Issues in Accessing and Using Administrative Data

Kelly Maxwell, Child Trends

FEBRUARY 2017

OPRE REPORT #2017-24

Policymakers are increasingly interested in using administrative data to address pressing, policy-relevant questions. The federal Office of Management and Budget (OMB), for instance, issued a memo in 2014 that encouraged agencies to use and share administrative data and provided guidance related to using administrative data for statistical purposes (M-14-06).[1] Building on this, the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families, U.S. Department of Health and Human Services focused its 2015 Innovative Methods meeting on the promises and challenges of using administrative data in social policy research.

This brief is based on a panel presentation at that meeting, *Gaining Access and Maintaining Confidentiality*. The purpose of this brief is to provide an overview of the multiple aspects of access to consider when using administrative data for social policy research. It includes discussion of access to data, importance of relationships, considering and valuing both confidentiality and access to data, and building capacity for the use of administrative data. As an overview, it is intended to raise awareness of issues rather than extensively describe access issues or offer strategies for overcoming challenges in accessing data. Multiple other resources are available regarding the use of administrative data. One example is the Child Care and Early Education Research Connections Working with Administrative Data webpage,[2] where administrative data resources are summarized and updated quarterly.

The purpose of this brief is to provide an overview of the multiple aspects of access to consider when using administrative data for social policy research.

[1] https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf

[2] http://www.researchconnections.org/content/childcare/understand/administrative-data.html

Dr. Kelly Maxwell is a senior research scientist and co-director of early childhood development at Child Trends who helps states improve their early care and education systems through research and evaluation.

LAYERS OF ACCESS

Access is one of the major issues in using administrative data for research purposes. Although one might think first about access as receiving the data from an agency, there are many layers of access. Four layers are highlighted below.

Layer 1: Determining availability. The first layer of access is determining whether the data even exist. Researchers might think that a federal or state agency collects certain data, but it is important to speak with program staff to ensure that the specific data of interest are available. The policy question of interest might require data about the capacity of early care and education programs to serve more children. Although the state licensing agency might have data on related variables (e.g., total enrollment and licensed capacity of the program), it may not have all the specific administrative data needed to address the question (e.g., how many more 2-year-olds could be served by licensed programs).

Layer 2: *Receiving the data*. This layer of access includes many steps that may require extensive time to accomplish. Three major steps are highlighted here. First, researchers and program staff need to determine the specific data variables needed to address the question of interest. Second, they will need to establish a data sharing agreement and have it approved by the proper authorities. Third, they need to determine how best to format the data in a way that the research team can use.

Layer 3: *Merging multiple data sets*. The data needed to address the particular question of interest often exist in separate systems within the same agency or in two or more agencies. Once the research team has the various individual data sets with the variables of interest, they then need to successfully merge the data in order to access all of the data needed to answer a question of interest. Merging can be complicated if the unit of analysis (e.g., child) does not match the level of the data (e.g., family) or if the research team has to match cases from multiple data sets using several data variables (e.g., name, date of birth) because a unique identifier was not available.

Layer 4: *Understanding what the data really mean (and don't mean).* This is arguably the most important layer of access and will take a significant investment of time for both researchers and the program staff who have and use the data. Researchers should understand the variable definitions, the quality of data collection and data entry, and the policy context in which the data were collected. If an agency, for instance, changed its policy for "full-time attendance" from 25 to 16 days per month in 2014, then the "full-time attendance" variable does not mean the same thing over time. Without knowing what the data mean, researchers could misinterpret the study findings. Researchers must work closely with program staff to review relevant documents, discuss policies, and ask questions to learn about the data considered for analyses.

IMPORTANCE OF RELATIONSHIPS

All of the layers of access require strong relationships among researchers, program staff, data managers, and other data stewards. Researchers ideally conduct studies using administrative data in partnership with program leaders and staff, working together at every stage from identifying or refining the research questions to interpreting the findings. Each partner makes an important contribution to the research. The program staff bring their expertise on the program, policies, and data, while the research team brings its expertise on research design, data analysis, and the relevant literature. Strong relationships are characterized by qualities such as clear communication, trust and mutual respect, and attention to the preservation of institutional knowledge given expected changes in staff. Researchers who create and maintain these strong partnerships are better able to analyze administrative data to appropriately address meaningful social policy questions.

A *BOTH-AND* NOT *EITHER-OR* APPROACH

Researchers and program staff sometimes may feel an *either-or* push between gaining access to administrative data and maintaining the confidentiality of data. At times, access to data may be limited because of confidentiality concerns. It may also be challenging at times to put into place the procedures necessary to protect the confidentiality of data. However, *both* confidentiality *and* access are important and necessary to conducting good research. The usefulness of administrative data is dependent on persistence in addressing concerns that may arise about enabling access while maintaining confidentiality; trusting relationships may be particularly important for working through such concerns. Everyone involved in conducting research with administrative data on both the program and research side needs to carefully consider the applicable federal and state regulations regarding data confidentiality and data sharing—and work together to reach the shared goal of appropriately using administrative data to address important social policy questions. Products like the M-14-06 OMB memo help agency leaders better understand how to appropriately share and use data for research purposes.

BUILDING CAPACITY

To take advantage of the opportunities for using administrative data for social policy research and to overcome some of the challenges in accessing administrative data, we need to build capacity. This section describes why this is important and offers some ideas about what is needed.

The amount of available data is growing, in part because of an emphasis on data-driven decision making as well as the increase in the availability of "big data." Technological advances are also making it easier to analyze data, which may spur more researchers to consider studies that rely on administrative data. At the same time, the laws and policies that govern the use of data are increasing and will likely continue to evolve at a rapid pace in the coming years. It is challenging to keep up with all the changes and understand recommended practices in using administrative data for research purposes. Thus, we need to build capacity related to the use of administrative data for social policy research. We need to consider *multiple roles, multiple levels*, and *multiple areas* of capacity building.

Multiple roles. We need to build the capacity of people in various roles related to using administrative data for research purposes. This includes researchers, agency and program leaders, data managers, program staff who collect the data, staff who oversee the program or agency, and funders. Capacity building looks different for each role. For example, staff who work directly with data may need in-depth information about data systems and procedures for ensuring high-quality data. Agency leaders, on the other hand, may need support in understanding federal or state guidelines about storing and sharing data. Researchers may need support in understanding appropriate techniques for matching data from the same individual across multiple data sets or for learning how best to ensure data security when storing a data set.

Multiple levels. We need to consider building capacity at the local, county, regional, state, and federal levels. Although the administrative data may be accessed for research at one specific level (e.g., state agency), individuals at each level may influence the data's collection and use. Thus, it is important to build capacity across these levels to support a shared understanding of the issues in using administrative data for research purposes. For instance, local or county staff who collect the data may need support about how best to record the information consistently and completely. Communicating across the multiple levels can also build capacity by allowing individuals at each level to understand the full range of issues from data collection to data storage, data use, program reporting requirements, and federal regulations. Improving people's understanding of *how* the data are used can also help individuals at multiple levels understand their role in supporting high-quality data collection and use, as well as the benefits that data can provide across levels.

Multiple areas. This section suggests three areas to address when building capacity to use administrative data for research purposes;

> *Data systems*. Do the programs and agencies have the basic systems to support their storage and use of administrative data? Sometimes program staff are limited in their ability to use their administrative data because of limitations in their data system. These limitations can be related to various facets of a data system, such as whether the front end of the system allows staff easy access to the data or whether major systems have common identifiers to facilitate linking records. Programs and agencies can work to build or improve their data systems to better take advantage of the promise of administrative data.

Knowledge and expertise. Administrators, program staff, and data managers need a strong knowledge of the data and policies affecting the data. Building capacity can be as simple as creating or updating a data dictionary that defines the data variables in each data set. Program, regional, state, and federal agency staff need expertise to utilize their data. This expertise can be provided by in-house staff who have research expertise or by external research partners who can help them analyze their data to address questions of importance to them. Researchers must have knowledge about methods and analytic techniques to appropriately analyze administrative data. In addition to having this professional expertise, *everyone* involved in using administrative data must keep up-to-date with the laws and policies that govern data. Tools like [ACF's Confidentiality Toolkit][3] and websites like [The Privacy Technical Assistance Center][4] can be very helpful.

Everyone involved in using administrative data must keep up-to-date with the laws and policies that govern data.

Culture. Organizations of all kinds benefit from developing a culture that values and uses data to make decisions and inform practice. A strong data culture within an organization or field makes it easier to use administrative data for research purposes. Organizations vary widely in their capacity to collect, document, use, and share data. Federal and state funders and partners can play an important role in supporting this data culture at the state and local levels through regulations, grants, and technical assistance. The Head Start Leadership, Excellence, and Data Systems[5] project, for example, offers resources to help local Head Start programs strengthen their data culture.

---

[3] https://www.acf.hhs.gov/sites/default/files/assets/acf_confidentiality_toolkit_final_08_12_2014.pdf
[4] http://ptac.ed.gov/
[5] https://www.acf.hhs.gov/opre/research/project/head-start-leadership-excellence-and-data-systems-hs-leads

CLOSING REMINDER

Finally, a key point to remember is that there is a person behind each data point. It can be easy to forget that when we are wading through spreadsheets or statistical reports; however, everyone who collects, oversees, or uses administrative data should remember the individual children, families, service providers, and other people who *provide* the data. Remembering the people behind the data will help foster appropriate data safeguards and strengthen the rigor and relevance of research using administrative data.

WANT TO LEARN MORE?

This brief is based on a roundtable discussion from OPRE's 2015 meeting, *The promises and challenges of administrative data in social policy research*. Video of the roundtable discussion, *Gaining Access and Maintaining Confidentiality*, is available from https://www.youtube.com/watch?v=KivNnO_52a8. To access the online meeting archive, including a detailed schedule, meeting materials, and presentation slides, please visit the OPRE Innovative Methods Meeting website at www.opremethodsmeeting.org. The site also includes materials from other innovative methods meetings that OPRE has organized and will be updated to include future meetings.

This brief was prepared for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. It was developed under Contract Number HHSP23320095651WC. RTI International assisted with preparing and formatting the brief. The ACF project officers are Anna Solmeyer and Nicole Constance. The RTI project director is Anupa Bir.