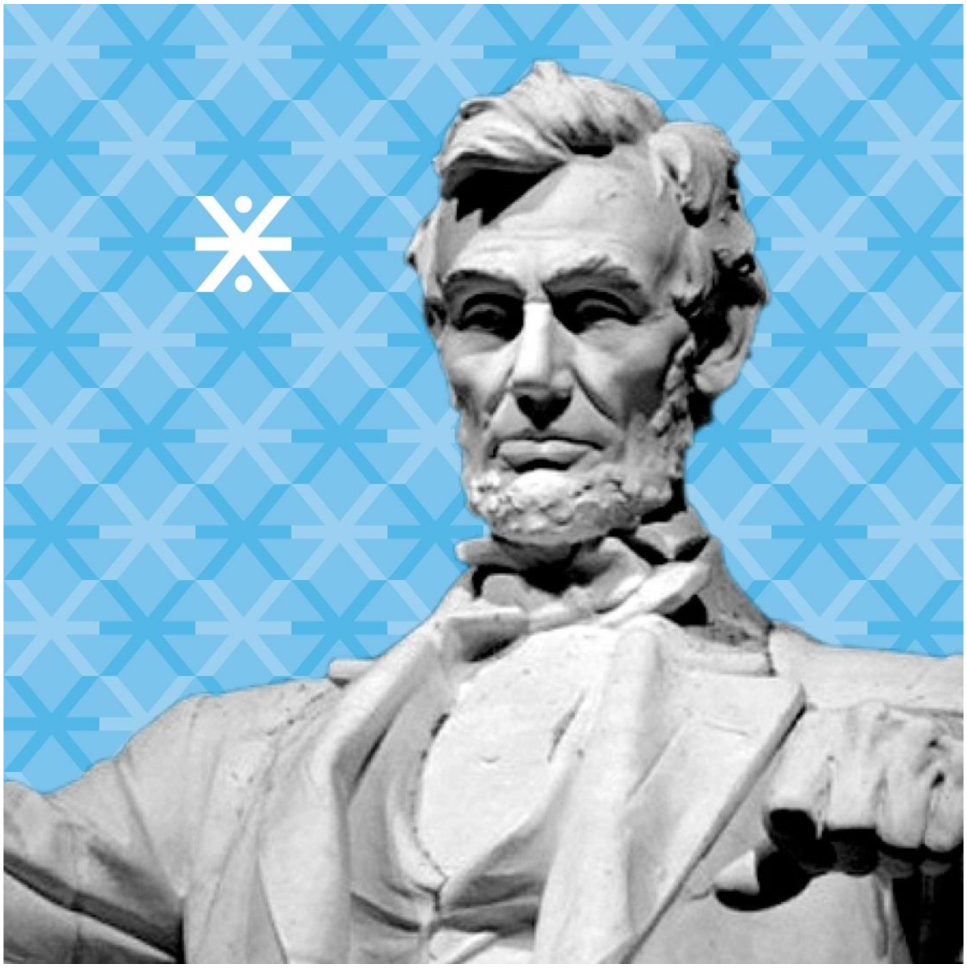
summit

complexity simplified.

Leveraging Administrative Data for Analysis and Evaluation

September 14, 2017



CONTACT

China Layne, Ph.D.
Manager, Data Analytics and Research
202.407.8300 Tel
509.984.8943 Fax
China.layne@summitllc.us

Anthony Curcio
Principal
202.407.8300 Tel
509.984.8943 Fax
Anthony.curcio@summitllc.us

AUTHOR

China Layne, Ph.D.

Table of Contents

EXECUTIVE SUMMARY 1

INTRODUCTION 2

GETTING STARTED USING ADMINISTRATIVE DATA FOR ANALYSIS AND EVALUATION 3

- **Review and Edit Administrative Data to Ensure Data Quality** 3
- **Link Disparate Data Sources to Create Integrated Dataset** 4
- **Leverage External Data Sources to Supplement Administrative Data** 4
- **Select Appropriate Analytic Methods** 4

ADVANCING THE USE OF ADMINISTRATIVE DATA FOR ANALYSIS AND EVALUATION 5

- **Consider Administrative Data First for Evaluation and Research** 5
- **Standardize Processes for Transforming Administrative Data into Research Datasets** 5
- **Consider Evaluation and Research When Designing Data Collection Systems** 6

CONCLUSION 6

ABOUT SUMMIT CONSULTING, LLC 7

Executive Summary

In the last few decades, federal and state agencies, as well as non-profit organizations, have increasingly used administrative data to conduct program evaluation and research. These agencies recognize that administrative data provide a more efficient and cost-effective alternative to primary data collection or external data. Several Federal agencies, including the U.S. Departments of Labor (DOL), Education (ED), and Health and Human Services (HHS) and the U.S. Census Bureau have been early adopters of integrating administrative data into evaluation and research. Seeing the direction of these changes, other organizations are asking if and how they can use their administrative data to support program analysis and evaluation.

Getting Started Using Administrative Data for Analysis

- Ensure Data Quality
- Integrate Disparate Sources
- Link to External Data Sources
- Select Appropriate Analytic Methods

In this white paper, Summit Consulting, LLC (Summit) discusses several best practices for how organizations can use their administrative data in analysis and research. The paper draws on our of experience conducting evaluation and analysis for several Federal agencies and our expertise with the administrative data of these agencies. We present the following four issues organizations should address as they begin using their administrative data in specific evaluation and research projects:

- Ensure data quality
- Integrate disparate sources
- Link to external data sources
- Use appropriate analytic methods

We also discuss the following three ways organizations can more fully integrate administrative data into their analysis and evaluation agenda:

- Consider administrative data first
- Standardize administrative data processes
- Factor analysis needs into data collection

With these best practices as a guideline, agencies and organizations of all sizes will be able to begin using their administrative in specific analysis and evaluation projects immediately. In addition, these guidelines can help organizations develop internal practices that will ensure future administrative data will support a robust research agenda.

Introduction

In the last few decades, the federal government has increasingly encouraged agencies to use administrative data to conduct program evaluation and research. Through multiple initiatives, such as the Program Assessment Rating Tool (PART) and Office of Management and Budget (OMB) memorandum M-10-01, the last two administrations encouraged agencies to conduct evidence-based assessments to guide budgeting and program implementation.[1][2] To support these analyses, the administrations stressed using administrative data as a preferable alternative to collecting primary data. Likewise, a bipartisan group of lawmakers established the Commission for Evidence-Based Policymaking, which promotes using existing data to conduct evidence-based research of government programs.

Several agencies and organizations have been early adopters of integrating administrative data into analysis and evaluation. From 1998 to 2012, an alliance of nine states maintained the Administrative Data Research and Evaluation (ADARE) project to provide long-series state-level administrative data for use in evaluation. DOL has promoted the use of administrative data to improve labor programs since at least 2010 through initiatives such as the Chief Evaluation Office's (CEO) Data Analytics Unit (DAU) and the Employee Benefits Security Administration's (EBSA) Data Integration (DI) project. This year, the Administration for Children and Families (ACF) at HHS is exploring the possibility of linking evaluation data to external administrative data sources to capture longer-term outcomes of the evaluated programs.

Type of Initiative	Sponsoring Agency	Detail
Single agency	Department of Labor	**Data Analytics Unit** focuses on short-turnaround analyses
		Administrative Data Research and Analysis project focuses on mid- to long-term evaluations
		Data Integration (DI) initiative linked data from several enforcement databases across 10+ years to enable long-term evaluation
Multi-agency	Alliance of nine states	**Administrative Data Research and Evaluation**[3] provided high-quality, long-series state-level administrative data for researchers to evaluate the impact of important federal programs
	Institute of Education Sciences (IES) and state and local education departments	**National Center for Analysis of Longitudinal Data in Education Research** leverages state and district administrative data files to examine the long-term effects of policies on students in a district or state.
Public–private collaboration	U.S. Census Bureau, Chapin Hall, and John and Laura Arnold Foundation	**Using Linked Data to Advance Evidence-Based Policymaking** is a grant competition to help states or localities pair administrative data with data from the U.S. Census Bureau to conduct evaluations of policies and programs.

[1] https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-01.pdf
[2] https://georgewbush-whitehouse.archives.gov/omb/performance/
[3] http://www.ubalt.edu/jfi/adare/index.cfm

In coming years, public agencies and non-profit organizations will likely rely even more on administrative data for evaluation and analysis. The change in executive branch administration will likely involve a shift in priorities for many agencies. In addition, both public agencies and non-profit organizations have experienced budget reductions and other resource constraints that may continue in the near-term. These circumstances will encourage organizations to look beyond primary data collection for programmatic analysis. As an alternative, administrative data have the following important advantages for use in analysis and evaluation:

- Because organizations already collect the data as part of managing their programs, using the data for analysis and evaluation is the **most cost-effective option**.
- As organizations collect a wide range of information directly about the programs (including participants, services provided, and outcomes), their data serves as the **most relevant source of information on their programs**.
- Unlike a one-time survey, organizations' administrative data is both the **most up-to-date source** of information on their programs and can provide longitudinal data on the programs.
- As organizations' administrative data includes information on both the program participants and providers, this data likely comprises the **largest sample of the relevant population** than any other single public or proprietary data source.
- With proper preparation and forethought, administrative data **can provide causal results in evaluation** similar to using primary data collection.

In this white paper, Summit discusses how both public and non-profit organizations can begin using their administrative data for program analysis. The paper draws on our experience helping several Federal agencies successfully leverage their data for analysis and evaluation. We discuss four specific issues organizations should address to begin using their data for analysis effectively. We also propose three solutions for organizations wanting to integrate administrative data into their analysis and evaluation agenda more fully.

Getting Started Using Administrative Data for Analysis and Evaluation

There are at least four issues agencies and organizations should address to effectively use their administrative data for analysis: data quality, disparate sources, external data, and analytical methods.

Review and Edit Administrative Data to Ensure Data Quality

Administrative data can have serious data quality issues, including duplicate cases, invalid values, incorrect formatting, internal inconsistencies, outlier values, and missing information. Some of these data quality issues (e.g. invalid values and incorrect formatting) can make the data difficult to use for analysis. More importantly, these data quality issues can compromise the accuracy and reliability of any analysis using the data (e.g. missing values, internal inconsistencies, and outlier values). To resolve this challenge, you should perform extensive data quality review and cleaning of the data before analysis.[4] At a minimum, the issues identified above should be identified and, where possible, corrected through data editing, imputation, or deletion. In cases where data quality issues cannot be corrected, you should flag the problem values in your dataset. All variables in the data should be reviewed for data quality, even variables you do not expect to have issues or use. Finally, any data quality review and cleaning work conducted should be thoroughly recorded in the analysis dataset (e.g. labels for revised variables

---

[4] https://www.summitllc.us/blog/improving-administrative-data-quality-for-research-and-analysis

and values) and accompanying documentation. This will ensure that future analysts can replicate the analysis dataset and results.

Link Disparate Data Sources to Create Integrated Dataset

Administrative data is often distributed across multiple tables or databases. In most cases, each database may have its own topic (e.g. participant demographics or participation information) and unit of analysis (e.g. participant or service record). In some cases, different databases may have the same information formatted differently (e.g. participant birthdate and participant age). Because of this, all information needed to conduct an analysis or evaluation may not be in one, single-format dataset. The disparate sources can make the data difficult to use for analysis, and differing formats of the data across tables can increase the likelihood of using the data incorrectly in analysis. To resolve this challenge, your organization should integrate the separate administrative data sources. You will need to know the unit of analysis and unique identifier for each table and how the various tables relate to each other. Once the tables are linked, you will want to review the variables and de-duplicate or standardize any duplicate information that may be coded differently.

Leverage External Data Sources to Supplement Administrative Data

Organizations should leverage external data sources to supplement their administrative data. External data can help administrative data support more in-depth analysis, provide more contextual information about the agency's regulatory or participant population, or enable a wider variety of research and evaluation topics. In recent studies for several Federal agencies, Summit has linked agency administrative data to external data sources such as Experian, STR Global, and the American Community Survey.

To best use external data in research, organizations should consider the full range of available data: administrative data from other organizations, public use surveys, and proprietary data collections. Likewise, organizations should ensure their administrative data is ready to link to common external data sources. This includes confirming the data includes commonly used unique identifiers (e.g. Employer Identification Number [EIN], Social Security Number [SSN]) and/or identification information in standardized formats (e.g. state and county information in Federal Information Processing Standard [FIPS] codes).

Select Appropriate Analytic Methods

Administrative data are generally observational rather than generated from an experimental design. As such, the program participants are not representative of the entire, target population. In observational data, participants are not randomly assigned to the program. They sometimes self-select into the program or are selected by the program for specific reasons. An analysis using this data may produce biased estimates of program impacts or other effects meant to be attributable to the entire population. To resolve this challenge, your organization should employ statistical methods to produce unbiased analysis results using observational data. Generally, the methods developed for quasi-experimental impact evaluation (e.g. propensity score modeling, regression discontinuity design) can be used effectively with administrative

Quasi-Experimental Methods for Administrative Data

- Regression Discontinuity Design
- Interrupted Time Series
- Propensity Score Modeling
- Exact Matching
- Multilevel Fixed Effects
- Synthetic Control

data.[5] As your organization considers administrative data for analysis, it will help to understand the source of bias in your data (e.g. self-selection into the program, selection for investigation) and any aspects of the program that may help support a quasi-experimental method (e.g. quantitative threshold for program participation, information on non-participants).

Advancing the Use of Administrative Data for Analysis and Evaluation

The previous four recommendations will help organizations more effectively use their administrative data for specific analysis projects. However, organizations should consider how to incorporate administrative data more fully into *all* their planned analyses. The following three recommendations will help ensure that the organization's data can support a full and robust analysis agenda.

Consider Administrative Data First for Evaluation and Research

Agencies should consider administrative data first and fully for all research and evaluation projects. Administrative data have many advantages over primary data collection. Moreover, we can use this data for a wide range of evaluation and research projects. However, if organizations only consider administrative data as a fallback option, they are less likely to use the data at all.

To determine whether administrative data is suitable for analysis, Summit uses a triage method developed by Dr. David Levine, Professor of Business Administration at University of California Berkeley. The triage method focuses on three questions about the administrative data:

- Is the data *large enough* to detect program effects and support robust analysis?
- Has the agency *collected the data long enough* to allow for multi-year analysis?
- Does the data include a *natural randomization or* information for a *potential control group*?

These three questions can help organizations determine if their administrative data can support evaluation and research. The organizations should consider primary data collection *only after* determining their administrative data is insufficient or unsuitable for research. Additionally, organizations could request researchers show in proposals or analysis plans that administrative data is unsuitable for the project before proposing primary data collection.

How Can Administrative Data Be Used?

- Full Program Evaluation
- Rapid Cycle Evaluation
- Basic Research
- Needs Assessment
- Predictive Modeling
- Business Intelligence Tools
- Prototyping
- Sampling Frame for Data Collections
- Public Use Files for External Researchers

Standardize Processes for Transforming Administrative Data into Research Datasets

Organizations should standardize and fully document the processes used to transform their administrative data into research-ready datasets. This work allows agencies to proactively establish best practices for working with administrative data. These best practices can ensure that researchers use the data appropriately, decrease the amount of time needed to familiarize themselves with the data and

[5] https://www.summitllc.us/blog/impact-evaluations-using-administrative-data-what-methodology-should-i-use

eliminate duplicative efforts of new researchers in working with the data. These benefits increase the utility of the agency's administrative data for research and evaluation.

Agencies can use the work of current research projects to develop, refine, and document best practice methods for using their administrative data. The organizations could request researchers provide a detailed summary of the data quality review and cleaning process used to create the research dataset, including a list of any data quality issues found and corrected. The project's final report could include this information as an appendix or supplemental document. For example, to conclude each research study, Summit provides a summary of the data quality review and cleaning process in the final report.

What Should Be Included in a Best Practices Manual?

- The data's collection process, structure, size, and unit(s) of analysis
- Any known data quality issues and their resolutions
- Steps for linking data internally and creating a flat file with important variables

Organizations can use the information submitted by researchers to develop a best practices manual for working with administrative data. The manual should include enough detailed and comprehensive information to enable a researcher unfamiliar with the data to easily create a research-ready dataset. The agency could distribute this best practices manual to all researchers at the start of new projects.

Consider Evaluation and Research When Designing Data Collection Systems

Agencies should work to integrate evaluation and research concerns when designing administrative data collection instruments and processes. This work will help ensure the agency's administrative data can support research and evaluation needs. Organizations should elicit input and feedback from researchers and experienced data users at each step of the data collection process: initial design, planned re-designs, and periodically over the course of the program's implementation.

At each point, input from researchers and data users will help organizations think proactively about how the data collection system can better support research and evaluation. For instance, this input can help organizations design instruments that better capture information important to evaluation, such as outcomes and program or participant characteristics. For example, last fall, Summit provided public comments on the research opportunities opened by EBSA's proposed changes to the Form 5500 questionnaire and collection process.

Conclusion

As an alternative to primary data collection, public agencies have increasingly used administrative data to conduct program evaluation and research. Several federal agencies have been early adopters of integrating administrative data use into evaluation and research. In this white paper, Summit discussed seven specific ways agencies and organizations can effectively and efficiently use their administrative data for analysis and evaluation. Summit drew on experience helping several federal agencies to propose four specific issues organizations should address as they begin to use their data and three recommendations for how agencies can more fully integrate administrative data into their analysis and evaluation agenda.

About Summit Consulting, LLC

Summit is a specialized analytics advisory firm that guides federal agencies, financial institutions, and litigators as they decode their most complex analytical challenges. Summit's staff of economists, econometricians, and research scientists use quantitative techniques to assist our clients as they model risk, evaluate program performance, and predict future performance.

At Summit, we solve complex analytical challenges with unparalleled customer service and extensive client collaboration. The solutions are complete only when they are understood by our clients and solve their problems. Our distinct capabilities include program evaluation, applied statistics and economics, mortgage finance, financial services, federal credit modeling and forecasting, and litigation analytics.

Summit hosts a solutions-focused academic environment and is dedicated to staying at—and pushing—the forefront of analytics best practices. To that end, our staff members present research at conferences worldwide and regularly partake in intensive in-house technical trainings. Our principals, academics, and research scientists are recognized experts in their fields, and they are capable of leading large and small solution teams.

www.summitllc.us