Data Competencies Public Policy Students Need to Know

Cynthia Osborne, Ph.D. & Laura Bellows, MPA
Child and Family Research Partnership
LBJ School of Public Affairs
University of Texas at Austin

APPAM Spring Conference: Washington DC

CHILD & FAMILY
RESEARCH PARTNERSHIP

Motivation

- The Child & Family Research Partnership specializes in making rigorous policy research relevant for policymakers and practitioners

- We regularly partner with state government agencies to conduct evaluation and research work

- Large-scale administrative and survey data drive the bulk of our work

- Knowledge of how to use state and federal data is increasingly important given the focus on evidence-based policy making

CHILD **&** FAMILY
RESEARCH PARTNERSHIP

Top Five Data Competencies

In addition to rigorous **statistical analysis** and a basic understanding of **statistical programs** and **database management**, we propose that all public policy students should learn the following five data competencies:

1. Collect data for decision-making
2. Investigate data quality
3. Understand data organization
4. Learn about combining data
5. Keep data secure

CHILD & FAMILY
RESEARCH PARTNERSHIP

Framework

For each data competency we:

- Identify the "problem;"

- Use an example from our own experience to demonstrate the necessity of the competency;

- Discuss the challenges and needs to teaching this information to students;

- Provide recommendations

CHILD & FAMILY
RESEARCH PARTNERSHIP

Collect data for decision-making

- **Problem:** Data points necessary for accountability or evaluation research are not always available in administrative databases. Data are frequently collected for compliance; states are often limited to collecting what is legislatively mandated.

- **Examples**:
  - The Texas Child Support Division collects the amount of child support owed and paid by obligors, but does NOT collect the income of either parent.
  - Texas education only recently collected a student-teacher link

- Adding new data points is expensive and often necessitates negotiations among many partners

- Without the data, accountability and decision making are incomplete

CHILD & FAMILY
RESEARCH PARTNERSHIP

Collect data for decision-making

Recommendations:

- Students must be trained to understand the theoretical framework that guides the research and informs the policy decision making. Theory guides the need for specific data points.

- Simply knowing statistical methods is not enough for rigorous policy research.

- As students move into positions of authority, they can push for adequate measures to be captured.

CHILD & FAMILY
RESEARCH PARTNERSHIP

Investigate Data Quality

- **Problem:** Administrative data quality can be poor, especially in newer systems. For decisions to be made with the data it must be clean.

- **Examples:**
  - Education data linking teachers and students is inaccurate.
  - Home Visitors: The same home visitor (ID 43) appears to work for two different home visiting programs in three different counties.

Visitor ID	Program ID	County	Program
43	18	Cherokee	Parents as Teachers
43	22	Dallas	Parents as Teachers
43	26	Nueces	Parents as Teachers
43	2	Nueces	Early Head Start

Source: Texas Home Visiting Data System, 3/31/2014

CHILD & FAMILY
RESEARCH PARTNERSHIP

Investigate Data Quality

- Programmers do not understand constructs being measured; important to pair someone with substantive knowledge with programmers, and to include both in the quality control

- For most research, academics can "toss" bad data and outliers

- For accountability and data-driven decision making, all data must be included

- Students should learn to treat all administrative data with skepticism

- Recommendation: Rather than have students work with clean analytic files, challenge them to find issues with "dirty data"

CHILD & FAMILY
RESEARCH PARTNERSHIP

Understand Data Organization
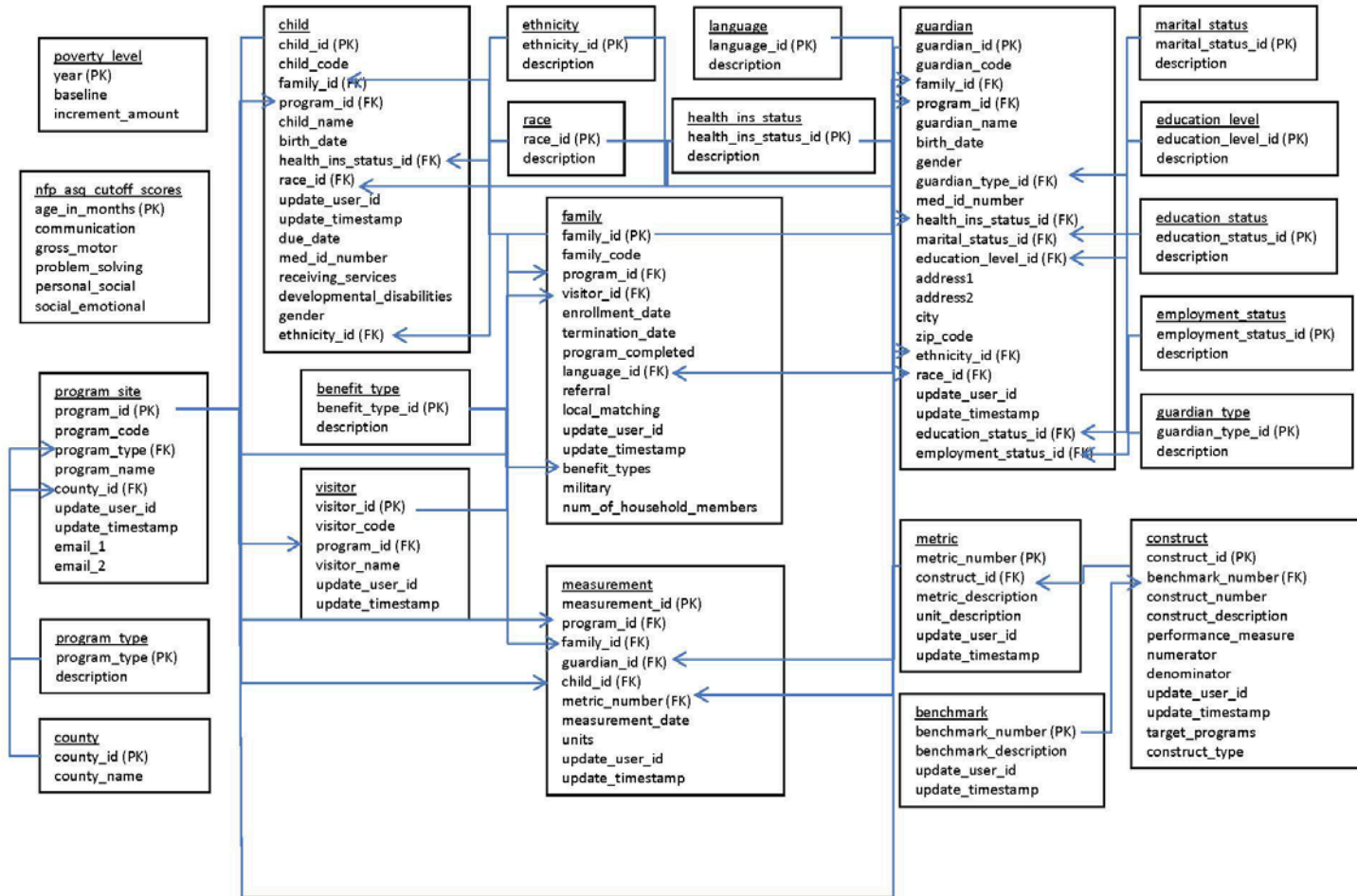
- **Problem:** Data are not clean and are generally stored in relational databases. Almost all of our methods courses use clean "flat" files for data analysis.

- **Examples:** Below

CHILD & FAMILY
RESEARCH PARTNERSHIP

Understand Data Organization

Received from TEA to describe the relationships between some of their data tables.

Understand Data Organization

Schema of HHSC Database

Understand Data Organization

- Not all students need to learn SQL and database management, but students need to understand the layout and shape of databases and how they are updated and merged for analysis.

- Understanding the shape of the data informs analysis for decision making

CHILD & FAMILY
RESEARCH PARTNERSHIP

Learn About Combining Data

- **Problem**: Individual datasets are insufficient for data-driven decision making. Data driving decision-making means using *all* available data and breaking down silos across agencies:
  - Texas Home Visiting and DFPS to find child maltreatment
  - Child Support and Work Force to identify employment
- Combining data *within* an agency also presents challenges:
  - Texas Education Agency has different student identifier in each dataset

CHILD & FAMILY
RESEARCH PARTNERSHIP

Learn about Combining Data

- Ideally each data file has an identifier that is unique to the person and is well known by that person (e.g. SSN). This permits the same identifier to be used across multiple agencies.

- Yet, this approach raises serious privacy and security concerns

- Within an organization, one identifier should be used that is only meaningful to that organization (e.g. UTID) and can be matched securely

- Recommendation: require students to learn merge or join techniques and link data from multiple sources – ideally using "fuzzy match" techniques

CHILD & FAMILY
RESEARCH PARTNERSHIP

Keep Data Secure

Data security is the single most important aspect of research! Treat all data as if it is an SSN

- Identifiable data are only available to a limited few with controlled access

- De-identified data stored securely with strict data protocols (no data are completely de-identified)

- Require training for all staff and students, even those who do not access secure data

- Remind students that data represent real people

- Set data security standards for the most sensitive data in a dataset

- Understand tension between security and access

15

CHILD & FAMILY
RESEARCH PARTNERSHIP

Conclusion

- Data are increasingly necessary for research and decision making

- Students need to understand the challenges and responsibilities that accompany using data – as well as the techniques that foster quality research

- We are educating the next generation of leaders who can make smart decisions that allow data to drive decision making and inform policy

CHILD **&** FAMILY
RESEARCH PARTNERSHIP