Using Administrative and Survey Data to Build Evidence[1]

I. Executive Summary

This white paper describes how various forms of evidence rely on survey and administrative data to influence a wide variety of policy and program decisions.

Many different types of evidence are vital to the formulation and development of effective policies and programs, including the activities outlined in the authorizing legislation of the Commission on Evidence-Based Policymaking ("the Commission"). Broadly speaking, these types of evidence can be grouped into two categories:[2,3]

- Foundational Evidence, which includes aggregate indicators, population descriptions, trends and correlations, and estimated specific treatment effects; and
- Policy-Specific Evidence, which includes performance measures; process evaluations; implementation evaluations; impact evaluations (using experimental/randomized control trials or quasi-experimental methods); and cost, benefit-cost, cost-effectiveness, and Regulatory Impact analyses.

Collectively these make up a *portfolio of evidence*, each component of which addresses specific features and characteristics of an issue. A strong portfolio of evidence can provide valuable insights into key policy and programmatic decisions. There is an association between each of these types of evidence, and the types of administrative and survey data that are best suited to generating that evidence. While not uniform, these data fall into four broad categories:

- Large-scale administrative data, which contain comprehensive information on the population or a key sub-population (e.g., Internal Revenue Service (IRS) tax data);
- Program-specific administrative data, which generally cover much smaller sections of the population (e.g., Second Chance Act grant data);
- General purpose survey data (e.g., the Current Population Survey); and
- Study-specific survey data (e.g., the Moving to Opportunity demonstration survey data).[4]

Over the last several years, the Federal Government has emphasized the need for building evidence and embedding its use into a wide range of policy and program formation. In order to determine whether a particular evidence-based program structure is appropriate for a given program area, the degree of consistency in program implementation, the breadth of the program area, and the quality and nature of

[1] This white paper is intended to provide the Commission on Evidence-Based Policymaking with background information on topics relevant to the Commission's work. The paper was prepared by staff from OMB, with assistance from staff at other Federal agencies.

[2] These categories of evidence and the range of components within them are interrelated and may overlap. They complement, inform, and rely on each other, often directing each other's further development, with some studies containing elements of more than one type of evidence.

[3] To ease exposition, this paper excludes several valuable types of data that can be used to inform policymaking, such as ethnographic and qualitative data and experimental data collected in a lab setting. One particularly promising source of evidence is observational data such as environmental information collected by satellites and traffic pattern data captured by sensors. As the Commission's focus is on administrative and survey data, we have not included these data in our discussion; however, they could be included in future discussions.

[4] National Bureau of Economic Research. *Moving to Opportunity for Fair Housing Demonstration Program: Survey Instruments.* http://www.nber.org/mtopublic/instruments.html.

the possible evidence base all must be considered. The availability of high-quality administrative and survey data is an especially powerful tool to build portfolios of evidence, support programmatic changes, and identify where new data are needed for future decisions.

Improving the quality of and access to data is likely to help translation of evidence into decision-making, regardless of the program structures and evidence base in a given policy area. A proliferation in data generation promises to continue for some time, and new technologies are providing the ability to utilize it. This confluence of factors presents an opportunity for the government to make more systematic and effective use of data and to develop and promote the most effective public policies.

The next section outlines how this paper relates to the Commission's mission. The ensuing sections describe different data types and their potential strengths and weaknesses, different components of an evidence portfolio and how they relate to these types of data, and a series of noteworthy case studies. The case studies illustrate how administrative and survey data can be used to generate a portfolio of evidence and influence program design across a variety of programs (e.g., programs directly administered by the Federal Government or formula grant programs) and policy areas (e.g., higher education or juvenile justice). While these case studies draw on a breadth of data and types of evidence, many important pieces of evidence are omitted in each of the examples by necessity. The conclusion identifies key takeaways. The appendix provides examples of existing successes and promising ongoing work.

II. Background

The Commission's authorizing legislation directs the Commission to consider how administrative and survey data "may be integrated and made available to facilitate program evaluation, continuous improvement, policy-relevant research, and cost-benefit analyses by qualified researchers and institutions."[5] The legislation also directs the Commission to consider "how best to incorporate outcomes measurement, institutionalize randomized controlled trials, and rigorous impact analysis into program design"[6] and "how data and results of research can be used to inform program administrators and policymakers to improve program design."[7] This paper suggests an evidence framework for the Commission to consider as it examines these broad and interrelated charges and activities.

Making better use of the administrative and survey data that the government already collects is one promising strategy to facilitate building a high-quality portfolio of evidence at a reasonable cost.

III. Data Types

Consistent with the Commission's charge, this paper focuses on data originating from administrative and survey sources. Administrative and survey data are much more useful for evidence-building purposes when accessible to researchers, evaluators, and statisticians at the individual level[8] because they can be linked to other data sources and because they can be analyzed in detail. Without individual-level data, many linkages are infeasible and analysis is limited. Personally-identifiable individual-level data are tightly controlled and not publicly released as a part of these evidence-building activities.[9]

---

[5] See Pub. L. 114-140 § 4(a)(1), 130 Stat. 317, 318 (2016).

[6] See Pub. L. 114-140 § 4(a)(3), 130 Stat. 317, 318 (2016).

[7] See Pub. L. 114-140 § 4(b)(2)(I), 130 Stat. 317, 319 (2016).

[8] This includes both person-level and establishment-level data.

[9] See Pub. L. 107-347 §§ 502(5)(A), (9)(A), 116 Stat. 2899, 2962-63 (2002) and the discussion of the use of data for statistical purposes in Office of Management and Budget Memorandum M-14-06, "Guidance for Providing and Using Administrative Data for Statistical Purposes," available at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

Administrative data are data collected by government entities for program administration, regulatory, or law enforcement purposes. These data are usually collected for the full universe of individuals, businesses, or communities affected by a particular program or regulation. They may also include financial and other data associated with program operations. Federal and state administrative data include rich information on labor market outcomes, health care, criminal justice, housing, pollution levels, and other important topics in addition to program participation. Since administrative data are collected to meet the needs of the relevant program or regulation—not the needs of research design—they will often lack information important for a given evaluation or other statistical use, such as demographic details needed to understand how policies and programs affect different groups within the population. In addition, it may be costly or even impossible to make administrative data usable for evidence-building purposes, especially if the original data are incomplete, inconsistent, invalid (i.e., they do not represent what they are supposed to represent), or poorly documented.[10]

Administrative data may be generated and maintained at the Federal, state or territorial, local, tribal, or non-governmental (e.g., grant recipient) level. In some cases data may be maintained at multiple levels, although for different purposes. This variability may complicate efforts to access particular datasets, since different data owners have different processes, policies, and orientations. The end result may be that while, in theory, data are available on the universe of participants, in practice, only a subset of those data may be accessed. They may also be collected under agreements with program participants or partners that limit their use to certain purposes to preserve individual privacy, business confidentiality, or other interests. There may be great variability in data quality between and within administrative datasets. For example, there may not be standard variables, documentation, data collection or input methods, data exchange standards, or data validation and cleaning processes associated with a particular administrative dataset. Office of Management and Budget (OMB) Memorandum M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes*, promotes increased agency use of administrative data for evidence-building and improved data stewardship practices designed to enhance privacy and data quality.[11]

In contrast to administrative data, the original goal of ***surveys*** is to produce certain statistical measures of specific outcomes of interest. The design of surveys begins with this goal in mind. For example, surveys measure population characteristics of interest, or they measure specific outcomes directly connected to policy questions. They are also a key evidence source for evaluations, performance measurement, and research. A major advantage of surveys (and a corresponding limitation of administrative data) is the ability of research or survey questions to drive the design, rather than being limited to existing data in what questions can be answered. Participation in individual and household surveys is mostly voluntary, and most provide a pledge of confidentiality to encourage participation and reliable responses. Importantly, surveys allow researchers to measure constructs generally not available

[10] For a more extensive discussion of the advantages and disadvantages of administrative and survey data, see:

Blank, R. M., Charles, K. K., & Sallee, J. M. (2009). A Cautionary Tale about the Use of Administrative Data: Evidence from Age of Marriage Laws. *American Economic Journal: Applied Economics*, 1(2): pp. 128-49

Prell, M., Bradher-Fredrick, H., Comisarow, C., Cornman, S., Cox, C., Denbaly, M., et al. (2009). *Working Paper: Profiles in Success of Statistical Uses of Administrative Data.* Federal Committee on Statistical Methodology.

National Research Council. (2009). *Reengineering the Survey of Income and Program Participation*: *Panel on the Census Bureau's Reengineered Survey of Income and Program Participation* (C. F. Citro & J. K. Scholz, Eds.). Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press.

Prewitt, K. (2010). Science Starts Not after Measurement, but with Measurement. *The Annals of the American Academy of Political and Social Science*, 631(1), pp. 7-16.

[11] Office of Management and Budget Memorandum M-14-06, "Guidance for Providing and Using Administrative Data for Statistical Purposes," available at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

in administrative data, such as opinions, experiences, and information about sensitive behaviors from both program participants and, critically, non-participants as well.

One intrinsic feature of surveys is that they require locating sampled respondents, ensuring participation in the survey, and constructing and administering the survey itself. Since each of these activities is expensive, evaluations, performance measurement, and other research based on surveys typically draw on small samples. Sometimes, despite careful planning, sample sizes are not large enough for the resulting studies to have sufficient statistical power to reliably detect policy-relevant effects of programs, even when these effects exist. Survey data can also be prone to the limitations inherent in all data gathering, such as response bias and other problems related to low respondent cooperation, recall errors, errors introduced by interviewers, and measurement error.

There are at least four categories of survey and administrative data. All are commonly used throughout Federal, state, and local governments:[12]

1. General Purpose Survey Data are typically collected through the Federal Statistical System,[13] are implemented solely for statistical purposes, and generate information about broad population characteristics. They tend to cover large, but not universal, samples of the populations of interest. In addition, they tend to be well *curated*: agencies collecting survey data often anticipate wide and varied uses of the data, and facilitate such uses by maintaining standards of data quality and documentation. Two subcategories are notable:

    - Large Cross-Sectional Survey Data provide representative estimates for large populations of people or organizations. They typically contain a large number of variables, enabling comparisons among narrowly-defined subpopulations. The surveys are often repeated with a consistent structure, yielding consistent datasets that make estimation of aggregate trends possible.

    - Large Longitudinal and Panel Survey Data follow samples over time, taking multiple collections from sample members. They usually have smaller samples than their cross-sectional counterparts, but the multiple observations of each member provide a powerful tool for studying changes in individual members over time (e.g., before and after a policy is enacted).

2. Study-Specific Survey Data are typically collected and associated with a specific evaluation or study, and may be started and developed either directly, through agency evaluation functions, or indirectly by a funding recipient. They have great variability in size, scope, and consistency. They may or may not include questions that are designed to correspond to questions available in other data sources (e.g., general purpose surveys or other study-specific surveys). While study-specific surveys tend to include tight controls on measurement quality, the curation of the

[12] Note that this is not a comprehensive listing of all data types. To ease the exposition, we have not included several valuable types of data that can be used to inform policy-making, such as ethnographic and qualitative data, and experimental data collected in a lab setting. One particularly promising source of evidence is observational data such as environmental information collected by satellites and traffic pattern data captured by sensors. As the Commission's focus is on administrative and survey data, we have not included these data in our discussion; however, they could be included in future discussions.

[13] In its Annual Report to Congress on Statistical Programs of the U.S. Government for Fiscal Year 2015, OMB defines the Federal Statistical System as the 13 principal statistical agencies and an additional approximately 130 other statistical programs housed within agencies across the Federal Government that support program planning and evaluation functions, or that are an outgrowth of administrative responsibilities. OMB, Statistical Programs of the U.S. Government Fiscal Year 2015 Report, available at https://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2015.pdf

data may be variable. Crucially, they may or may not be available for re-use in other applications. However, they can often:

- Fill in gaps in available administrative or general purpose survey data,
- Serve as a vehicle for gathering similar data for non-program participants,
- Tie to the specific outcomes of interest of the evaluation, and
- Be matched to large-scale administrative datasets to enhance the utility of both the survey and administrative data.

3. Large-Scale (Wide-Coverage) Administrative Data include some sources that have near-universal coverage of the population, key sub-populations, or entities—a scale that cannot generally be matched by survey collections. This makes them very valuable for several applications, such as providing sampling frames for surveys, supporting small-area estimation and non-response adjustments, and matching to other datasets to connect later outcomes to earlier interventions. These data are particularly promising for research, evaluation, and statistical purposes because they cover such a large portion of the population and contain valuable outcome information that is of interest to multiple policy and program areas. But in order to be useful in creating evidence, they typically require a great deal of curation, often by a statistical or analysis team such as the Census Bureau, the Statistics of Income Division at IRS, a Federal evaluation office, or an individual research team.

4. Program-Specific Administrative Data reflect the data collected as particular programs are carried out. They often contain rich information about program operations, receipt of services by program participants, and indicators of participant outcomes (during program enrollment, but typically not after). They are key resources for program evaluation, but they do not operate on the near-universal scale of large-scale administrative datasets. Like large-scale administrative data, they often require significant additional curation to be analytically useful. These data often do not cover long-term outcomes for program participants or contain complete information about those who drop out of services. They also do not include information on individuals not participating in the program who might form a comparison group for evaluation purposes. Also, programs with complex eligibility or ranking mechanisms may not retain information on key variables used to make selection decisions, which is useful for evaluating voluntary program participation decisions.

IV. Components of the Evidence Portfolio

Evidence encompasses a broad range of information that employs the different data sources and measurement activities discussed above that can serve to provide insight into policy and programmatic decisions.[14] Examples below are divided into two broad categories: foundational evidence and policy-specific evidence. Foundational evidence comprises a range of components that establish our understanding of the social, economic, behavioral, and other conditions that form the context for programs and policies. Policy-specific evidence is used to provide metrics and evaluate policy or program effectiveness. These categories of evidence and the range of components within them are interrelated and often interdependent. They complement, inform, and rely on each other, often

[14] See the definitions of "evidence" and "intended use" and their role in the performance management process in OMB Circular A-11 (2015), Section 200, page 200-213, available at: https://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/s200.pdf

directing each other's further development. Some studies contain elements of more than one type of evidence. When taken together, sets of information compose a portfolio of evidence.

Components of Foundational Evidence

1. Aggregate Indicators[15] represent a broad, general level of evidence, and act as high-level signals of relevant issues and as a barometer of changing conditions. Some examples include national- and state-level health insurance coverage, crime, school completion, air travel, and residential construction, as well as other designated Principal Federal Economic Indicators (PFEIs).[16]

2. Population Descriptions, Trends, and Correlations are a specific type of aggregate indicators focused on the populations or entities that policies may affect. This type of evidence includes a wide variety of analyses ranging from calculations of averages for selected subpopulations and related cross-tabulations, to analyses of variance across several related measures, to multiple regression analyses and other multi-dimensional decompositions of trends. The results can provide indications about whether, how, and for whom various policies might be needed and be effective, although analyses in this evidence component usually cannot draw firm causal conclusions.

3. Estimated Effects of Specific Treatments, not Policy-Specific are a third type of foundational evidence that includes a diverse range of studies that employ quasi-experimental and (less often) experimental methods to support causal conclusions, but without reference to a particular policy or program. A great deal of high-quality work—originating in many social science fields and employing many different analytical tools—falls into this category. Such studies develop a basic understanding of how and why different outcomes occur. The implications for policy may be immediate or indirect. These would include, for example, studies to estimate the price elasticity of the supply of labor, or how test performance is related to anxiety. The word "treatment" refers to any causal factor that might influence an outcome of interest. These studies result in specific findings that are not policy-specific, because the treatments under study are not attached to any particular policy.

Components of Policy-Specific Evidence

1. Performance Metrics, such as data from program monitoring, capture a program's operational aspects. When a relationship can be established between these metrics with program delivery fidelity (through an implementation evaluation), or with effectiveness and outcomes (through impact evaluations), they can be useful in providing a near-term feedback loop on program performance, as well as a tool to assess and communicate to policymakers the potential impact of budgetary or operational changes.

2. Implementation (or Retrospective) Evaluations can complement assessments of impacts on outcomes by measuring how well programs or aspects of programs deliver services relative to program design, professional standards, or regulatory requirements. Implementation evaluations can also assess how variations in program implementation or design affect high-quality service delivery.

---

[15] As an example of the important role of aggregate indicators, see Chapter 5 (Social Indicators) of the *Analytical Perspectives* volume of the President's Fiscal Year 2017 Budget, available at: https://www.whitehouse.gov/sites/default/files/omb/budget/fy2017/assets/ap_5_indicators.pdf.

[16] For a listing of the PFEIs and a schedule of their release dates, see OMB, Schedule of Release Dates for Principal Federal Economic Indicators for 2016, available at https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/pfei-schedule-of-release-dates-2016.pdf.

Implementation evaluations and performance metrics may measure the same things. However, implementation evaluations typically measure in greater depth, often collecting data specifically for evaluation purposes, and typically just one time; in contrast, performance metrics typically rely on less detailed data collected for administrative purposes and collected over time.

3. Estimated Effects of Specific Treatments, Policy-Specific (i.e., Impact Evaluations) assess how the intervention being evaluated affects outcomes, whether these effects are intended or unintended. The proper analysis of impact requires a counterfactual of what those outcomes would have been in the absence of the intervention.

    - Experimental/Randomized Control Trials exploit randomization between "treatment" and "control" groups to generate statistically unbiased effects of an intervention. Well-designed and well-executed randomized experiments with sufficient samples are the strongest design for determining the effects of a given policy or program. However, quasi-experiments can sometimes allow for larger, more diverse samples with less sampling error and more straightforward generalizability.

    - Quasi-Experimental Evaluations exploit variation across factors such as geography, time, or participants to measure the effects of a policy intervention or "treatment" on populations of interest. Quasi-experiments may employ statistical methods such as propensity score matching to construct a comparison group not receiving a policy intervention that is similar in key dimensions to the treatment group receiving it. That helps disentangle the causal effects of the policy or program under study from the impact on the outcome variable of factors that determine whether someone receives the policy intervention. Quasi-experimental evaluations often do not include randomization, and therefore need to be well designed to minimize potential sources of bias. However, they can be especially useful in circumstances where program design (such as statutory entitlement to benefits) can make randomization infeasible, and in ex-post impact studies of ongoing programs where enrollment was not determined randomly. Large-scale administrative data files sometimes make possible combinations of quasi-experiments that generate far more rigorous evidence than is possible with smaller datasets.

4. Cost, Benefit-Cost, Cost-Effectiveness, and Regulatory Impact Analyses assess costs or compare costs with estimates of benefits or effectiveness. Done well, these analyses allow for comparisons of costs and/or effectiveness across interventions. These analyses are often integrated into policy-specific evidence components and can depend on both administrative and survey data.

Table 1 depicts the relationship between data types and evidence components. For each evidence component listed in the first column, the second column shows the data types that are typically used to generate it. All data types are important contributors to the evidence portfolio, as illustrated in the third column.

Table 1. Relationship between Evidence Components and the Underlying Data Types

Evidence Component	Underlying Data Types	Policy-Related Example
Foundational		*Unemployment*
Aggregate Indicators	Large Cross-Sectional Survey Large-Scale Administrative	The unemployment rate is estimated via the Current Population Survey at the national, state and local levels, and broken out by age, sex, race, disability status, and other characteristics.
Population Descriptions, Trends and Correlations	Cross-Sectional Survey Large Longitudinal Survey Large-Scale Administrative	Analyses show that there are demographic differences in economic, health, and social correlates of unemployment. Long-run trends show a widening divide in unemployment rates between less and more educated workers as well as workers with and without disabilities. Cyclical downturns also seem to hit less educated workers and workers with disabilities especially hard.
Estimated Effects of Specific Treatments, not Policy-Specific	Cross-Sectional Survey Longitudinal Survey Study-Specific Survey Large-Scale Administrative Program-Specific Administrative	Studies of workers who are displaced from their jobs because of layoffs or plant closings demonstrate that the job skills of such workers often include significant components that are specific to the industry but not to the employer.[1]
Policy-Specific		*Job Training*
Performance Metrics	Program-Specific Administrative Large-Scale Administrative Study-Specific Survey	All Federal job training programs authorized under the Workforce Investment Act (WIA) have used a common set of performance metrics since 2005, including measures of employment and earnings. At the time of this writing, the WIA has been reauthorized and revised under the Workforce Innovation and Opportunity Act, and new performance metrics are being established to reflect the goals of the Act (see page 11). These performance metrics are generated by matching program administrative data on WIA participants to Quarterly Wage records—a large-scale administrative dataset.
Implementation Evaluations	Program-Specific Administrative Study-Specific Survey	The new Workforce Innovation and Opportunity Act will be the focus of a multi-data implementation analysis to assess specific management decisions and outcomes at the state and local level.
Estimated Effects of Specific Treatments, Policy-Specific:[2] Quasi-Experimental	Longitudinal Survey Study-Specific Survey Large-Scale Administrative Program-Specific Administrative	Using linked program data and wage records on participants in sector-focused programs found that participants had higher earnings and employment than a matched sample of other career center participants.[3]
Estimated Effects of Specific Treatments, Policy-Specific:[2] Experimental/ Randomized Control Trial	Study-Specific Survey Program-Specific Administrative Large-Scale Administrative	Using linked program data and wage records, evaluations from multiple sites found that providing re-employment services to Unemployment Insurance claimants speeds reemployment and reduces Unemployment Insurance benefits.[4] The Department of Labor (DOL) conducted a Job Corps evaluation that also demonstrates how administrative data allows for cost-effective long-term follow-up.[5]
Cost, Benefit-Cost, Cost-Effectiveness, and Regulatory Impact Analyses	Longitudinal Survey Cross-Sectional Survey Study-Specific Survey Large-Scale Administrative Program-Specific Administrative	DOL conducted an effectiveness and cost-benefit analysis of the Registered Apprenticeship (RA) program in 10 states to assess the effectiveness of RA with regard to the earnings and net benefits received by apprentices.[6]

[1] See, for example, Neal, D. (1995) Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics*, 13(4): 653-677.

[2] Otherwise known as impact evaluations.

[3] Gasper, J., & Henderson, K. (2014). *Sector-Focused Career Centers Evaluation: Effects on Employment and Earnings after One Year*. Prepared for NYC Center for Economic Opportunity by Westat, available at http://www.nyc.gov/html/ceo/downloads/pdf/CEO-Sector_Based_Approaches_Evaluation_Report-2014_final.pdf.

[4] Michaelides, M., Benus, J., Poe-Yamagata, E., Shen, T., Bill, N., Carrington, H., & Tirumalasetti, D. (2012). *Impact of the Reemployment and Eligibility Assessment (REA) Initiative*. Prepared for the U.S. Department of Labor by IMPAQ International.

[5] Burghardt, P., McConnell, S., Schochet, P. Z. (2006). *National Job Corps Study and Longer-Term Follow-up Study: Impact and Benefit-Cost Findings Using Survey and Summary Earnings Records Data*. Prepared for the U.S. Department of Labor, Employment and Training Administration by Mathematica Policy Research.

[6] Kleinman, R, Liu, A., Mastri, A., Reed, D., Reed, D., Sattar, S., & Ziegler, J. (2012). *An Effectiveness Assessment and Cost-Benefit Analysis of Registered Apprenticeship in 10 States*. Prepared for the U.S. Department of Labor, Employment and Training Administration by Mathematica Policy Research (which uses a variety of Federal and state administrative data, including Unemployment Insurance wage records, and data from the Current Population Survey).

V. Building the Evidence Portfolio: Examples

Evidence is currently incorporated into policy decision-making, program design, and program implementation in multiple ways. The following are examples from several policy areas illustrating how evidence from the entire portfolio can be integrated and used to inform different types of decision-making. Many of these examples meet the quality standards of Federal "what works" repositories,[17] such as the Department of Education's *What Works Clearinghouse*. The examples purposely cover a variety of programs (e.g., programs directly administered by the Federal Government or formula grant programs) and policy areas (e.g., higher education or juvenile justice), but are certainly not exhaustive. For instance, the education example focuses on the Free Application for Federal Student Aid (FAFSA). Implications for other higher education programs are not discussed, so excellent examples of seminal evidence in the K-12 domain, like the Chetty et al.[18] teacher value-added study, are not included (even though they also highlight the potential of linking large-scale administrative datasets to enable studies that can affect state and Federal programs). Similarly, while these examples draw on a breadth of data and types of evidence, they are not meant to be exhaustive literature reviews. By necessity, many important pieces of evidence are omitted in each of the examples. Given the central role of survey and administrative data in the Commission's work, we highlight the role of these data in both generating evidence and integrating the evidence into policy and program decision-making.

1. *Job Training Programs*

Evidence on job training programs and the labor market is broad, deep, and stretches back decades. Labor market statistics,[19] typically from Bureau of Labor Statistics (BLS) surveys and sometimes supplemented with administrative data, form the foundation of the evidence portfolio and create a foundation for research and policy-making. Within the workforce development field, there is also a rich history of Federal and state partnerships to combine administrative and survey data to generate new evidence. The Longitudinal Employer-Household Dynamics (LEHD) program, in particular, is one such example in which the Census Bureau combines Federal and state administrative and survey data on employers and employees with data from programs, to produce new, cost-effective, public-use information that helps to fill critical gaps in knowledge and provide indicators needed by state and local authorities.

Under the Local Employment Dynamics Partnership, states agree to share Unemployment Insurance earnings data and the Quarterly Census of Employment and Wages data with the Census Bureau. The LEHD program combines these administrative data, additional administrative data, and data from censuses and surveys. From these data, the LEHD program creates statistics on employment, earnings, and job flows at detailed levels of geography and industry, and for different demographic groups. In

[17] "What works" repositories, sometimes referred to as research clearinghouses, synthesize evaluation findings in ways that make research useful to decision-makers, researchers, and practitioners in the field. They also make evaluation results easily accessible to the public and improve the transparency of evaluation results. Information in the repositories also indicates the implementation contexts of programs and strategies evaluated, and areas where more innovation or more evaluation is needed.

[18] Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), pp. 2633-2679.

[19] These statistics include the national unemployment rate, alternative unemployment measures, the distribution of unemployment spells, the variation in unemployment rates between local areas, business dynamics, and labor turnover. Some of the data underlying these statistics are generated through agreements with states.

addition, the LEHD program uses these data to create partially synthetic data[20] on workers' residential patterns.

The evidence portfolio in the area of workforce development and job training covers a range of issues and programs, and has resulted in many studies that influence program and policy directions, including studies on the returns to education and training;[21] the importance of closely relating training to specific jobs and occupations;[22] the extent to which providing re-employment services to Unemployment Insurance claimants speeds reemployment and reduces benefits;[23] the importance of tailoring job training approaches to individual needs;[24] the advantages of work-based training,[25] particularly subsidized on-the-job training[26] and registered apprenticeships;[27] and the benefits of coordinating strategies for workers (or future workers) facing many barriers to work.[28] This portfolio covers a broad

[20] Synthetic data are a type of statistical disclosure control method whereby observed dataset values are replaced by artificial data values generated to be representative of the same target population as the observed dataset.

[21] See, for example:

Ewert, S., & Kominski, R. (2014). *Measuring Alternative Educational Credentials: 2012*. Washington, DC: U.S. Census Bureau (which uses data from the Survey of Income and Program Participation), available at https://www.census.gov/prod/2014pubs/p70-138.pdf.

Jacobson, L. S., LaLonde, R. J., & Sullivan, D.G. (2005). Estimating the Returns to Community College Schooling for Displaced Workers. *Journal of Econometrics*, 125(1-2) (which uses administrative earnings records and community college transcripts from Washington State).

Jepsen, C., Troske, K., & Coomes, P. (2009). *The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates*. University of Kentucky Center for Poverty Research Discussion Paper Number 2009-08 (which uses administrative data from the state of Kentucky).

[22] See, for example:

Maguire, S., Freely, J., Clymer, C., Conway, M., & Schwartz, D. (2010). *Tuning Into Local Labor Markets: Findings from the Sectoral Employment Impact Study*. Public/Private Ventures, Philadelphia, PA, available at http://ppv.issuelab.org/resources/5101/5101.pdf.

Gasper, J., & Henderson, K. (2014). *Sector-Focused Career Centers Evaluation: Effects on Employment and Earnings After One Year*. Prepared for the New York City Center for Economic Opportunity by Westat, available at http://www.nyc.gov/html/ceo/downloads/pdf/CEO-Sector_Based_Approaches_Evaluation_Report-2014_final.pdf.

[23] See, for example, Michaelides, M., Benus, J., Poe-Yamagata, E., Shen, T., Bill, N., Carrington, H., & Tirumalasetti, D. (2012). *Impact of the Reemployment and Eligibility Assessment (REA) Initiative*. IMPAQ International.

[24] For a review of evidence on job training, see U.S. Departments of Labor, Commerce, Education, and Health and Human Services (2014, July) *What Works in Job Training: A Synthesis of the Evidence*, available at https://www.dol.gov/asp/evaluation/jdt/jdt.pdf.

[25] See, for example, Lerman, R.I., McKernan, S., and Riegg, S. (2004). The Scope of Employer-Provided Training in the United States: Who, What, Where, and How Much? In *Job Training Policy in the United States*, (Christopher J. O'Leary, Robert A. Straits, and Stephen A. Wandner, Eds.) Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, pp. 211-244 (which draws on survey data from the National Employer Survey, the Survey of Employer-Provided Training, the National Household Education Survey, and the Survey of Income and Program Participation).

[26] See U.S. Departments of Labor, Commerce, Education and Health and Human Services, (2014, July) *What Works in Job Training: A Synthesis of the Evidence*, available at https://www.dol.gov/asp/evaluation/jdt/jdt.pdf.

[27] See, for example, Kleinman, R., Liu, A., Mastri, A., Reed, D., Reed, D., Sattar, S., & Ziegler, J. (2012). *An Effectiveness Assessment and Cost-Benefit Analysis of Registered Apprenticeship in 10 States.* Prepared for the U.S. Department of Labor, Employment and Training Administration (which uses a variety of Federal and state administrative data, including Unemployment Insurance wage records, and data from the Current Population Survey) by Mathematica Policy Research.

[28] See, for example, Bloom, D., and Michalopoulos, C. (2001). *How Welfare and Work Policies Affect Employment and Income: A Synthesis of Research*. New York: Manpower Demonstration Research Corporation (MDRC), available at http://www.mdrc.org/sites/default/files/full_393.pdf.

range of evidence, spanning from analysis of statistical series to specific impact evaluations of Federal job training programs.

This rich portfolio of evidence helped inform the development of the Workforce Innovation and Opportunity Act (WIOA), which reauthorized the Nation's employment, training, adult education, and vocational rehabilitation programs authorized under the WIA.[29] WIOA incorporates many evidence-informed strategies (and supports for states to implement evidence-informed strategies) throughout the statute.[30] For example, WIOA encourages states to adopt strategies that are tailored to a specific industry sector, encourages integration and coordination with apprenticeship programs, and makes it easier for states to support on-the-job training. Beyond the integration of these approaches, WIOA incorporates data usage and evidence-building in a variety of ways, at the Federal and state levels. WIOA requires the measurement of states' performance on a set of common indicators of performance for all of the core programs. The states' performance on these indicators will be determined, in part, using WIOA administrative data with Unemployment Insurance employment and wage data. WIOA also requires the targets for the indicators of performance to be set (and actual performance adjusted) using a statistical adjustment model, which itself incorporates a variety of administrative and statistical survey data on attributes like participant demographics and the state unemployment rate. WIOA requires states to implement high-quality evaluations of the core WIOA programs, and to cooperate with Federally-led evaluations of those programs, including the provision of administrative data.

2. *Education*

The Federal Government plays an active role in furthering the education of Americans, from early in their lives through pre-K programs, during their childhood through elementary and secondary education, and into adulthood through postsecondary student aid programs. Throughout these stages, data play an integral role in tracking needs and progress. In its Digest of Education Statistics, the National Center for Education Statistics (NCES) provides an annual encapsulation of key aggregate statistics such as family household characteristics, school enrollment rates, teacher employment, and state and local educational expenditures, as well as educational attainment rates in the population. Its National Postsecondary Student Aid Study, which integrates student aid administrative data with robust survey data on demographics and student experiences, is a primary source of information used by the Federal Government, researchers, and higher education associations to analyze student college financing and debt, and to inform public policy on programs, like the Pell Grant program and Stafford loans. And, in part as a result of Federal support to improve local education data, statewide longitudinal data systems recording the progress of children from kindergarten through college are being developed from administrative data sources. These resources have been used for influential research on topics ranging from disparities in educational outcomes by family income to the effects of universal pre-kindergarten, charter schools, intensive tutoring programs, and community college remediation programs.[31]

---

[29] Workforce Innovation and Opportunity Act (WIOA), 29 U.S.C. § 3101 note.

[30] See, for example, the DOL's overview of WIOA, available at https://www.doleta.gov/wioa/Overview.cfm.

[31] See, for example:

Papay, J. P., Murnane, R. J., & Willett, J. B. (2014). *Income-based Inequality in Educational Outcomes: Learning from State Longitudinal Data Systems*. NBER Working Paper No. 20802. Available at http://www.nber.org/papers/w20802.pdf.

Andrews, R. J., Jargowsky, P.,& Kuhne, K. (2012). *The Effects of Texas's Targeted Pre-Kindergarten Program on Academic Performance*. NBER Working Paper No. 18598. Available at http://www.nber.org/papers/w18598.pdf.

One particular example in which evidence from a wide range of these data and evidence types played a key role in improving services is in the simplification of the FAFSA process. The simplification of FAFSA is an example in which changes in the implementation of one process can have an effect on a range of programs. The FAFSA process is governed directly by the Department of Education (ED), according to statutorily prescribed parameters, so the Federal Government has somewhat direct control over changes to the FAFSA (subject to statutory and resource constraints).

Consequently, as evidence began to accumulate about the potential benefits of simplifying the FAFSA, ED was able to act. This evidence includes the large body of literature on socio-economic and demographic differences in college enrollment and the returns to education (using a variety of survey and administrative data); broad behavioral sciences research (often employing administrative data) demonstrating that making parts of the process, like questions, forms, access, and user experience, easier and more intuitive tends to increase adoption;[32] and studies of FAFSA simplification and supports, such as the H&R Block Study by Bettinger et al.[33] This study used a combination of tax return data from H&R Block, college enrollment and degree data from the National Student Clearinghouse, and student aid data from National Student Loan Data System at ED to show that making the financial aid application process simpler could increase rates of applying for aid and enrolling in college.

By highlighting the importance of simplification, the H&R Block study generated evidence that motivated ED to simplify the online FAFSA form by allowing applicants to skip questions that are not relevant to them and automatically retrieve needed tax information. These changes have helped reduce FAFSA completion time by two-thirds over the last eight years, to about 20 minutes. Researchers are continuing to study ways to improve the financial aid process. For example, Castleman and Page tested approaches to increase college persistence and found that sending students text messages with offers of assistance with the financial aid process, along with reminders of important deadlines and about maintaining satisfactory grades, can be effective in freshman to sophomore persistence at two-year colleges.[34]

3. *Home Visiting*

In addition to advancing educational opportunities and outcomes, the Federal Government has long taken an interest in safeguarding the health, positive development and well-being of young children. To foster coordination, collaboration, and integration of Federal data efforts related to child and family well-being, a working group of 23 Federal agencies, the Federal Interagency Forum on Child and Family

---

Abdulkadiroğlu, A., et al. (2011). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots. *The Quarterly Journal of Economics*, 126(2), pp. 699–748.

Fryer, R. G., Jr. (2014). Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments. *The Quarterly Journal of Economics*, 129(3), pp. 1355-1407.

Calcagno, J. C., & Long, B. T. (2008). *The Impact of Postsecondary Remediation Using a Regression Discontinuity Approach: Addressing Endogenous Sorting and Noncompliance.* The National Center for Postsecondary Research Working Paper. Available at http://www.postsecondaryresearch.org/i/a/document/8162_CalcagnoLongRevised.pdf.

[32] For a broad literature review touching on each of these literatures, see Page, L. C., & Scott-Clayton, J. (2015). *Improving College Access in the United States: Barriers and Policy Responses.* NBER Working Paper No. 21781, Available at http://www.nber.org/papers/w21781.pdf.

[33] Bettinger, E., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment. *Quarterly Journal of Economics*, 127(3), pp. 1205-1242.

[34] Castleman, B. L., & Page, L. C. (2014). *Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence.* EdPolicyWorks Working Paper Series No. 29. Charlottesville, VA: EdPolicyWorks. Available at http://curry.virginia.edu/uploads/resourceLibrary/29_Freshman_Year_Financial_Aid_Nudges.pdf.

Statistics (the Forum), was established in 1997. The Forum's annual report, *America's Children: Key National Indicators of Well-Being*, provides a summary of trends in national indicators of child well-being collected from 22 administrative and survey data sources from across government agencies.[35] In addition to providing statistics in an easy-to-use, non-technical format, the purpose of the report is to stimulate discussions among policymakers and the public, exchanges between data providers and policy communities, and improvements in Federal data on children and families. The report presents key indicators in seven domains: family and social environment, economic circumstances, health care, physical environment and safety, behavior, education, and health.

Social scientists have extensively studied the correlation between aspects of early child well-being and later life outcomes, using many data sources. Traditionally, some of the best information about these correlations has come from cross-sectional and longitudinal surveys, such as the National Health and Nutrition Examination Survey and the National Longitudinal Survey of Youth. A common finding is that the social and economic disadvantages of parents often transfer to their children. But only careful and detailed work using a variety of methods can disentangle the effects of particular childhood experiences, and identify the pathways by which a particular effect takes place. A vast number of studies have explored these processes in a long and broad literature.[36]

One particular set of lessons from these studies is that positive early childhood experiences can have a great impact on children. These lessons have spurred policies and programs aimed at addressing early childhood experiences, either by preventing childhood traumas (e.g., poverty/deprivation, abuse/neglect) or by encouraging protective factors (e.g., early education, positive parenting skills, physical and emotional health). Home visiting is one strategy to increase protective factors with a substantial evidence-base. Home visiting programs enroll women prenatally on a voluntary basis, and provide support services to families through the first few years of their child's life and sometimes through their child's entry into kindergarten. Nurses, social workers, educators, or other professionals or paraprofessionals deliver curriculum and supports to families through a range of home visiting models.

Rigorous evaluations of early childhood home visiting programs have indicated sustained, long-term effects on important outcomes such as improving school readiness, decreasing child maltreatment, and increasing family economic self-sufficiency. It is this evidence that supported the creation of the Maternal, Infant, and Early Childhood Home Visiting Program (MIECHV; Federal Home Visiting Program), one of the Administration's tiered evidence initiatives, jointly administered by the Health Resources and Services Administration and the Administration for Children and Families. By statute, the MIECHV program requires grantees to invest at least 75 percent of the funds in home visiting programs that through rigorous evidence have been shown to have impacts in one of eight outcome domains, with a portion of funds available for innovations that grantees are required to rigorously evaluate.[37] The

---

[35] Federal Interagency Forum on Child and Family Statistics. (2015). *America's Children: Key National Indicators of Well-Being*. Washington, D.C.: U.S. Government Printing Office. Available at http://www.childstats.gov/index.asp.

[36] See, for example:

Brooks-Gunn, J., & Duncan, G. (1997). The Effects of Poverty on Children. *The Future of Children* 7(2), pp. 55-71.

Duncan, G., & Magnuson, K. (2013). The Long Reach of Early Childhood Poverty. In Yeung, W., & Yap, M., (Eds.) *Economic Stress, Human Capital, and Families in Asia*. Netherlands: Springer, pp. 57-70.

Bornstein, M., & Bradley, R., Eds. (2014). *Socioeconomic Status, Parenting, and Child Development*. Routledge.

[37] See 42 U.S.C. § 711(d). For example, one state is using Medicaid data to create a propensity score matched comparison group to examine the impact of their promising approach, and another state is using internal administrative data to create a historical control to assess the impact of an enhancement of their home visiting program.

Department of Health and Human Services (HHS) Administration for Children and Families Office of Planning, Research and Evaluation (OPRE) has sponsored a thorough, transparent, and systematic review to determine which early childhood home visiting models qualify as evidence-based.[38] Many of the evidence-based models tested impacts, in part, using administrative data, particularly medical records and child welfare records.[39]

To continue to build the evidence-base, the statute mandated[40] a national evaluation of the MIECHV program, overseen by HHS. The Maternal, Infant and Early Childhood Home Visiting Evaluation (MIHOPE) is a random assignment evaluation of approximately 4,000 families across 12 states and 87 local agencies.[41] The MIHOPE evaluation is distinctive in its use of questionnaire data, in-person assessments, and administrative data to assess impacts. In addition to the study-specific survey data, MIHOPE will leverage administrative data on child welfare, health (i.e., Medicaid), and employment and earnings (National Directory of New Hires) to assess multiple outcomes. A companion study called MIHOPE Strong Start (MIHOPE-SS), in partnership with the Center for Medicare & Medicaid Innovation, is examining the impacts of early childhood home visiting on birth outcomes and infant health in those families who enroll in home visiting prenatally.[42] A brief baseline survey was necessary because little quality administrative data of baseline risk factors for premature birth existed; however, outcomes in MIHOPE-SS are all being assessed using administrative data, specifically vital statistics and Medicaid. Both of these evaluations are examples of the importance of collecting direct study-specific survey data on evaluation participants when necessary to effectively evaluate a program, and leveraging existing administrative data, where available and appropriate.

The use of data is integral to MIECHV design. The statute allowed the design of the program's performance measurement to complement the current state of the program. Initially, recognizing the need for grantees to design programs that reflected their unique needs, HHS allowed grantees flexibility to set their own performance measures to inform their unique program design and administration. As MIECHV has matured, HHS has proposed standardizing performance measures to tell the national story of the program. The performance measure data collection in states is creating administrative data that, if included in state early childhood integrated data systems, have the potential for long-term evaluation of outcomes, such as school readiness.[43]

---

[38] Avellar, S., Paulsell, D., Sama-Miller, E., Del Grosso, P., Akers, L., & Kleinman, R. (2014). Home Visiting Evidence of Effectiveness Review: Executive Summary. OPRE Report #2014-59, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, available at http://homvee.acf.hhs.gov/HomVEE_Executive_Summary_2014-59.pdf.

[39] For example, Durham Connects, Early Head Start, Healthy Families America, and Nurse Family Partnership.

[40] 42 U.S.C. § 711(g).

[41] 42 U.S.C. § 711(g). *Mother and Infant Home Visiting Program Evaluation (MIHOPE), 2011-2015.* Office of Planning, Research & Evaluation, Administration for Children and Families, available at http://www.acf.hhs.gov/programs/opre/research/project/maternal-infant-and-early-childhood-home-visiting-evaluation-mihope.

[42] *Mother and Infant Home Visiting Program Evaluation – Strong Start (MIHOPE-Strong Start), 2012-2016.* Office of Planning, Research & Evaluation, Administration for Children and Families, available at http://www.acf.hhs.gov/programs/opre/research/project/mother-and-infant-home-visiting-program-evaluation-strong-start-mihope-ss.

[43] Jordan, E., & King, C. (2015). Stacking the Blocks: A Look at Integrated Data Strategies. In *Rising to the Challenge: Building Effective Systems for Young Children and Families, a BUILD E-Book*. BUILD Initiative, available at http://www.childtrends.org/wp-content/uploads/2015/08/2015-35BuildChap7.pdf.

4. *Housing Vouchers*

There is a rich body of descriptive evidence on geographic differences in poverty, income, race, ethnicity, jobs, and other demographic/socio-economic characteristics, largely generated through surveys and (increasingly) the analysis of administrative data. Some of these data, such as state poverty rates and state unemployment rates, play a direct role in allocating Federal funding or determining eligibility for programs such as Medicaid or Unemployment Insurance. These data also have documented how disadvantages vary by geography, and a related broad literature has generated evidence on the impact of disadvantaged communities on social, economic, and health outcomes.

This policy area has several examples of studies that would not be possible without access to very large administrative datasets. In particular, Chetty and Hendren used large-scale administrative tax files to examine the effects of tax policy on intergenerational mobility, as well as the effects of neighborhoods on children's outcomes.[44] Armed with these rich data, the authors are able to show that when families move from one city to another, the future earnings (and other outcomes) of their children are substantially affected by the change in neighborhood quality, and that the size of the effect depended substantially on the age of the children at the time of the move. Although this study is neither an experimental study nor an evaluation of any program, the huge sample sizes and rich outcome data allow for a series of quasi-experiments that together generate strong evidence. Large-scale administrative data files open up a new world of opportunities for both randomized experiments and quasi-experiments.

Those same administrative tax files also played a pivotal role in moderating the current wisdom on how place may be associated with earnings and employment outcomes. A number of studies from the mid-2000s examined the Department of Housing and Urban Development's (HUD) "Moving to Opportunity" (MTO) experiment, which provided randomly selected and assigned families the opportunity to move from high-poverty housing projects to low-poverty neighborhoods. These studies consistently failed to observe significant economic impacts associated with the MTO program. When Chetty, Hendren, and Katz matched administrative tax data to the evaluation survey data from the MTO experiment, they identified that the MTO program had large, long-term economic effects not previously identified.[45] Specifically, when the data were stratified by age the authors found that children under 13 who "moved to opportunity" experienced statistically significant increases in future earnings. But, they also observed slightly negative effects on children older than 13, positing that the results may relate to a disruption effect associated with changing neighborhoods as adolescents. These types of follow-up studies that link administrative data to survey or administrative data from earlier experiments/pilots are a promising way to assess the difficult but important long-term effects of many government programs.

These findings, that "moving to opportunity" can have substantial effects on younger children living in high-poverty neighborhoods, have supported proposals to expand the housing voucher program. Other related research undertaken by HUD has looked at policy-specific factors and has played an important role in shaping housing voucher and related policies at HUD. For example, HUD is proposing to require agencies with concentrations of voucher holders in high-poverty neighborhoods to set fair market rents at the ZIP code level rather than metropolitan-wide; this change is based upon preliminary evidence that

---

[44] *See, e.g.*, Chetty, R., & Hendren, N. (2015). *The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates*. Cambridge, MA: Equality of Opportunity Project, Harvard University. Available at http://scholar.harvard.edu/files/hendren/files/nbhds_paper.pdf.

[45] Chetty, R., Hendren, N., & Katz, L. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4), pp. 855-902. Available at http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.20150572.

this approach can be more effective in enabling voucher holders to move to lower-poverty neighborhoods. HUD is also working to make voucher portability moves easier for Public Housing Authorities (PHAs). In addition, evidence from MTO suggesting that families were more successful at leasing units if the counseling agency identified potential units for them to lease-up supported proposals for a pilot program in the FY 2017 Budget, to increase administrative funding for PHAs in order to provide this counseling.

In addition to the role this evidence has played in housing voucher policies, evidence on the effects of neighborhoods and communities has been used to support a range of place-based initiatives, which in turn have generated demand for evidence on the effectiveness of those initiatives. Survey data often lack the sample sizes necessary to assess the effectiveness of these initiatives, but administrative data on outcomes such as earnings, test scores, and criminal records offer new opportunities to learn about the effectiveness of these programs.

5. *Food Assistance*

Effectively reaching those most in need is a critical concern of the Supplemental Nutrition Assistance Program (SNAP). The Food and Nutrition Service (FNS) uses administrative data—specifically, a sample of certification data drawn annually to measure payment accuracy (the quality control, or QC, data) to examine in depth the participants it serves. The latest annual SNAP Characteristics report developed with these data shows that well over half of the benefits (58 percent) go to households at or below 50 percent of the poverty line, and almost all of the benefits (93 percent) are going to households at or below 100 percent of the poverty line.[46]

While the QC data allow a national picture, other efforts that link administrative data with the Census Bureau's American Community Survey (ACS) data allow for estimates of smaller areas. Researchers at the Economic Research Service recently studied how to create smaller area estimates by working with FNS and the Census Bureau to link survey data containing demographic and annual income information from the ACS to administrative records from New York's SNAP program over 2008-2012.[47] They found that 27 percent of SNAP recipients live in deep poverty (less than half of the poverty threshold); these recipients garnered 32 percent of SNAP benefits. These results allow FNS to consider operational and policy changes that improve access to SNAP benefits for those in greatest need. For example, New York during this time period used a program simplification strategy aligning SNAP and Temporary Assistance for Needy Families rules, making the application process easier for the state and low-income families.

Several studies have linked administrative and survey data to assess the effect of SNAP on food insecurity[48] and have found a strong association between program participation and reductions in food insecurity. Census Bureau data have also been used to demonstrate the positive effects of SNAP in reducing poverty.[49] Other studies have used both cross-sectional and longitudinal data to show how

---

[46] Gray, K., & Kochhar, S. (2015). *Characteristics of Supplemental Nutrition Assistance Program Households: Fiscal Year 2014*. USDA: FNS, Report No. SNAP-15-CHAR. Available at http://www.fns.usda.gov/sites/default/files/ops/Characteristics2014.pdf.

[47] Scherpf, E., Newman, C., & Prell, M. (2015). *Improving the Assessment of SNAP Targeting Using Administrative Records*. ERR-186, USDA, Economic Research Service. Available at http://www.ers.usda.gov/media/1844884/err186.pdf.

[48] Mabli, J., Ohls, J., Dragoset, L., Castner, L., & Santos, B. (2013). *Measuring the Effect of Supplemental Nutrition Assistance Program (SNAP) Participation on Food Security.* Prepared for the USDA, FNS by Mathematica Policy Research. Available at http://www.fns.usda.gov/sites/default/files/Measuring2013.pdf.

[49] Tiehen, L., Jolliffe, D., & Smeeding, T. (2013). The Effect of SNAP on Poverty. *University of Kentucky Center for Poverty Research Discussion Paper Series*, *DP2013-06*. Available at http://www.irp.wisc.edu/publications/dps/pdfs/dp141513.pdf.

food insecurity affects learning and childhood development.[50] This evidence helped inform the Summer Electronic Benefits Transfer for Children (SEBTC) demonstration, which provided benefits on an electronic debit card to children eligible for free and reduced-price school meals to purchase food during the summer months when away from school. Rigorous evaluations of the demonstration found that SEBTC reduced "very-low food security" among children by one-third and improved participants' diets. The study used random assignment into treatment and control groups, combined with EBT transaction data, dietary recall surveys, household interviews, and the Food Security Supplement Survey to collect data from a sample of over 50,000 children.[51]

Another critical question of interest to FNS is how well the programs address participants' nutrition needs and encourage healthy choices. FNS has completed a number of studies, and others are ongoing. One study considered the impact on fruit and vegetable consumption of three different SNAP-Education program interventions.[52] That study used a process evaluation, which described how each program implemented its intervention and documented the challenges each encountered, and a set of impact evaluations, which used surveys of participants' parents and caretakers. The evaluations found significant positive impacts from two of the interventions, but not the third. These findings helped FNS focus nutrition promotion resources on the effective approaches.[53]

In addition to FNS's use of administrative data and primary data collection in its own research, it also partners with the Economic Research Service and academics in a number of venues. The two agencies are supporting research efforts to examine SNAP participants' food purchasing patterns using data from the National Household Food Acquisition and Purchase Survey (FoodAPS). FoodAPS is the first nationally representative survey of American households to collect unique and comprehensive data about household food purchases and acquisitions. It includes SNAP households, low-income households not participating in SNAP, and higher-income households. SNAP administrative data were used to construct the FoodAPS sample frame for SNAP participants, enabling the survey to collect information on a hard-to-reach population at a lower cost with greater accuracy. In addition, administrative records were used to obtain data on SNAP benefits, SNAP purchases, and locations of SNAP-authorized food retailers,

[50] Studies on associations between food insecurity and children's education or development are reviewed in Coleman-Jensen, A., McFall, W., & Nord, M. (2013). *Food Insecurity in Households with Children: Prevalence, Severity, and Household Characteristics, 2010-11*. USDA, Economic Research Service, EIB 113. Available at http://www.ers.usda.gov/media/1120651/eib-113.pdf.

[51] Collins, A., Briefel, R., Klerman, J., Wolf, A., Rowe, G., Enver, A., et al. (2014). *Summer Electronic Benefits Transfer for Children (SEBTC) Demonstration: Evaluation Findings for the Third Implementation Year.*. Alexandria, VA: Prepared for the USDA, FNS by Abt Associates, Mathematica Policy Research, and Maximus. Available at http://www.fns.usda.gov/sites/default/files/ops/sebtc2013.pdf.

[52] Long, V., Cates, S., Blitstein, J., Deehy, K., Williams, P., Morgan, R., et al. (2013). *Supplemental Nutrition Assistance Program Education and Evaluation Study (Wave II)*. Altarum Institute. Alexandria, VA: USDA, FNS. Available at http://www.fns.usda.gov/sites/default/files/SNAPEdWaveII.pdf.

[53] In addition, the School Nutrition and Dietary Assessment (SNDA) series uses large-scale, nationally-representative data collection through surveys, administrative records analysis, dietary recalls, stakeholder interviews, and on-site observation to examine the nutrition profile of school meals. The WIC Participant Characteristics (WIC PC) series analyzes administrative data to examine an extensive number of variables, including breastfeeding rates, use of the science-based food packages, and BMI measures, and can track these critical statistics over time because the data are collected biannually.

reducing respondent burden.[54] These efforts created a great deal of potential for research uses while adding more information about the respondents.

6. *Criminal Justice*

The Federal Government operates the Bureau of Prisons, where it can directly apply lessons it learns in the area of criminal justice. However, criminal justice is a policy area where the majority of funding and program operation is at state and local levels, typically with little Federal oversight. Consequently, the Federal role, in particular that of the National Institute of Justice (NIJ), is more to encourage and facilitate the building and use of evidence, rather than directly integrating evidence into the way in which programs are run.[55] This evidence can then be used by groups such as the Washington State Institute of Public Policy and the Pew-MacArthur Results First initiative (working with 22 states) to influence state and local policy in the area of criminal justice.

One of the important ways that the Federal Government facilitates the building of evidence in the area of criminal justice is by helping make large survey and administrative datasets available to researchers, evaluators, and programs. One such survey is the National Crime Victimization Survey (NCVS) administered by the Bureau of Justice Statistics (BJS). The NCVS surveys 90,000 households twice a year on the frequency, characteristics, and consequences of criminal victimization. Researchers use NCVS data to identify subpopulations that are at high risk for particular types of crimes and devise policies to reduce risk factors.[56] Participating states can use these data as well as Uniform Crime Report incident-based crime data from the FBI's National Incident-Based Reporting System (NIBRS) to measure the effects of policy and legislative changes on crime rates.[57]

Since 1995, BJS has administered the National Criminal History Improvement Program that, among other accomplishments, helped all states achieve full participation in the FBI's Interstate Identification Index. This critical operational network allows criminal justice agencies in the U.S. to exchange automated criminal history records (records that chronicle offenders' contacts with the justice system, i.e. "rap sheets"). Recently, BJS constructed an automated process that standardizes these variable Federal and state records and creates unified researchable databases that can support a variety of

---

[54] For more information, see FoodAPS National Household Food Acquisition and Purchase Survey. USDA Economic Research Service, available at http://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey.aspx.

[55] One of NIJ's vital roles is to provide program evaluation support to Department of Justice agencies and programs which then translates into funding to state and local governments. Across the Office of Justice Programs (OJP), NIJ serves as the evaluator for a number of OJP's high-dollar, high-profile initiatives, such as the Second Chance Act, the Byrne Criminal Justice Innovation program, the Wraparound Victim Legal Assistance Network, and the Linking Systems of Care project. Across these evaluations, NIJ's grantees are evaluating multiple sites that are funded to provide services to communities and to victims. Results from the evaluations help OJP program offices to understand the effectiveness of their funding, as well as to guide policymakers in decisions on future funding to improve public safety. These evaluations also have major implications for services and investments made by state and local governments.

[56] See, for example:

Lanier, D., & Dietz, T. (2012). Time Dynamics of Elder Victimization: Evidence from the NCVS, 1992 to 2005. *Social Science Research*, 41(2), pp. 444-463.

Schreck, C. (2014). Juvenile Victimization. *Encyclopedia of Criminology and Criminal Justice*, pp. 2785-2793. New York: Springer.

[57] Currently, the NIBRS data are reported by one-third of local enforcement agencies nationwide. These agencies are found in 37 states, with 16 states being "complete" reporters (meaning that all the agencies in the state report NIBRS data). BJS is working with the FBI to systematically add enough additional law enforcement agencies to the set of NIBRS-reporting agencies to make this sample nationally representative.

research and evaluation of recidivism patterns and sentencing.[58] In the future, BJS will seek to integrate its survey and administrative data into a unified system, leveraging existing criminal justice system record management systems. This system will permit several valuable analyses, for example, to determine which crime prevention programs correlate with changes or reduction in the incidence and types of crime.

In addition, BJS maintains high-quality databases on prison inmates, their entry and exit from the prison system, and the facilities maintained to hold them. These databases, assembled from administrative sources, support ongoing characterizations of the prison population, allowing policymakers to anticipate needs and develop strategies for successful rehabilitation.[59] Also, criminal records are a critical resource for studies and evaluations that test the effect of preventive interventions on outcomes such as earnings, education, and crime. These interventions vary from policies aimed at reducing recidivism in early childhood, where outcomes on crime may not appear until decades later.[60]

7. *Natural Resources and Environmental Quality*

Agencies implementing natural resource and environmental policies collect a range of administrative data, from state-level enforcement information to firm-level emissions data. The U.S. Environmental Protection Agency (EPA) publishes annual emissions data reported from industrial and Federal facilities in the Toxics Release Inventory (TRI), which is used to monitor pollution prevention activities across the country and to ensure public accountability for permitted air, land, and water emissions. EPA makes the longitudinal TRI data—collected from more than 20,000 facilities covering 650 chemicals—publicly accessible for use by local communities, stakeholders, and researchers. The latest annual report discloses that more than 85 percent of chemicals by volume reported to TRI are recycled, treated, or recovered for energy value.[61]

---

[58] Standardization of variables and the development of related schemata are vital tasks performed in the curation of data. As in this example, standardization is necessary to combine data from disparate sources. This necessity is a common obstacle in efforts to incorporate administrative datasets with other survey or administrative datasets. In addition to accelerating evidence-building, standardization also makes it possible to assess data quality by enabling comparisons with other measures, both within the country and internationally.

[59] One promising strategy is the formation of Therapeutic Communities (TCs) among prisoners to combat addictions to drugs and alcohol. TCs are separate residential drug treatment programs in prisons or jails that emphasize participation by all members of the program in the overall goal of reducing substance use and recidivism. The "communities" they support are the key instruments in making and solidifying changes in behavior. According to a systematic meta-evaluation of 55 studies by Washington State Institute for Public Policy, TCs are an effective way to reduce (post-release) recidivism rates; they are especially effective when coupled with "swift and certain" punishment approaches. For more information, see National Institute of Justice, "Incarceration-based Therapeutic Communities for Adults," available at www.crimesolutions.gov/PracticeDetails.aspx?ID=52 and Drake, E. (2012). *Chemical Dependency Treatment for Offenders: A Review of the Evidence and Benefit-Cost Findings*. Olympia, WA: Washington State Institute for Public Policy, available at http://www.wsipp.wa.gov/ReportFile/1112/Wsipp_Chemical-Dependency-Treatment-for-Offenders-A-Review-of-the-Evidence-and-Benefit-Cost-Findings_Full-Report.pdf.

[60] See, for example:

Miller, M., Fumia, D., & He, L. (2015). *The King County Education and Employment Training (EET) Program: Outcome Evaluation and Benefit-Cost Analysis*. Olympia, WA: Washington State Institute for Public Policy, available at http://www.wsipp.wa.gov/ReportFile/1621/Wsipp_The-King-County-Education-and-Employment-Training-EET-Program-Outcome-Evaluation-and-Benefit-Cost-Analysis_Report.pdf.

Reynolds, A., Temple, J., White, B., Ou, S., & Robertson, D. (2011). Age 26 Cost–Benefit Analysis of the Child-Parent Center Early Education Program. *Child Development*, 82(1), pp. 379–404.

[61] Environmental Protection Agency. (2016). *TRI National Analysis 2014*. Available at https://www.epa.gov/sites/production/files/2016-01/documents/tri_na_2014_complete_english.pdf.

TRI data are currently used to support a variety of program implementation and research purposes. Federal, state, and local governments utilize TRI data to establish environmental performance targets and monitor program performance.[62] For example, one EPA strategic performance goal from the Office of Air and Radiation tracks toxicity-weighted emissions of air toxics, relying on TRI data.[63] EPA is also developing resources to use TRI data to inform permit writers and program personnel.[64]

While EPA's TRI data are not comprehensive for all emissions that occur in the U.S. each year, the data provide a useful national indication of trends for waste management and emissions over time. Third-party researchers have matched TRI data to other data sources in order to investigate a variety of environmental health issues. For example, Hendryx and Luo matched TRI data for carcinogenic pollutant discharges to county-level hospitalization rates in 20 states, identifying associational evidence of higher admissions for certain cancers.[65] Similarly, Chen et al. examined associations between cardiovascular disease and zinc emissions using 20 years of TRI data.[66]

Other researchers have utilized TRI data to explore the efficacy of EPA pollution prevention programs, which explicitly aim to reduce emissions. For example, researchers from Abt Associates and EPA jointly developed a quasi-experimental evaluation of projects aimed at reducing pollution at the facility source with a difference-in-difference framework, estimating that on average, facilities that intentionally develop source reduction projects reduce emissions of targeted chemicals by 9 to 16 percent in the first year, relative to non-participating facilities.[67] Other similar panel research, using a decade of observations from TRI, identified that effects of pollution prevention projects can be short-lived, but sustained multi-year pollution prevention in facilities can generate 35 to 50 percent reductions in releases.[68] Still other researchers rely on TRI data in quasi-experimental designs to estimate the economic impacts of environmental policies. For example, in 2015, a team of researchers linked TRI data with information on housing transactions, birth outcomes, and firm closures to identify and compare the impacts of new industrial plants with permitted pollution on proximate home values and infant health.[69]

While much of the research using TRI data is associational, in recent years researchers have begun to apply TRI administrative data in quasi-experimental evaluation designs to build stronger evidence around certain environmental policies. In addition to TRI, EPA maintains extensive permitting and enforcement databases that can be used for evidence-building purposes in environmental policy.

---

[62] Environmental Protection Agency. (2013). *The Toxics Release Inventory in Action: Media, Government, Business, Community, and Academic Uses of TRI Data.* Available at https://www.epa.gov/sites/production/files/documents/tri_in_action_final_report_july_2013.pdf.

[63] Environmental Protection Agency. (2016). *FY 2015 Data Quality Records.* Available at https://www.epa.gov/sites/production/files/2016-02/documents/fy15-data-quality-records.pdf.

[64] Environmental Protection Agency. (2016). *FY 2017 Congressional Budget Justification*, p. 1062. Available at https://www.epa.gov/sites/production/files/2016-02/documents/fy17-congressional-justification.pdf.

[65] Hendryx, M. & Luo, J. (2013). Cancer Hospitalizations in Rural-Urban Areas in Relation to Carcinogenic Discharges from Toxics Release Inventory Facilities. *International Journal of Environmental Health Research,* 23(2), pp. 155-169.

[66] Chen, B., Luo, J., & Hendryx, M. (2015). Zinc Compound Air Releases from Toxics Release Inventory Facilities and Cardiovascular Disease Mortality Rates. *Environmental Research* 142, pp. 96-103.

[67] Ranson, M., Cox, B., Keenan, C., & Teitelbaum, D. (2015). The Impact of Pollution Prevention on Toxic Environmental Releases from U.S. Manufacturing Facilities. *Environmental Science and Technology* 49(21), pp. 12951-12957.

[68] Harrington, D, Deltas, G., & Khanna, M. (2014). Does Pollution Prevention Reduce Toxic Releases? A Dynamic Panel Data Model. *Land Economics* 90(2), pp. 199-221.

[69] Currie, J., Lucas, D., Greenstone, M., & Walker, R. (2015). Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings. *American Economic Review* 105(2), pp. 678-709.

VI. Variety of Factors Influence the Integration of Evidence Into Program Design

Administrative and survey data support a suite of evidence-building activities, ranging from regular performance measurement and monitoring to periodic formal program evaluations. As the examples in the previous section illustrate, a wide range of administrative and survey data help develop all components of a portfolio of evidence, and a full and robust portfolio of evidence is important to informing a wide variety of program decisions. There are also a wide range of structures available to support the development of a robust portfolio of evidence. This variety makes it challenging to prescribe a one-size-fits-all approach to developing a portfolio of evidence or for how that portfolio should feed into program design and decision-making. This calculus is complicated further by the variation in what evidence is already available, and what types of evidence can reasonably be generated in a given policy area. All of this points to flexibility being a key consideration in any attempt to navigate the optimal methods to embed evidence-building into programs and the use of evidence in program design.

A key factor is the degree to which state and local jurisdictions are permitted flexibility in their implementation of Federal programs, based on local needs and conditions. This spectrum of flexibility has important implications for how to embed evidence generation and use into program design. The FAFSA example illustrates how a program that is directly administered by the Federal Government can apply lessons based on evidence (within statutory confines) so that those changes will be implemented consistently for those who access the program. The criminal justice example illustrates the other end of the spectrum, where there is a great deal of flexibility based on state and local needs and laws. In these cases (where the Federal Government provides neither a regulatory framework nor the majority of funding), the Federal Government primarily serves as an instigator and facilitator to support states and local areas in building and using evidence to inform policy decisions.

Most programs fall in between these extremes, consistent with the principles of cooperative federalism. The WIOA example highlights a common model of program design, in which states and localities have a great deal of flexibility (within certain statutorily prescribed constraints) to tailor their implementation, evidence-building, and use of evidence to fit local needs and circumstances. Many block grants also fit into this category. In these cases, the Federal Government sets standards and guidelines that must be flexible enough to accommodate varying needs of states and local areas. In these areas the availability of administrative data for evidence-building purposes, and evidence-building activities themselves, are frequently up to state discretion and are governed by state-level policies and laws that the Federal Government has somewhat limited power to influence. This highlights the fact that the ability of Federal Government managers to encourage or compel valid and reliable data collection—and eventual reporting to an accessible central repository—is a key issue in both establishing and maintaining administrative datasets that are useful in evidence-building endeavors. This ability to compel valid and reliable data collection remains a central issue regardless of whether programs are delegated to states or other entities.

There are also more standardized programs, like Housing Vouchers or SNAP, where programs are tightly constrained through regulations and other program parameters. Tightly scoped competitive grants can also fall into this category. Where implementation is tightly standardized due to Federal requirements, flexibilities may be helpful to test new approaches, to build evidence of effectiveness, and to scale up implementation of effective approaches. Federal waivers and other demonstration authorities can be useful flexibility tools in these contexts. These flexibilities are especially important if the program is mandatory or an appropriated entitlement, where any change in implementation could have budgetary impacts. The set-aside for testing innovative approaches within the MIECHV program is a good example of how flexibilities within a fairly standardized program can help build evidence.

In all of these cases, the collaboration between the Federal Government and state or local partners and implementers is an important factor in increasing the quality of administrative data and evidence generation, and effectively using evidence to improve programs. The quality of all of these practices greatly improves when those who are actually implementing the program are invested in the outcomes of the associated evidence efforts. The Federal role is frequently one of supporting state- and local-level evidence generation and implementation, through establishing standards, providing technical assistance, and disseminating best practices.

A second key factor that influences the strategy of how best to incorporate evidence-building and use into program design is the breadth of the program and how it fits within the larger policy area. As the examples above highlight, a broad portfolio of evidence is useful to influence policy. However, policy areas may cover multiple programs and funding streams, which themselves may have a variety of different program structures. Similarly, a program itself may cover either a very tightly defined set of strategies (e.g., home visiting, FAFSA) or a broad range of actual interventions (WIOA, criminal justice).

In cases where programs are narrowly prescribed *and* the policy area is narrowly defined (or the program is the major funding stream in that policy area), it may be easier to explicitly incorporate requirements into the program that facilitate building or using evidence. Home visiting is a good example of a tightly-defined policy area where, in addition to providing funding for evidence-based programs and evidence-generation, the MIECHV program also provides a centralizing policy lens that may encourage states to dedicate a larger portion of funding from other sources to these evidence-based practices.[70] In areas where multiple programs influence the policy area, it probably makes more sense to pursue flexible, cross-program approaches to building and using evidence. Higher education, workforce development, and housing policy are all examples where multiple programs come together to support a policy area.

In these multi-program areas, there are a variety of program and funding structures that can support effective and efficient targeting of evidence-building and use activities to the most promising areas. Tools such as evaluation set-asides, transfers between programs, and centralized evaluation and research funding can be especially useful since they allow the flexibility to look across programs (and look at themes that span multiple program areas). These tools allow evidence-builders to target the most interesting or promising areas of development across the range of programs within the policy area, rather than being limited to what can be tested within the confines of a single program.

A third key factor is the quality and depth of the portfolios of evidence in a given program or policy area. Some program areas, such as home visiting programs and early childhood education programs, have rich evidence-bases. Other program or policy areas may not currently have similarly rich evidence-bases, but are well-suited to further evidence-building. Other program areas are more difficult to evaluate or measure performance. Some program areas have rich policy variation across geography or time or households that lend themselves to credible, large-scale quasi-experiments. In other program areas, such policy variation is absent or selection is such a big issue that credible quasi-experiments are challenging. Randomized experiments with sufficient sample sizes are possible in some program areas, but in others, such randomization is infeasible or challenging to do with the sample sizes necessary for credibility. In some program areas, evaluating the overall effectiveness of a program is useful and

[70] Michalopoulos, C., Lee, H., Duggan, A., Lundquist, E., Tso, A., Crowne, S., et al. (2015). *The Mother and Infant Home Visiting Program Evaluation: Early Findings on the Maternal, Infant, and Early Childhood Home Visiting Program*. OPRE Report 2015-11. Washington, D.C.: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Available at http://www.acf.hhs.gov/sites/default/files/opre/mihope_report_to_congress_final_1.pdf.

actionable; in other program areas, the most important and actionable questions center around small variations in interventions or how interventions are implemented.

The recurring theme here is that evidence-based policymaking is challenging and multi-dimensional. In order to determine whether a particular evidence-based program structure is appropriate for a given program area, the degree of consistency in program implementation, the breadth of the program area, the quality and nature of the evidence-base (and possible evidence-base) all must be considered. That dimensionality has led to entire commissions devoted to specific aspects of evidence-building, such as standards for economic analyses or principles and practices of evaluation functions. In some cases, Congress has established requirements on evaluations in a particular program or policy area. While these requirements are well-intentioned, they may limit the flexibility of the researchers or evaluators to apply the most appropriate method to answer the question that Congress intended to have answered. In these cases, the suitability of the resulting evidence to answer key policy questions may be compromised. Flexibility is key to being able to match the appropriate program design and evidence-building process to a given program area.

All that said, one theme remains universal—the availability of high-quality administrative and survey data is an especially powerful tool to build portfolios of evidence, support programmatic changes, and identify where new data are needed for future decisions. Improving the quality of data and access to it is likely to help translation of evidence into decision-making regardless of the program structures and evidence-base in a given policy area.

VII. Summary of Themes

As societal needs change and issues become more complex, program officials, researchers, and institutions are faced with increased demands for more complete information to make smart and strategic decisions that are grounded in evidence. Over the last several years, the Federal Government has emphasized the need for building evidence and embedding its use into a wide range of policy and program formation. At the same time, a proliferation in data generation promises to continue for some time, and new technologies are providing the ability to utilize it. This confluence of factors presents an opportunity for the government to make more systematic and effective use of data and to develop and promote the most effective public policies.

Building a stronger base of evidence for policymaking requires the maintenance of high-quality, well-curated, accessible data sources that underlie the different evidence components. Ideally, these complementary data sources and evidence components will be used in concert to take advantage of their varying strengths. As we have described earlier, the government is already producing or facilitating a great deal of this work and developing more and better methods and tools on an ongoing basis. The examples we have highlighted in the body of this paper provide just a small sample of the valuable evidence-building work that is underway.

Key takeaways identified in the paper include:

- Traditional aggregate measures establish baselines and inform broad directions for social, economic, and environmental policy. But the data underlying these measures, which originate primarily in large statistical surveys, could be strengthened and improved to provide even clearer signals to policy-makers through the use of administrative datasets.
- In other cases, entirely new datasets are being formed through the careful curation of the administrative data that exists in disparate locations and within programs that have not traditionally curated their data. The efforts by BJS to build analytical databases from rap sheet

repositories and prison records are just one such example, and several others are included in the appendix. There is potential to build on these efforts to create more and better evidence.

- There is potential for linkages between survey and administrative datasets to allow researchers to exploit the best aspects of the different data sources while minimizing their weaknesses. The work discussed above combining SNAP administrative records with the American Community Survey, as well as the linkages described in the appendix, such as the collaboration between the Centers for Medicare & Medicaid Services and HUD and the Supplemental Poverty Measure being created by BLS and the Census Bureau, are all exemplars of this promising potential.

- The work by Chetty and his collaborators using IRS tax data and the re-use of existing study-specific survey data illustrates the power of the large-scale administrative datasets that have been created by the Federal system, but are exceedingly difficult for even the most qualified researchers to access. This example also illustrates the power of retaining and re-using high-quality study-specific data, in this case the HUD MTO evaluation data. There is tremendous potential for combining these and other similar datasets to facilitate a broad type of evidence-building activities. Devising a way to scale up access to these powerful sources of evidence would almost surely lead to a significant expansion of the knowledge base required to formulate effective policies.

- Program-specific evaluations are vital to the mission of building an effective government by informing decisions to expand on what works and improve or abandon what does not. Sometimes program-specific survey and administrative data can be used in combination with large-scale administrative datasets to facilitate efficient evaluations, as is the case in many workforce evaluations and the MIECHV evaluation efforts. In other cases, these evaluations combine a wide range of survey and administrative data to efficiently capture a broad range of information. A prime example is the SEBTC evaluation, which used a combination of EBT administrative data, study-specific surveys, and the Food Security Supplement Survey. However, these efforts are often carried out on a "one-off" basis, incurring a high cost. With more systematic re-use of intermediate products and a clearer path to the available data, these valuable efforts could be made easier and less expensive.

The strides that have been made in each of these areas result from concerted efforts over several years by agencies, components of the Federal Statistical System (FSS), and Federal evaluation offices. This progress is likely to continue as innovation, collaboration, and newly developing technical expertise are further expanded. The FSS is a natural nexus for much of the technical infrastructure given its valuable assets, including its extensive existing infrastructure, its accumulated experience in curating data and provisioning its access, and its ample store of existing data on which it can build. In order to fully realize this potential, consideration must be paid to how the Federal evaluation offices, external researchers, and other state and local entities can leverage this valuable infrastructure to generate further evidence. The Commission may consider how it can aid this progress, by further encouraging areas of growth, and by taking steps to address issues that remain impediments, as described in the paper, "Barriers to Using Administrative Data for Evidence-Building."

Appendix

Emerging Uses and Potential for Expanded Application

As demonstrated by the examples in the main text, data are already being used to build portfolios of evidence in a wide variety of settings. However, many opportunities to use data for evidence-building remain unexploited, either because the data are not yet available for the potential application, or because their application is only recently emerging. This appendix highlights several examples of such potential and/or emerging uses, focusing on the use of administrative data, since they have been a major source of untapped data use potential for years, a target of concerted facilitation efforts in the recent past, and a component of the Commission's charge.[71] In order to capture the breadth of the developments relevant to this discussion, only brief descriptions of the various projects are provided. Expanded application of administrative data has the potential to enrich the data underlying every component of the evidence portfolio.

- The Department of Housing and Urban Development (HUD) has collaborated with the Census Bureau, the Centers for Medicare & Medicaid Services (CMS), and most recently the National Center for Health Statistics (NCHS), to combine data and expertise to study relationships between housing, health risk behaviors, and health in order to use housing as a platform to improve quality of life. One outcome of these collaborations is new availability of linked survey and administrative datasets for researchers.

- HUD is collaborating with the Department of Education and the White House Social and Behavioral Science Team to test the impact of varying messengers and messages informed by behavioral insight on Free Application for Federal Student Aid (FAFSA) completion rates and post-secondary enrollment among young adults in public and assisted housing.

- HUD is also comparing the data obtained through administrative tax assessments with several of the variables captured by the American Housing Survey to explore the extent to which the administrative data source might supplement or improve upon the survey's results.

- HUD has developed a data licensing project through which research organizations can access personally identifiable information in HUD administrative data for policy-relevant research purposes without HUD funding. Participating organizations must provide adequate safeguards for confidentiality and destroy the personally identifiable information when the license expires.

- CMS has a strong and longstanding infrastructure for making administrative data available for statistical uses with strong privacy protections, including by outside researchers. CMS has added to this infrastructure with the new CMS Virtual Research Data Center. In addition, in 2014 CMS for the first time released Medicare utilization and cost data summarized at the physician level, letting Americans compare their own doctors' practice patterns with national norms. Likewise, the CMS "Hospital Compare" feature provides information on how almost every U.S. hospital performs with respect to clinical quality metrics (such as whether patients receive timely and effective care) and hospital readmission rates.

---

[71] See, for example:

Office of Management and Budget. Memorandum M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes*. https://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

Chapter 7, "Building Evidence with Administrative Data," in *Analytical Perspectives, Budget of the United States Government, Fiscal Year 2016*, pp. 65-73. https://www.whitehouse.gov/sites/default/files/omb/budget/fy2016/assets/ap_7_evidence.pdf.

- The Bureau of Justice Statistics is working with the Census Bureau to determine how well the National Corrections Reporting Program data can be linked to other files. This will enable researchers to follow the experiences of individuals as they move through the criminal justice system. Such files, linked via the Census Bureau's specialized record linkage center, may include a variety of sources tracking receipt of government benefits (Temporary Assistance for Needy Families or TANF, housing support, Medicare), work history (Longitudinal Employer-Household Dynamics, tax records), and other vital events (death master file). Such a rich, linked dataset would allow a variety of research and evaluation avenues to be pursued, for example in examining interventions, recidivism patterns, and sentencing.

- The Social Security Administration is participating in a joint pilot demonstration project with the Department of Education**,** the Department of Health and Human Services (HHS), and the Department of Labor (DOL) to test interventions that improve the health, education, and post-school outcomes of children who receive Supplemental Security Income benefits.

- The National Center for Education Statistics (NCES) has been working with the Census Bureau to seek a way to create measures of socio-economic status characterizing the students attending different schools. In its experimental phase (working with 2,000 schools in Ohio), the project applies spatial interpolation methods to the Census Bureau's American Community Survey (ACS) data to calculate a poverty index (i.e., an income-to-poverty ratio) based on ACS household responses and NCES public elementary school points.

- The Census Bureau and Bureau of Labor Statistics (BLS) are leading a multi-agency effort to improve the Supplemental Poverty Measure (SPM), designed to complement the official poverty measure, which is based on outdated assumptions and does not take into account most government transfer programs, and hence, cannot by itself be used to evaluate the impact of anti-poverty programs. In order to estimate the effectiveness of targeting resources toward the disadvantaged, the SPM integrates household income and expenditure information from national survey data with administrative data from a variety of Federal programs that help families, households, and individuals meet their basic needs.

- BLS is using information on employment at foreign-owned U.S. enterprises made available through an agreement with the Bureau of Economic Analysis to associate establishments with their parent enterprises. The goal of this work is to allow for a greater tracking of employment effects from inward foreign direct investment, providing vital information to inform international direct investment policies.

- BLS is also evaluating the feasibility of merging its National Compensation Survey (NCS) data with tax filings from the Internal Revenue Service. Although there are legal hurdles to overcome, such a file linkage could potentially provide the NCS with accurate and low-cost information on the provision of health insurance by employers, as well as premiums paid.

- The Employment and Training Administration and the Chief Evaluation Office at DOL have gained approval to use data from the National Directory of New Hires (NDNH) in several high profile evaluations, including the Workforce Investment Act Gold Standard Evaluation, Transitional Assistance Program for active duty military persons, Enhanced Transitional Jobs Evaluation, YouthBuild Evaluation, and Middle Class Tax Act Evaluations of Unemployment Insurance.

- The Employment and Training Administration at DOL is using BLS data from firms in the Occupational Information Network (O*NET) occupational skills data system that is a critical source of information about the skills in demand by employers. This system is used to help inform training program needs, and provides information to customers about the skills and credentials required for various programs.

- DOL's Workforce Data Quality Initiative grant program funds development of state workforce longitudinal databases to provide a comprehensive picture of workers' skill development and earnings throughout their careers.[72]

- DOL has recently created the Chief Evaluation Office Data Analytics Unit, which provides analytical services and capabilities for DOL programs, as well as products available for the public, typically using DOL administrative data.

- HHS has established a comprehensive microsimulation model of the major governmental tax, transfer, and health programs (SNAP, SSI, WIC, TANF, LIHEAP, EITC, etc.) in order to understand the potential outcomes of public policy changes such as tax reform and health reform. The Transfer Income Model combines administrative data on program rules with survey data from the Census' Current Population Survey Annual Social and Economic Supplement and also uses data from the ACS and Survey of Income and Program Participation (SIPP). Most importantly, the model can examine how these programs interact with each other and the cumulative benefits that are provided to individuals and families. The model estimates the number of families and individuals eligible for a number of programs and can be used to demonstrate how these rules translate to the population over many years, quantifying how many individuals were eligible for these programs and how many actually received benefits to inform policy and program decisions.

- The Administration for Children & Families (ACF) uses NDNH data in several of its employment and training evaluations. The NDNH is the source of W-2 wage data for the joint ACF-DOL Subsidized and Transitional Employment-Enhanced Transitional Jobs Demonstration, the Pathways to Advance Careers and Education evaluation, the Health Professions Opportunity Grants evaluations, and the Job Search Approaches evaluation.

- The Agency for Healthcare Research and Quality (AHRQ) recently used aggregate benchmarks from National Health Expenditure Accounts and tax expenditure estimates to align and supplement household survey data from the Medical Expenditure Panel Survey (MEPS). By combining the richness in characteristics of the MEPS with the accuracy on key variables of administrative sources, these combined data can serve as an enhanced resource for conducting benefit incidence or microsimulation analyses.

- The AHRQ Evidence-Based Practice Center Program conducts many systematic reviews, including assessments of program interventions, and has compiled a Methods Guide for Effectiveness and Comparative Effectiveness Reviews,[73] which describes guidelines for assessing

[72] U.S. Department of Labor, Employment and Training Administration. *Workforce Data Quality Initiative (WDQI) Grant Information*. Available at http://www.doleta.gov/performance/workforcedatagrant09.cfm.

[73] Agency for Healthcare Research and Quality. (2014). *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. AHRQ Publication No. 10(14)-EHC063-EF. Available at https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=318.

a body of evidence, including a chapter on how to assess the risk of bias of an individual study to answer questions about effectiveness.[74] The guide also examined how to consider the role of these contextual factors in assessing the effectiveness of program or policy interventions.[75]

- NCHS has increased the flexibility of the sample design for the National Health Interview Survey to respond to increasing demand for state-level health outcomes, particularly to inform the implementation and evaluation of the Affordable Care Act.
- The Department of Agriculture (USDA) Food and Nutrition Service's evaluation of ten Supplemental Nutrition Assistance Program (SNAP) unemployment and training pilots. The pilots were authorized by Congress in 2014, and will make use of state administrative data on SNAP participation and unemployment linked with project-derived survey data on participant characteristic and food insecurity to assess intervention impact.
- The Economic Research Service (ERS) is using USDA administrative data on grain storage levels and forecasts, as well as crop conditions, to study the volatility of grain derivative prices.
- ERS is combining administrative data collected by the Rural Business-Cooperative Service (RBS) with business establishment data from the National Establishment Times Series data and other data sources to assess impacts of selected RBS grant and loan programs on rural business survival and growth.
- ERS is developing methods to address the gap between farm surveys conducted at the level of the operation and farm program administrative data, which typically reports on the basis of subsets of the operation, such as contracts based on farm fields. This work will facilitate research to examine the implications of treating subsets of the operation as independent decision-making units (and ignoring interactions between those units), when analyzing the effects of program participation decisions on program goals and thus program effectiveness.
- ERS is combining conservation program administrative data with spatially explicit datasets on groundwater, satellite-based crop residue estimates, and other environmental and socio-economic variables to study: 1) environmental drivers of conservation program participation, 2) the legacy effects of program participation, and 3) the role of conservation programs in drought mitigation.

---

[74] Viswanathan, M., Ansari, M., Berkman, N., Chang, S., Hartling, L., McPheeters, L., et al. (2012). *Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions*. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews, Publication No. 12-EHC047-EF. Available at https://www.effectivehealthcare.ahrq.gov/ehc/products/363/1384/White-Paper_Global-Health-Frameworks-1-22-13.pdf.

[75] See, for example:

Guise, J., Chang, C., Viswanathan, M., Glick, S., Treadwell, J., Umscheid, C., et al. (2014). *Systematic Reviews of Complex Multicomponent Health Care Interventions*. Agency for Healthcare Research and Quality Publication No. 14-EHC003-EF. Available at https://www.effectivehealthcare.ahrq.gov/ehc/products/578/1878/health-care-interventions-review-report-140303.pdf.

Shekelle, P., Maglione, M., Luoto, J., Johnsen, B., Perry, T. (2013). *Global Health Evidence Evaluation Framework.* Agency for Healthcare Research and Quality Publication No. 13-EHC008-EF. Available at https://www.effectivehealthcare.ahrq.gov/ehc/products/363/1384/White-Paper_Global-Health-Frameworks-1-22-13.pdf.

Rojas Smith, L., Ashok, M., Dy, S., Wines, R., Teixeira-Poit, S. (2014). *Contextual Frameworks for Research on the Implementation of Complex System Interventions.* Agency for Healthcare Research and Quality Publication No. 14-EHC014-EF. Available at https://www.effectivehealthcare.ahrq.gov/ehc/products/490/1882/contextual-frameworks-complex-interventions-report-140318.pdf.

- The National Science Foundation (NSF) is developing and using text-mining tools to link NSF data on research awards to patent data from the United States Patent and Trademark Office to identify patents that could have links to NSF's research funding.
- NSF is investing in the development of text mining tools that can be used to explore research portfolios of Federal agencies investing in science and engineering research and education to gain an understanding of the type and nature of the research being funded and to inform decisions regarding future funding.