Determining the Feasibility of Using State Early Care and Education Administrative Data

CCEEPRA
Child Care and Early
Education Policy and
Research Analysis

Determining the Feasibility of Using State Early Care and Education Administrative Data

OPRE Research Brief #2017-17

February 2017

Submitted by:	Van-Kim Lin, MSPH, Kelly Maxwell, PhD, and Nicole Forry, PhD, Child Trends
Submitted to:	Ivelisse Martinez-Beck, PhD, Project Officer Office of Planning, Research and Evaluation Administration for Children and Families U.S. Department of Health and Human Services Kathleen Dwyer, PhD, Project Officer Office of Planning, Research and Evaluation Administration for Children and Families U.S. Department of Health and Human Services
Contract number:	HHSP23320095631WC
Project director:	Kelly Maxwell Child Trends 7315 Wisconsin Avenue Suite 1200W Bethesda, MD 20814

Determining the Feasibility of Using State Early Care and Education Administrative Data

The purpose of this resource is to help early childhood researchers determine the feasibility of using administrative data in their research. *Administrative data* refers to information about individual children, families, and/or service providers that is collected and maintained as a part of program operations. Administrative data can describe and inform the implementation of policies and programs. This resource provides questions for researchers to consider and discuss with agency leaders before finalizing a research plan that uses administrative data. This resource has been designed for use by researchers who are new to the analysis of administrative data, as well as seasoned users of administrative data who are expanding their research to include new types of administrative data (e.g., expanding to a new state or new agency).

The questions generated for this resource were developed through conversations with grantees and researchers who had experience analyzing state administrative data. Two other resources may be helpful to researchers interested in using early care and education administrative data: *Developing Collaborative Partnerships with State Agencies to Strengthen Research Using Early Care and Education Administrative Data* and *Considerations in Preparing to Analyze Administrative Data to Address Early Care and Education Related Research Questions.*

We have organized this resource into three sections, each of which covers a different set of questions to help determine the feasibility of using administrative data: 1) data stewardship and management, 2) data contacts and coordination, and 3) data usability. By presenting the three sections and the corresponding questions in this order, we do not mean to imply that researchers will confront

these questions sequentially. On the contrary, we expect an iterative process: asking some questions will prompt additional questions, with researchers needing to discuss multiple issues across the three sections simultaneously.

Why use administrative data?

Before delving into the main content, it is useful to review briefly why researchers may want to consider using administrative data. Administrative data can be used by agency staff for multiple purposes, including determining program eligibility or payments, generating standard reports for funders and federal agencies, and analyzing data to better understand whether the agency is meeting its goals. Researchers are often interested in analyzing administrative data because they include information that might be difficult for individuals to self-report (e.g., length of program participation) and because they are collected for an entire population (e.g., all users of a particular program). In addition, analyzing administrative data may be a cost-effective way to answer policy-relevant research questions, as it can reduce the amount of primary data collection needed.

Researchers can use administrative data in a variety of ways. For example, researchers may choose to rely solely on administrative data to answer policy-relevant research questions. Other times, researchers may supplement primary data that they collect by linking it with administrative data records or by using administrative data side-by-side with another source of data. In the first case, a linked primary-administrative dataset can provide researchers with additional variables to include in their statistical models. In the second case, one of the datasets can be used to contextualize or elaborate the results of the other. Researchers may be able to answer the questions of interest using data from a single administrative dataset or may need to link administrative data from multiple programs or agencies.

It is important for researchers to have a good understanding of the information included in relevant datasets to determine whether an agency has the data researchers are looking for, how (and if) researchers could access the data, and whether the data would be appropriate and sufficient to answer research questions of interest.

As part of this feasibility assessment, researchers should consider the **logistics and time** needed to access and understand the data. The amount of time needed to gather information about the feasibility of accessing and using administrative data varies, but could take several weeks or months. When considering a possible project, researchers should begin as early as possible. Particular activities that might take more time than anticipated include:

- receiving approval for using administrative data for research purposes;
- developing and approving a data-sharing agreement;
- receiving the data from the agency that houses or manages the data; and
- creating or clarifying information for a data dictionary or codebook.

The Perspective of a State Child Care Administrator

Leigh Bolick
South Carolina State Child Care Administrator
Division of Early Care and Education
South Carolina Department of Social Services

Using early care and education administrative data for research purposes has a number of advantages. State Child Care Administrators and other program staff are much more likely to be familiar with this data than with the process of setting up formal evaluations or research projects that can be more complex and far more costly. However, there are many issues that researchers must consider in determining whether a state's administrative data will truly be useful for study.

As stated in the article, states differ widely in how they are structured. In some cases, program staff are very well versed in working with administrative data and have a certain degree of autonomy in making it available (within the parameters of their agency's data-sharing agreements) to researchers and data analysts. In other states, the data may be heavily monitored by information technology staff or may even be located in another agency. Regardless of the state's structure, researchers can be assured that it is almost always best to approach program staff initially about potential research projects; they consider themselves to be the actual "owners" of the data and are therefore generally more receptive if they are consulted first. It is also very important for them to be seen as the experts of their own data throughout the research process.

The most important consideration may well be the quality of the data collected. It is important that researchers establish relationships with the program and information technology staff who regularly work with the data. If possible, researchers should try to meet with staff doing the data collection at the actual point of collection (often local or county staff) to become more familiar with how the data are gathered. It is also a good idea to meet with those involved in ongoing quality control assessments or data mining efforts to determine whether the researcher can be confident that the data quality is sufficient to be used for the intended purpose.

Finally, researchers should help the agency, to the extent possible, share the data while complying with ever-increasing concerns about confidentiality, security, and regulations established by policymakers in an era of data breaches. In the future, researchers may even be required to access the data on site at the agency using agency-assigned equipment.

To comply with the reauthorization of the Child Care and Development Fund (and with additional priorities set forth in other federal programs), states will be required to provide more data on the children and families they serve, improve accountability for their use of public dollars, and increase transparency to the public regarding the information they collect. This will promote an exciting environment in which research using administrative data sets can flourish, and state staff can pursue relationships to help them answer the questions that policymakers are determined to resolve.

Data policies and procedures

Identifying the potential administrative datasets that may be available to answer a given research question is a critical piece of assessing the feasibility of a research project that uses administrative data. Multiple datasets or sources may be explored before deciding whether, which, or how many dataset(s) are most appropriate for answering a research question. Although researchers may learn some information about the availability of data from reviewing public reports, they will also need to talk with state agency staff who know the program policies and data to learn enough about the administrative data to determine the feasibility of using it for research.[1] (For more information about identifying appropriate state agency staff, see *Section 2: Data contacts and coordination.*) This section offers some possible questions and considerations when regarding the stewardship and management of data, organized into four categories: data availability and stewardship, data sharing, data storage, and data security.

Data availability and stewardship

Questions for consideration: What data do state agencies collect? What data are reported to the federal government? What data reports does the state regularly generate? What additional data may be collected that are relevant to the question?

Before talking with agency staff, we suggest that researchers review agency websites and other online resources to learn as much as possible about the data available related to government programs of interest (e.g., child care subsidy program, TANF) as well as the supporting policies (e.g., the CCDF Policies Database provides an inventory of subsidy policies for all states and territories). Reports submitted to the federal government by state agencies may shed light on the types of information available in each administrative data set, though researchers should bear in mind that state agencies may collect more data than what is required to report to the federal government. Agencies may also regularly generate their own reports for state reporting or internal use. Reviewing these reports may help identify relevant datasets. Researchers will need to talk with agency staff to gather more detailed information about the data, particularly because most of the reports generated by agency staff provide aggregate data, and researchers will most likely want individual-level data. Rather than asking a broad and likely overwhelming question like, "What data do you collect?" researchers may want to ask staff for a brief overview of the various datasets that are related to the topic of interest. It may also be helpful to ask staff about the kinds of reports they regularly generate from their data. Staff may be comfortable describing these reports, but not sharing them, during these early conversations.

Examples of federal reports:

- Child Care Development Fund (CCDF) Quality Performance Report (QPR): http://www.acf.hhs.gov/programs/occ/resource/program-instruction-ccdf-acf-pi-2012-02

- Temporary Assistance for Needy Families (TANF) state annual report: http://www.acf.hhs.gov/programs/ofa/resource/policy/pi-ofa/2001/pi200106

Questions for consideration: Who or what agency is responsible for or authorized as the steward of the data? If another organization houses or manages the data, how are the data shared between them? How often are data updated?

One entity may be responsible for and authorized to steward data, but the data may be housed by another entity. For example, the state child care licensing agency may be responsible for collecting child care regulatory information, but the data may be stored electronically in another division or agency, like

---

[1] Cochenour, M., Corey, C., & Irvine, S. (n.d.). Early childhood data governance in action! An introduction. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/programs/slds/pdf/EC_DataGovernance.pdf

an information technology (IT) department or data management agency. In these instances, it will be important to understand the relationship between the two agencies and the procedures in place to access the data. The researcher will likely need to work with both groups to determine the feasibility of accessing data. Researchers also may want to find out how data are shared between the two groups. Specifically, researchers may want to inquire whether the agency housing the data recodes or reformats collected data when storing data in their system. It is recommended that researchers involve an individual from each agency or group to answer questions about the data.

Researchers will need to understand the frequency with which data are collected and whether historical data are overwritten or saved. If planned analyses rely on time-specific variables (e.g., the date a child enters the subsidy program, the date a program receives technical assistance), knowing the frequency and timing of data updates will be important. For example, a research team may want to examine change in quality over time for QRIS programs participating in a technical assistance project. To examine the relationship between the number of TA visits and change in quality, the research team will need to know how often the TA visit data are updated in the data system.

Question for consideration: Who can authorize the use of data for research purposes?

Because one or more agencies may oversee the data of interest, researchers will want to clarify whether one or more entities must approve the use of the data for research purposes. For instance, the agency authorized as the data steward may be able to approve the request for the use of data for research purposes. If one agency is the data steward and other agency houses the data, then they both may need to grant permission to use the data for research. Policies vary from state to state; it is important to clarify the permission process as part of determining the feasibility of using administrative data.

Question for consideration: How easy is it for agency staff to extract the data?

Because an agency's data may be housed or managed by another entity, the research team will want to know how easy it is for agency staff to extract the data and create research-ready data files. For example, it is possible that data management or IT staff manage the data and send reports to agency program staff. If it is difficult for agency program staff to access or extract the data, it may be difficult for researchers to access the data as well. If staff can extract the data, it will be important to understand how long it takes to process a request for a research file.

Data sharing

Questions for consideration: Has the agency ever shared, linked, or transferred data to another organization? How was the data shared, linked, or transferred?

While some states may regularly share data with outside organizations, others may not have any previous experience with data sharing. Researchers should ask about the state's previous data sharing efforts. If data have been shared, researchers may want to find out how that was accomplished (e.g., on a USB, or a secure data portal) in order to set up the systems needed to access the data in advance. Learning this information early can help the research team minimize delays. If data have been shared with researchers in the past, the current research team may want to ask those researchers for tips, lessons learned, and any data documentation they might have.

Questions for consideration: Does the agency have a data-sharing agreement template? If not, who in the agency would be responsible for developing one?

State agencies that have shared data previously may have a more streamlined process for sharing data with the research team than agencies that haven't. The research team may want to consider the time it takes to

set up a data-sharing agreement with a state agency, particularly for those that have not shared their data previously. It would be beneficial for the state agency and research team to identify the key individuals from both teams who will be developing, reviewing, and approving the data-sharing agreement.

Question for consideration: What information does the data sharing agreement require?

Agencies may vary in what they include in their data-sharing agreements, so the research team may want to clarify each component necessary for the data-sharing agreement to be approved. If an agency does not have a data-sharing agreement template, it may be useful for the researcher to provide a template or share an example of a data-sharing agreement. Potential components of the agreement might include:

- who is requesting the data (either specific individuals or an organization) and who needs to sign the data sharing agreement (e.g., University Institutional Review Board or Research Office);

- what data will be accessed;

- the purpose of the project;

- how the data will be used and reported, and whether anyone needs to approve or review the report before it become public;

- what federal protections apply to the data;

- how the data will be physically secured;

- how long the research team will have access to the data;

- how the research team will dispose of the data after the project is complete; and

- how errors, issues, or modifications to the data will be reported to the agency.

Researchers may want to ask someone in the agency to informally review the draft data-sharing agreement before submission, in order to minimize revisions.

Question for consideration: Do multiple data-sharing agreements need to be established to use the data?

If the datasets of interest are in multiple locations or managed by multiple organizations, multiple data-sharing agreements may be needed. Therefore, researchers will need to work with agency staff to identify all the required data-sharing agreements. Depending on the number of these agreements and the time expected to finalize them, the researcher may need to determine whether it is feasible to move ahead with the project.

Questions for consideration: In what format are data shared? Does the format allow for the data to be changed or manipulated by the researchers?

Researchers may want to ask the agency which format will be used to share the data (e.g., Excel, ASCII, CSV, SAS, or Stata file) so that the team can be sure that the data will be in a useable format or could be converted into a preferred format for statistical analysis. For instance, some purchased software programs only allow for individual printed reports to be exported from the software. In that case, printed sheets or electronic scans of the data can be shared. The printed/scanned data would then need to be re-entered by a research team into an electronic format to be analyzed. As another example, if the data are shared with a research team in an Excel spreadsheet, the research team may need to ensure that the statistical package they have can utilize data stored in an Excel file. Otherwise, the team may need software to transfer data into a statistical package in order to analyze it (e.g., StatTransfer). Researchers will also want to ensure that

their statistical package allows for the number of observations in the dataset to be transferred, because state databases may be quite large.

Questions for consideration: What technology should researchers have to receive the data securely? Will researchers need to set up a secure file transfer protocol (SFTP) or something similar in order to receive the data?

Researchers may be asked to receive the data through secure means. Sometimes a password-protected file sent by email is sufficient. Other times, they may need to set up specific technology, such as a secure file transfer protocol (SFTP), in order to receive the data. Researchers can talk with experts such as IT support staff, data security experts, and Institutional Review Board staff to ensure that they have the appropriate procedures in place for receiving and storing the data.

Data storage

Question for consideration: How are the data stored (e.g., paper or electronic)?

Researchers may want to inquire about how data are stored before determining the feasibility of using the data for planned analyses. If data are stored electronically, the research team will want to know whether the electronic form of the data can be transferred to or used with systems available to the research team. For example, the system that stores state-level data may have a larger memory capacity than that of the statistical program the researchers use to transfer or analyze data. Therefore, the researchers may need to make decisions about how to transfer and analyze the data, whether by dividing the data into smaller datasets (which can take time), choosing to upgrade their statistical software's memory, or other similar measures. If the research team cannot find a solution, the data may not be useable for planned analyses. If the data are stored on paper, the research team will need to consider the time and effort needed to enter data into an electronic form to be analyzed for research purposes. Coding data from paper forms may also be time-intensive for researchers, who will most likely need to review and code the paper forms in the agency's office.

Question for consideration: How should the data be securely stored when shared with another organization?

The organization that houses or manages the data may have rules and regulations for how the data should be kept secure when shared with external users. Researchers should ask if the agency has data security and data sharing policies and, if so, should request copies to review. It is also possible that the agency does not have data security policies. Researchers should be proactive in identifying data security issues to address with agencies and offering to create draft data security policies.

Data security

Question for consideration: Does the dataset of interest include any personally identifiable information (e.g., social security number, birthdate, address)?

Researchers may want to consider whether they need any personally identifiable information (PII) data elements to complete their planned analyses and determine any IRB requirements or data protection challenges that this may present. If researchers do not need PII to address their questions of interest, they should not request it in a dataset. If, however, a researcher needs PII data and the state is considering sharing the information, then it may be useful to understand the various types of PII data available. Having more than one piece of personally identifiable information for an individual can be helpful when linking multiple datasets. For example, different programs or agencies may use different IDs, which may require the researcher to use PII to match the individuals across the multiple datasets. Having two pieces of information, like birthdate and full name, would make it easier for the researcher to confirm the identity across datasets.

Question for consideration: What level of security is needed for particular types of data?

Agencies may keep certain data secure at different levels of accessibility, so it is important for the research team to understand the security structure of the data in order to understand what processes they may need to follow to access it. For instance, data that does not have any personally identifiable information (PII) may have a less stringent requirement to access than data with PII. Researchers will want to consider the security level of the dataset of interest in order to anticipate the number of restrictions or the amount of time it may take to access the data. Additionally, the research team might also be able to adjust their analyses in order to use a less secure but more easily accessible dataset.

Question for consideration: What state or federal protections (e.g., FERPA or HIPAA) must be considered?

Federal regulations also add a layer of protection to the data the researchers want to access. For example, health data are subject to restrictions of the Health Insurance Portability and Accountability Act (HIPAA), and education data are protected by restrictions laid out by the Family Educational Rights and Privacy Act (FERPA). Researchers will want to ensure that they adhere to any state or federal regulations that apply to the data.

Question for consideration: How is the confidentiality of data maintained when data are exported or shared with others?

The agency may have rules and regulations about how data with personally identifiable information (PII) can stay confidential when shared with external users. The research team will need to make sure that they understand how data must be kept confidential, both in storage and in reporting. For instance, some agencies do not allow data that has less than five individuals in one subgroup to be shared with external users because the users could more easily identify these five individuals. Understanding the procedures to maintain confidentiality may help the research team determine whether the data are appropriate for the question or analysis plan.

Question for consideration: Does the dataset of interest include a unique identifier (e.g., case IDs, child IDs, family IDs)?

To protect the identity of program participants, agency staff may assign a unique identifier to programs and individuals. Researchers will want to determine whether a unique identifier exists and, if so, whether it can be used to match individuals or programs across multiple datasets. For instance, a child may be assigned a unique identifier that follows her across different state services (e.g., subsidy, or TANF). If the datasets do not include unique identifiers, then the research team may need to request to use personally identifiable information to match individual cases across multiple datasets.

Data contacts and coordination

To be successful, a researcher must be familiar with the program staff most knowledgeable about the data. Researchers should identify the appropriate staff and determine how best to coordinate with these individuals to access and understand the data. We recommend that researchers speak with both data managers and program staff who are knowledgeable about the data to understand the specific data elements as well as the policies/procedures that may affect data collection and management, potential data quality issues, and interpretation of findings from the analysis of administrative data. A state administrator or department leader may be helpful in pointing the researcher to the appropriate staff. These leaders may also need to give their approval for agency staff to work with the research team to determine the feasibility of a particular project.

In some instances, one staff member is knowledgeable about both program operations and data management. In other cases, researchers may want to consult with both program operations and data management specialists. After identifying the individuals who can help inform the planned analyses, researchers will want to determine how best to coordinate ongoing communication with these individuals, to respect their limited time and minimize interruptions while getting information needed for the research project.

Identifying contacts

Question for consideration: Which individuals are most familiar with how the program operates at a state and local level and are available to answer research-related questions?

There are a few different considerations in determining who will be most helpful in answering questions about the program or population of interest. First, is there one individual in an agency who can answer all the questions about the program, or are there multiple individuals who each focus on different aspects of a program's operations? If there are multiple individuals with different foci, is there a regularly scheduled meeting with these individuals that the researcher could join if he/she had cross-cutting questions about the program's operations? Second, is the most knowledgeable person available to answer research-related questions in a timely manner? If the person is very busy, can he/she delegate someone else who could be the contact for the research project?

Question for consideration: Who manages the data?

It is very helpful to partner with a data manager who is knowledgeable about how the data are stored and managed, the quality of the data, and the meaning of specific variables in the dataset. Data managers have unique knowledge and expertise regarding the storage, management, and reporting of data. Some data managers may also be knowledgeable about the program operations and relevant policies, but others may not be. Researchers may need to talk with both data managers and program staff to fully understand the data and the program/policy context. Research teams will need to determine on a case-by-case basis which individuals are the best contacts for a particular project.

Coordination

Questions for consideration: How do the individuals who are knowledgeable about the program work with the data managers? Are they in the same agency? Are they in the same building? Do they have regularly scheduled meetings?

Coordination among the research team, data managers, and program managers can help ensure that everyone has a shared understanding of the research questions and data used to answer the questions. If multiple data are collected and managed by two different groups or individuals, the research team may want to ask about joining an existing meeting to discuss research-related issues so that it is as efficient for staff as possible.

Question for consideration: How can researchers reduce the burden for data managers and program administrators who are helping with the research project?

Agency staff must balance multiple responsibilities, so it may be difficult for them to commit the time needed to support a research project. Before beginning a project, it may be useful to discuss with staff how regular communication could occur between research, program, and data management staff; how quickly program/data management staff would be able to respond to research-related questions throughout the duration of the study; and how the research team could support the program staff (e.g., share with them a data dictionary developed for the research project that could be used by the state agency. For

more information about how to build effective relationships with program administrators, see *Developing Collaborative Partnerships with State Agencies to Strengthen Research Using Early Care and Education Administrative Data.*) Researchers also should not assume that program or data management staff will be familiar with specific analytic methods and should avoid using research jargon.

Question for consideration: Who is the point person for the agency and research team?

Selecting a point person for the state agency and the research team is important so that communication flows efficiently between the two groups. If multiple people hold expertise in different areas, a point of contact for each area of expertise can help facilitate the communication process. In some instances, there may be multiple contacts for various aspects of the project (e.g., one point person to approve the research project and another to approve the data security policies).

Data usability

In addition to identifying datasets of interest and key contacts in the agency, researchers will want to learn as much as possible about the data they would like to analyze before moving forward with a research project. The research team might consider how the data are collected, stored, and secured; the availability of information about the variables in the data; and how others may have used the data.

Data collection and quality

Question for consideration: What data were collected and how were they collected?

The methods and purpose of data collection can affect the appropriateness of the data in answering research questions. Administrative data are collected primarily for programmatic, rather than research, purposes. Thus, data collected for program operations may not include all of the information needed to address a particular research question. For example, a researcher may wish to have family income as a continuous variable to use in analysis, but might learn from agency staff that they ask families to identify income only by category (e.g., < $20,000, $20,000-$50,000) because the program is interested in eligibility above and below a certain threshold. When information is not routinely used for program operations, those who collect data (e.g., caseworkers, licensing consultants) may not record information correctly or consistently. For instance, family demographic characteristics (e.g., home language, country of origin) can be important in addressing a research question related to service usage in specific sub-populations. If program staff do not regularly use or report these characteristics, they may not always ask families these questions or consistently record the information. We recommend that researchers clarify how data are collected, to help determine the feasibility of using administrative data for a particular research purpose.

Questions for consideration: Do local programs report to the state agency? If so, what data are collected at the local level? How are data reported to the state? How are data aggregated at the state level?

Information in state-level datasets is often compiled from local programs or local/regional government agencies. Thus, researchers may want to determine the data that states collect from local programs, how local programs report information to the state, and how data are aggregated at the state level. This allows researchers to understand how the state administrative dataset was created. They may also want to ask program staff their perceptions of the quality of the data gathered at the local level, as it may affect the quality of data at the state level. Researchers may also want to explore what data are stored at the local vs. state level, as this may affect how a research team can access the data for planned analyses.

Questions for consideration: Were any processes in place to ensure the quality of data collection and entry? If so, were any data quality concerns identified in this process?

Researchers may ask about whether processes or protocols are in place to ensure that the data collected were checked for accuracy. State agencies or local programs may have training manuals, how-to guides, or instructions for data collectors that researchers can review. If these protocols are not in place, the research team may want to ask program administrators about known issues with the data collected. If, for example, staff report that data about the reasons families leave the subsidy program are unclear because the definitions of response options have changed over time, then researchers may not want to include this particular variable in their analysis.

Researchers may also want to ask questions to understand the steps taken to ensure accuracy in data entry. If someone enters information from paper forms into an electronic data file, are a percentage of data verified by someone who did not enter the data? If data are entered directly into an electronic data system, does the computer flag numbers that are out of range (i.e., higher or lower than would be expected)?

Appropriateness of data for the research question

Question for consideration: What is the appropriateness of the administrative dataset for the research question?

Preliminary discussions about data may suggest to a researcher that particular research studies could be conducted using administrative data. It is important to revisit this issue, though, as the researcher learns more about the data. In considering this question, the researcher will need to determine whether a) the data are appropriate for the question and design; b) the questions or design could be modified to enable the use of administrative data; or c) the data are not appropriate for the question or design and the project should stop. Box 1 includes some examples of questions that could be addressed with administrative data.

Box 1. Examples of policy-relevant child care and early education questions that could be addressed with administrative data.[a]

- **Characteristics of child care subsidy recipients:** Which child care subsidy recipients use subsidies to help cover the costs of child care so that they can look for a job (considering characteristics such as income category, county, maternal age and education level etc.)? Are certain groups of subsidized families (e.g. immigrant families, families of children with special needs, families with the lowest incomes) more likely to use legally unregulated providers?

- **Family outcomes related to child care subsidy policies:** What percentage of families leave the subsidy program around the time of a required reporting event or redetermination, even if they are still eligible? Are subsidy recipients' incomes higher after engaging in a supported job search than in their previous job?

- **Access to high-quality child care among subgroups of children and families**: Are there similar percentages of infants and toddlers enrolled in high-quality child care as compared to preschoolers? Are there county or city variations in access for infants and toddlers?

- **Outcomes before and after a change in policy or practice:** What outcomes (e.g., increased use of high-quality care for low-income families) are associated with recent policy or practice changes (e.g., increasing the redetermination length for subsidy recipients)? Comparing trends over the 3-year period before and after a state implements tiered subsidy reimbursement, do more high-quality providers participate in the subsidy program when they have access to higher tiered reimbursement rates?

- **Improving the quality of care:** What types of professional development supported by the Child Care and Development Fund (CCDF) are associated with improved quality as measured by Quality Rating and Improvement System (QRIS) ratings? What are the characteristics of programs that improve their QRIS rating within a 3-year period?

- **The link between participation in early care and education programs and positive child outcomes:** Do children who attend high-quality early care and education programs for multiple years have greater growth on measures of children's development than do children who attend low-quality early care and education programs for multiple years? (Note that questions like this may require combining administrative data with new data collected on a sample of children and families.)

---

[a]The example questions are adapted from research questions answered by Child Care Research Capacity Cooperative Agreement grantees (http://www.acf.hhs.gov/programs/opre/research/project/child-care-state-research-capacity-cooperative-agreements-2001-2011) and questions developed by CCADAC for the Urban Institute's Implications of Child Care and Development Block Grant Reauthorization for State Policies brief series (http://www.acf.hhs.gov/programs/opre/resource/implications-of-child-care-and-development-block-grant-reauthorization-for-state-policies).

Question for consideration: Does the research question and design take into account the unit of analysis available in the administrative dataset?

It is also important to identify the unit of analysis (program, family, child) and ensure that the data are available at the unit of analysis. If, for instance, the unit of analysis of interest is child, then it will be important to ensure that variables such as family income or program quality can be tied directly to a child for analysis purposes.

Availability of information about the data

Questions for consideration: Is there a data dictionary or codebook available for the data? Are there variable names, descriptions of the variables, and values for the variables?

Planned analyses may be feasible without a data dictionary or codebook, but having these documents can help reduce the amount of time that the research team talks with program staff to understand a dataset or to analyze the data. State agencies may not have these documents, so the research team may need to develop a data dictionary, codebook, or user guide for their own research purposes, which can take time. The research team will want to consider whether the lack of documentation may hinder their ability to complete analyses in a timely manner. As part of the partnership with the state agency, the research team could share with agency staff any documentation it develops (e.g., data dictionary).

Questions for consideration: Were there any program-level modifications in the data during the time period of interest? If so, what are they? Is there documentation for how data were modified?

Because changes to data can influence how the data are used for research purposes, the research team will want to make sure that they understand how the data have been modified, if at all. Without a clear understanding of data modifications, planned analyses may not be feasible. For instance, a research team may want to include in its analysis information about pre-K teachers' professional development. The information available is a categorical variable that indicates whether the pre-K teachers met the professional development requirement. The requirement, however, may have changed over time. If the research team is using data from multiple years, the change in requirement changes the meaning of that particular variable.

Data modifications may be found in a user guide or in another form of documentation. If not, the research team may want to consult with program administrators or data managers about how data have been modified. In some instances, the research team may want the data in the raw, original form in which it was collected instead of in its modified form. Knowing how the data were modified, the research team can either request the original data or determine a way to alter modified data back to the original form.

Use by others

Questions for consideration: Has any other organization or agency used these data for analysis? If so, in what way? What research questions were answered? Are there available resources, reports, or briefs that could be used to inform the current analyses?

It may be useful to ask the agency staff if any previous external researchers have analyzed administrative data. If so, then the agency's previous experience with partnering on a research project may make it easier to use administrative data for a new research partnership. It may also be useful to talk with the previous research team to learn more about their experiences and how they analyzed the data. Previous users of the administrative dataset can help to provide tips for how the data can be used best for research purposes. Reading earlier reports may also help inform the current research efforts.

Summary

Researchers who are interested in using administrative data to address early childhood research questions need to consider various issues, such as the ones highlighted in this resource, to determine the feasibility of the project. This initial step in the research process will take some time—likely weeks to months—but is necessary before moving forward in understanding, accessing, and using administrative data for research. The questions included in this resource are meant to be a starting point, not a comprehensive list, to help researchers understand more about the administrative data they would like to use for research purposes, in order to determine the feasibility of a research project and to minimize the challenges they may face. Researchers interested in learning more about developing partnerships with state agency leaders may want to read *Developing Collaborative Partnerships with State Agencies to Strengthen Research Using Early Care and Education Administrative Data*. Researchers who have determined that it is feasible to use administrative data may be interested in reading *Considerations in Preparing to Analyze Administrative Data to Address Early Care and Education Research Questions.*