chapin Hall.[16] In other cases where identification of individuals or institutions is necessary for the research, the transparency and speed for researchers to obtain the necessary clearance can be promoted by improved communications about applying for these permissions and streamlining data sharing protocols between the multiple agencies whose data would be involved and home institutions of the researchers. The Actionable Intelligence for Social Policy (AISP) initiative at the University of Pennsylvania has produced several white papers to improve awareness of the considerations in promoting models of this type of partnership.[17]

In a recent *Atlantic* article titled "Can Government Play Moneyball?", John Bridgeland and Peter Orszag's rough calculations suggest that "less than $1 out of every $100 of [federal] government spending is backed by even the most basic evidence that the money is being spent wisely."[18] The hope is that with the data and methodological resources currently available and still improving, at least a "10x" extension of the impact of the Perry Preschool Project study will be possible for all government social policy.

---

[16] See the following link for an abstract of the project: http://successfulpathways.uchicago.edu/page/derek-neal-robert-george-university-chicago-pathways-adulthood-data-enclave-uc-pade-250k.

[17] "Actionable Intelligence for Social Policy," http://www.aisp.upenn.edu/.

[18] John Bridgeland and Peter Orszag, "Can Government Play Moneyball?" *The Atlantic*, June 19, 2013.

## REFERENCES

Anderson, M. L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103, no.484.

Baker, E. et al. 2010. "Problems with the Use of Student Test Scores to Evaluate Teachers." *Economic Policy Institute Briefing Paper #278*.

Baker, E. L., et al. 2010. "Problems with the Use of Student Test Scores to Evaluate Teachers." *Economic Policy Institute Briefing Paper* #278: 3

Bridgeland, J., and Peter Orszag. 2013. "Can Government Play Moneyball?" *The Atlantic,* June, 19.

"Data Sources for Social Scientists," *Cornell Institute for Social and Economic Researcher*. http://www.ciser.cornell.edu/ASPs/datasource.asp.

Dehejia, R., and SadekWahba. 2002. "Propensity Score Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84. no.1: 151–161.

Morgan, D. L. 2007. "Paradigms Lost and Pragmatism Regained. Methodological Implications of Combining Qualitative and Quantitative Methods." *Journal of Mixed Methods Research* 1, no. 1: 48-49.

Neal, D., and Robert George. "The University of Chicago Pathways to Adulthood Data Enclave (UC-PADE)." *The University of Chicago*. http://successfulpathways.uchicago.edu/page/derek-neal-robert-george-university-chicago-pathways-adulthood-data-enclave-uc-pade-250k.

"Network," *Actionable Intelligence for Social Policy*. *http://www.aisp.upenn.edu/network/*.

Heckman, J. J. 2005. "The Scientific Model of Causality." *Sociological Methodology* 35, no. 1: 1–97.

Heckman, J. J. et al. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66, no. 5: 1017–1098.

Heckman, J. J. et al. 2010. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94, no. 1-2: 114–128. doi:10.1016/j.jpubeco.2009.11.001.

Heckman, J. et al. 2010. "Analyzing Social Experiments as Implemented: A Reanalysis of the HighScope Perry Preschool Program." *Quantitative Economics* 1, no. 1: 1–46.

Hotz, V. J. et al. 1998. "Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development." *Report of the Advisory Panel on Research Uses of Administrative Data of the Northwestern University/University of Chicago Joint Center for Poverty Research*.

LaLonde, R. J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76, no.4: 604–620.

Lazer, D. et al. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343, no. 6176: 1203–1205. doi:10.1126/science.1248506.

Morgan, D. L. 2006. "Paradigms Lost and Pragmatism Regained. Methodological Implications of Combining Qualitative and Quantitative Methods." *Journal of Mixed Methods Research* 1 (1) (2007): 48–76. Patton, M. Q.. Evaluation for the Way we Work. *Nonprofit Quarterly* 13, no.1: 28–33.

"National Longitudinal Survey of Youth" *Norc at the University of Chicago. http://www.norc.org/Research/Projects/ Pages/national-longitudinal-survey-of-youth-1997.aspx*.

Patton, M.Q. 2006. "Evaluation for the Way we Work", *Nonprofit Quarterly* 13, no. 1: 28–30.

Prashant, K. 2011. *An Overview of Architectures and Techniques for Integrated Data Systems Implementation*. Retrieved from http://www.sp2.upenn.edu/aisp_test/wp-

content/uploads/2012/12/0033_12_SP2_Architectures_Techniques_Data_Systems_000.pdf.

Shadish, W. R., Thomas D. Cook, and Donald T. Campbell. 2005. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin, 2002.Smith, J. A. and Petra E. Todd. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125, no. 1-2: 305–353.

Stiles, P. G. and John Petrila. 2011. "Research and Confidentiality: Legal issues and Risk Management Strategies." *Psychology, Public Policy, and Law* 17, no. 3: 333-356.

Todd, P. and Kenneth I. Wolpin. 2006. "Assessing the Impact of a School Subsidy Program in Mexico Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling." *American Economic Review* 96, no. 5: 1384–1417.