Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services

OFFICE OF HUMAN SERVICES POLICY

Leveraging ACF Administrative Data for Evidence and Research

Authors: Robin Ghertner and Kendall Swenson

_____

HIGHLIGHTS

This brief analyzes information on administrative data collected by the Administration for Children and Families (ACF). It explores how agencies and researchers can leverage data to improve evidence and research on ACF programs and beneficiaries. Key highlights include:

- ACF offices, when administering programs, collect and process data that present opportunities to understand the programs and participants and to generate evidence of effectiveness.

- Datasets vary in how they can be used in research, depending on factors such as the inclusion of personally identifiable information and the ability to link with other data sources.

- Variability in dataset characteristics is largely due to differences in statutory and regulatory authority, as well as infrastructure for data collection.

- HHS is taking steps to expand the utility of administrative data for research and evaluation activities, within statutory authority.

_____

Introduction

Over the past several decades, federal agencies have heard increasing calls to make more and better use of their administrative data to expand the evidence base that guides federal policies and programs. The Office of Management and Budget issued several memoranda on using administrative data for evidence-building activities.[1] The Commission on Evidence-Based Policymaking, a nonpartisan effort to find ways to increase the availability and use of evidence in federal decision-

Administrative data are defined by the Office of Management and Budget as "administrative, regulatory, law enforcement, adjudicatory, financial, or other data held by agencies and offices of the government or their contractors or grantees…and collected for other than statistical purposes."[1]

---

[1] Office of Management and Budget. "Barriers to Using Administrative Data for Evidence-Building." White paper prepared by the Office of Management and Budget for the Commission on Evidence-Based Policymaking. Washington, DC: Office of Management and Budget, July 15, 2016. Office of Management and Budget. "Guidance for Providing and Using Administrative Data for Statistical Purposes." OMB M-14-06. Washington, DC: Office of Management and Budget, 2014. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2014/m-14-06.pdf

making, issued a report in 2017 that included recommendations for ways to expand the use of administrative data while protecting privacy.[2] Most recently, the Foundations for Evidence-Based Policymaking Act of 2018 requires agencies to develop and maintain comprehensive federal data inventories, among other efforts intended to expand access to program data.

To better understand the state of administrative data in the Administration for Children and Families (ACF), the Office of the Assistant Secretary for Planning and Evaluation (ASPE) partnered with the Office of Planning, Research, and Evaluation (OPRE) to produce the *Compendium of Administrative and Survey Data of the Administration for Children and Families,* available at aspe.hhs.gov/pdf-report/acf-data-compendium. The *Compendium* covers 19 survey and administrative datasets, providing basic descriptive information to inform future work, including content, accessibility, capacity to link with other datasets, presence of personally identifiable information (PII), use, documentation, and data quality. This brief focuses on the 12 administrative datasets, listed in Table 1.

Why Focus on Administrative Data?

The government already collects administrative data for programmatic purposes. Such data have significant potential as a cost-effective resource to inform research leading to better policy and program design.[3] The data often contain detailed program information that surveys cannot easily duplicate. Such data usually include records for all program participants (often millions), allowing for analyses of small subgroups and detailed geographic areas that are not possible with most survey data. Finally, many administrative data sources include PII, providing the potential to link records across multiple data sources and multiplying the capacity to answer important research questions.[4]

Administrative data are not without challenges and limitations in research applications. These data sources often lack broader social and demographic measures needed to answer important questions and support complex models. In addition, data not previously used for research purposes often have uneven quality and consistency. Administrative data often lack documentation to adequately inform research use outside of the agency. Finally, administrative data often have restricted access, and in many cases versions that would allow broader public access have not been developed.[5]

Data Characteristics Vary and Influence How the Data Can Be Used in Research

Every program office within ACF regularly collects and processes data intended to administer, implement, and monitor program operations. These data sources vary widely in the data fields they contain, accessibility for researchers, quality, and other factors. These characteristics influence the extent to which the data can be used for research and evaluation activities.

---

[2] Commission on Evidence-Based Policymaking. "The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking." Washington, DC: Commission on Evidence-Based Policymaking, September 2017. https://www.cep.gov/report/cep-final-report.pdf

[3] Office of Management and Budget. "Guidance for Providing and Using Administrative Data for Statistical Purposes." M-14-6.

[4] Fantuzzo, J., and Culhane, D. (Eds.). 2015. *Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government*. New York: Palgrave Macmillan.

[5] One of the goals of the Foundations for Evidence-Based Policymaking Act of 2018 (P.L. 115-435) is to encourage the development of more publicly accessible datasets from federal administrative data sources. https://www.congress.gov/bill/115th-congress/house-bill/4174

Table 1. ACF Administrative Datasets in *Compendium*

Dataset	Program	ACF Office	Populations
Adoption and Foster Care Analysis and Reporting System (AFCARS)	Title IV-E Foster Care	Children's Bureau	Children in foster care, children with finalized adoptions
Advocate Caller Application Database	National Domestic Violence Hotline	Family and Youth Services Bureau	Contacts made with the hotline
Child Care Information System (ACF-801)	Child Care and Development Fund (CCDF)	Office of Child Care	Families receiving subsidies
Debtor File	Title IV-D Child Support	Office of Child Support Enforcement	Individuals owing past-due child support
Federal Case Registry	Title IV-D Child Support	Office of Child Support Enforcement	People involved in child support cases, including children and parents
National Child Abuse and Neglect Data System (NCANDS)	Title IV-E Foster Care	Children's Bureau	Children reported to child protective services
National Directory of New Hires (NDNH)	Title IV-D Child Support	Office of Child Support Enforcement	Workers
National Youth in Transition Database (NYTD)	Title IV-E Foster Care	Children's Bureau	Youth exiting foster care and youth receiving independent living services
Refugee Arrivals Data System (RADS)	Refugee Resettlement	Office of Refugee Resettlement	Refugees and other eligible populations[6]
Residential Energy Consumption Survey (RECS)	Low Income Home Energy Assistance Program (LIHEAP)	Office of Community Services	Beneficiary households
Runaway and Homeless Youth Homeless Management Information System (RHY-HMIS)	Runaway Homeless Youth	Family and Youth Services Bureau	Runaway or homeless youth
TANF Data Reporting System	Temporary Assistance for Needy Families (TANF)	Office of Family Assistance	Beneficiaries

Source: *Compendium of Administrative and Survey Data of the Administration for Children and Families.*

---

[6] Other populations include asylees, Special Immigrant Visa holders, Cuban and Haitian entrants, and victims of human trafficking.

Personally Identifiable Information

Each dataset contains case-level data on individual persons or families, and five contain PII, such as social security numbers and dates of birth, which can provide matches to other data sources. For example, NDNH and TANF data contain PII such as social security numbers, dates of birth, and location, which can link data across these two systems as well as other data sources. In contrast, child welfare case data from AFCARS, NCANDS, and NYTD can be linked through unique identifiers that states provide, rather than PII. Because these identifiers cannot be traced directly to personal information, records can only be linked across these sources within the same state, not to records in other states or to other databases.

Geographic Information

All datasets but one (RECS) contain substate geographic identifiers, usually down to the county level. While state-level analysis can be useful, having data with lower-level geographic information permits a better understanding of geospatial differences in trends and can support more accurate modeling of local contextual influences. Linking geospatial data from other sources, such as those produced by the U.S. Census Bureau or the Health Resources and Services Administration's Area Health Resources Files, can also facilitate analysis. In addition, many ACF programs are administered by states, and states differ in how they implement programs and collect data. Distinguishing whether different results between states are due to program design, data collection, or actual outcomes can be challenging.

Data Access and Linking

The datasets range in how accessible they are to the public. The public can access aggregate data for seven datasets. Researchers can access restricted-use, case-level data for six datasets, including AFCARS, NCANDS, NYTD, NDNH, CCDF 801, and TANF data. ACF has established protocols for accessing AFCARS, NCANDS, and NYTD through the National Data Archive on Child Abuse and Neglect. Researchers can access restricted-access data through data archives, through federal research data centers, and occasionally directly through the agency that gathers the data. Even with restricted datasets, some information is withheld, suppressed, or altered to prevent reidentification of individuals.

Six of the datasets are or have been linked to other datasets, enhancing their utility for research and evaluation. AFCARS, NCANDS, and NYTD can be linked on common child identifiers. TANF and CCDF 801 data have each been linked to U.S. Census Bureau records for projects, although those links are not ongoing. RADS is linked to the Annual Survey of Refugees, serving as the main sampling frame for that survey data collection.

Data Quality

While the *Compendium* does not provide a systematic assessment of data quality across the datasets reviewed, it does include a general assessment of quality from those most familiar with the data resource, including practices in place to address quality issues such as validity, reliability, and item nonresponse. For many ACF datasets, data quality was judged highly overall. In many cases, practices are in place to promote data quality, including regular site audits and validation at submission. For example, the AFCARS Assessment Reviews provide formal assessment of the validity and reliability of data reported to AFCARS.

Administrative data can vary substantially in quality for a number of reasons. As is the nature of administrative data, fields used to manage programs, particularly those related to compliance or performance monitoring, tend to have the strongest quality. Fields that are more peripheral to program administration may have poorer quality. One limitation of many of the datasets is a lack of complete documentation, which is primarily the case for datasets not previously used for research. In addition, states or individual grantees commonly report data from their own administrative data systems, which can vary significantly in how measures are defined. For

example, data reported to NCANDS can vary because states use different definitions of child abuse and neglect.

Dataset Content Varies In Part Because of Differences in Program Authority and Collection Infrastructure

The variability in the content of ACF administrative datasets is primarily due to differences in program statutory and regulatory authority, as well as data collection infrastructure and arrangements.

Statutory and Regulatory Guidelines

All datasets are subject to different statutory and regulatory guidelines that govern data collection. These unique restrictions mean that changes in what data are collected require a lengthy approval process, making it harder for programs to adapt to changing data needs. Another consequence is that datasets covering different programs do not typically collect comparable information. In particular, statutes dictate what PII can be collected from program participants. Only five of the datasets collect PII, while for others ACF cannot collect such information. For example, RADS data is used to verify arrivals by using PII on refugees and other eligible populations in order to allocate funding to grantees for services. In contrast, ACF is not permitted to collect PII on families receiving child care subsidies as part of the ACF-801 data, per the reauthorization of the Child Care and Development Block Grant (P.L. 113-186).[7] As another example, regulations do not allow reporting of PII for the AFCARS, NCANDS, and NYTD datasets.

These statutory and regulatory rules also control the use of datasets for research. For example, the statute authorizing the NDNH strictly defines how data can be used for research. It states that the HHS Secretary may provide access to data "for research purposes found by the Secretary to be likely to contribute to achieving the purposes of [part A or part D of the Social Security Act], but without personal identifiers."[8]

Data Collection Infrastructure and Arrangements

Most ACF programs provide services through grantees. Often these grantees are states, such as in the case of TANF or foster care services paid for by Title IV-E funds, but sometimes the grantees are local communities or nonprofits, as in the case of Head Start and some grants from the Office of Refugee Resettlement for refugee assistance services. Because ACF rarely collects records directly from participants, programs depend on the administrative capacity of these grantees. Among other things, such capacity includes quality control to ensure collection and transmission of reliable and valid data. Records of TANF and CCDF recipients, for example, both contain information on employment status, but jurisdictions may interpret the precise time of a beneficiary's employment relative to the data collection date differently. For data submitted to NCANDS on children reported to child protective services and children subsequently identified as victims of maltreatment, states differ widely in their definitions of maltreatment and classification of victims.

While most ACF administrative data include records on all persons served, several programs allow states to submit representative samples instead. For example, TANF, CCDF, and NYTD allow states to submit samples, and many do. Sampling greatly limits the ability to follow individuals over time within a dataset and greatly reduces the number of cases that can be linked to other datasets.

---

[7] Before 2015, states reported PII. The CCDF Final Rule (45 CFR Part 98) of 2016 requires states instead to submit a unique identifying number that can link data for a family unit over time.
[8] 42 U.S.C. 653(j)(5).

HHS Is Taking Steps to Expand the Utility of Administrative Data Within Statutory Authority

Just as programs transform over time, so do data systems. And as policy priorities shift, so do needs for knowledge about program implementation and effectiveness. HHS is working to use data nimbly to improve programmatic and policy decision-making and will continue to look for opportunities to increase the use of administrative data for evidence-building activities.

Supporting Research Using Administrative Data

HHS is taking important steps to increase the usefulness of ACF administrative data to understand the effectiveness of human services programs and ensure accountability. HHS regularly leverages its existing capacity to use administrative data for program evaluation and performance enhancement. An example is the use of the Child Care Information System (ACF-801) to analyze the caseload patterns of families participating in child care subsidy programs. For example, one evaluation found that while some families receive subsidies continuously for a year or more, many families enter and exit subsidy programs multiple times.[9] This information helped policymakers when they recently enacted laws and regulations designed to reduce administrative burden during the eligibility certification and recertification processes.

ACF administrative data also prove valuable when policymakers estimate the costs and benefits of changes to new or existing programs. For example, HHS sponsors TRIM3, a microsimulation model that combines data from Census Bureau surveys and ACF administrative records to estimate how potential changes in eligibility affect the number of families eligible for assistance programs and tax credits. This same model allows HHS to respond to a congressionally mandated request to document the percentage of people who rely on government assistance for a significant portion of their income and how this share has changed over time.[10]

Linking Administrative and Survey Data

ACF has enhanced its administrative data by linking to other data sources. For example, HHS has linked ACF-801 data on child care subsidies with microdata from the Current Population Survey and the American Community Survey. This link, accomplished in collaboration with the Census Bureau, provides critical demographic and economic information about CCDF recipients.[11] ACF matches records from the TANF administrative data with the NDNH to document recipients' employment patterns before, during, and after participation in the TANF program.

Datasets without PII can be linked to other data sources with geographical identifiers. Datasets with substate geographic identifiers, such as the AFCARS and NCANDS data, can be combined with person-level administrative data records to relate ecological circumstances to client characteristics and outcomes. For example, recent work by ASPE has used county-level

---

[9] Swenson, K. 2014. "Child Care Subsidy Duration and Caseload Dynamics: A Multi-State Examination." U.S. Department of Health and Human Services. https://aspe.hhs.gov/system/files/pdf/77151/rpt_ChildCareSubsidy.pdf

[10] "Indicators of Welfare Dependence: Annual Report to Congress." https://aspe.hhs.gov/indicators-welfare-dependence-annual-report-congress

[11] Shantz, K. 2019. "Using Administrative Records to Evaluate Child Care Expense Reporting Among Child Care Subsidy Recipients." Census Working Paper SEHSD-WP2019-11. https://www.census.gov/library/working-papers/2019/demo/SEHSD-WP2019-11.html

substance use indicators to predict foster care caseloads.[12] In addition, AFCARS and NCANDS administrative records have been appended to the National Survey of Child and Adolescent Well-Being, combining the rich survey data with detailed administrative data and enhancing the utility of both for research.

HHS is also initiating work to explore how additional data linkages across self-sufficiency and related programs may create opportunities to learn about program participant outcomes, particularly after individuals exit programs. Increased data matching capacity could facilitate alignment of outcome measures across employment and self-sufficiency programs.

In many cases, even when federal statutes or regulations do not allow data linking at the national level, states can link data within their systems. For example, linking child welfare and Medicaid data may help states understand how services paid for by Medicaid relate to child welfare outcomes. NCANDS and AFCARS data cannot link to Medicaid records held by the Centers for Medicare and Medicaid Services. States, however, may link their own data for such research. Currently, ASPE and ACF are partnering on a demonstration project to support states in linking Medicaid and child welfare records to conduct research on how Medicaid behavioral health services relate to child welfare outcomes.

Another key area of HHS's efforts includes improving infrastructure to enhance and link datasets. For example, OPRE's TANF Data Innovation (TDI) project supports innovation and improved effectiveness of state TANF programs by curating and expanding the analysis of data already collected during program administration. Resources from TDI bolster ACF's effort to assess the strengths and limits of combining benefits data from the TANF Data Reporting System and earnings information from the NDNH to understand the employment of work-eligible adults in TANF households. ACF is also working with states to enhance the quality and use of data from TANF and related human services programs by providing technical assistance and pilot support through the TANF Data Collaborative.

Conclusion

Leveraging administrative data can increase our ability to understand the reach of programs, measure their impact on key outcomes, and address other policy and program questions that surveys or other data resources may not be able to answer. In the human services field, HHS has long analyzed administrative data to improve service provision and make informed policy decisions. This brief highlights characteristics of key administrative datasets that ACF manages and describes why they differ in important ways that influence how they can be used for research. We hope that these insights can inform policymakers as they seek to increase federal data capacity and implement the Evidence-Based Policymaking Act. Additionally, this information may be useful for analysts and researchers looking for cost-effective ways to answer research and evaluation questions.

---

[12] Ghertner, R., Baldwin, M., Crouse, G., Radel, L., and Waters, A. 2018. "The Relationship Between Substance Use Indicators and Child Welfare Caseloads." https://aspe.hhs.gov/pdf-report/relationship-between-substance-use-indicators-and-child-welfare-caseloads