

# E1: Utilizing Administrative Data Outside of Early Care and Education to Address ECE Policy Questions

Thursday, April 18, 2019

1:30 p.m. – 2:45 p.m. | Wilson/Roosevelt

## 1. Descriptive Information

### E1: Utilizing Administrative Data Outside of Early Care and Education to Address ECE Policy Questions

This session highlights various types of data outside of early care and education (ECE) that can be linked with ECE data to address policy-relevant questions. The session panelists will describe the research questions asked, methods used to analyze the data, and how the findings informed policy. They will also offer tips in linking various administrative data. The session will begin with an overview of administrative data outside of ECE that could be linked to ECE data to address various research questions. The first panelist will briefly describe work in Florida to link multiple types of administrative data to address questions of interest to state and local ECE policymakers. The second panelist will briefly describe work in Philadelphia to link non ECE data about children's risk factors with pre-K program data to help administrators expand pre-K in areas with high concentrations of children with multiple risk factors. The moderator will engage panelists and participants in a discussion of issues in linking data to address ECE research questions.

#### Facilitator

**Kelly Maxwell**, Child Trends

#### Panelists

**Kelly Maxwell**, Child Trends |  
*An Overview of Non ECE  
Administrative Data to Link with ECE  
Data to Address Policy-Relevant  
Questions*

**Roland Estrella**, University of Florida |  
*Florida Early Childhood and Health  
Data Repository: Improving Services  
with Birth to Five Data*

**Katherine Barghaus**, University of  
Pennsylvania | *Connecting Early  
Childhood Data to Health and Human  
Service Integrated Data Systems to  
Inform Policy and Practice*

#### Scribe

**Nazran Baba**, ICF

## 2. Documents Available on Website

1. CCADAC Resources handout, Maxwell
2. Utilizing Administrative Data

## 3. Brief Summary of Presentations

This talk will be about the various ways to use early childhood administrative data

- Child Care Administrative Data Analysis Center (CCADAC)
- Definitions
  - Administrative data: talks about information about children, families and service providers collected and maintained as part of regular program administration
  - Integrated Data System: System that combines administrative data across multiple programs and over time
  - Other uses of data beyond childcare; experiences using a range of administrative data; K-12, health, unemployment data, property tax records etc.
- We use certain kinds of administrative data outside of early childhood education that can be linked to early childhood data to address various research questions.

### • Summary of Presentation #1: Roland Estrella

- Florida Early Childhood and Health Data Repository
  - Working on an early childhood and health and data repository that tracks data on children and their parents from the time of conception through 6 years of age.

- Physical resources are needed to set up an environment like this. We had to work with the data owners and show the owners what would be involved. Strong communication channels need to be in place with data owners to build trust.
- We used national standards for our database administration –Federal Information Security Modernization Act (FISMA). This and what it involves has been communicated to the state owners.
- A lot of partnerships are needed to get the resources in place, buy servers, certify people and then to have conversations with each of the state agencies.
- Ultimately, we produce the data that can be used by the legislature. This results in multiple data sharing agreements that constantly need to be renewed and run through by various organizations such as IRBs.
- After all the infrastructure is in place, then we go into the fun stuff – the data.
- When we do longitudinal studies, we can link the mom with the child and follow the family across all the programs, and across the life course perspective. This will help us connect events that happen early in life with outcomes later on. This is the conceptual framework of the repository. This has been the result of a 20 year partnership to get vital records. The system will include data like:
  - Health screenings
  - Social services
  - Medicaid enrollment
  - Hospitalizations
  - Paid claims and encounters
  - Special medial services
  - Birth anomalies
  - Mental health
  - Child development
- ECE data sets include:
  - Child care providers
  - Children with disabilities
  - Homeless children
  - Subsidized child care
- When we build a repository of this nature it must be set up as a system that can handle a recurring intake of data.

- **Summary of Presentation #2: Katherine (Katie) Barghaus**

- Actionable Intelligence for Social Policy is an integrated data network. This was developed in Philadelphia.
- Context: there is an effort for Pre-K expansion in Philadelphia. The mayor allocated 60 million dollars to expand high quality Pre-K and we needed to identify how to strategically use this money to get the most out of it.
  - Research questions
    - What geographic areas have the most vulnerable kids?
    - What geographic areas have the highest quality of care?
    - To identify the areas we had to overlay the greatest need and smallest supplies zones to identify places to expand Pre-K.
  - How did we tackle it?
    - Fortunately, the data systems we needed were already set up in Philly.
    - Having longitudinal data models with all the relevant data elements is helpful for predicting school readiness.

- Data sources for early risks: institutionalized readiness data model was built from integrated data. Each indicator uniquely predicted mental development and social risk factors, and they had to be cumulative. Data were gathered from:
  - Vital statistics
  - Public health
  - Child welfare
  - School districts
  - Homeless shelters
- Data sources for quality: early child care data were collected at the state level.
  - Philly had data on the all providers to include capacity and QRIS. We were interested in high quality providers, defined as providers that have the top two of the highest QRIS ratings and are accredited etc.
- Findings:
  - 20% of the kids had two or more risks
- We now needed to know where they lived so we did some more mapping to identify which neighborhoods have the highest concentrations of kids with the most risk factors. We had many conversations with the city councils and we landed on using names that are familiar to the community – neighborhood names instead of typical survey classifications of zip codes, counties etc.
- We then ranked high quality providers by the amount of regulated capacity – seats that could be served for three and four year children to identify where the most need was.
- **This identified Pre-K deserts. They had to meet both criteria (risk factors, and lack of supply) and 23 neighborhoods were flagged. These would then become the strategic places to which the state would invest the limited funds of 60 million dollars.**
- Implications:
  - Information was used by the community to call for new providers
  - Information was used to evaluate the proposal responses via RFPs
  - Extra points for higher need neighborhoods were allocated given the evidence in the data
  - Used the data to inform case management: we were able to go back and identify a lot of these kids and figure out their touch points with the city through data (e.g. social welfare etc.). From this we identified ways to reach them. For example, have case workers follow up with these kids and families and inform them of the resources available. That was when we started to shift from planning to practice.
  - Theme: We cannot build a system and assume they will come. Sometimes we need to go to them. This applies to Philly specifically as the Pre-K is a first come first serve system.
- Today:
  - We are working to integrate the city Pre-K enrollment data into the integrated data system. We are also working to expand two-generation indicators as we have gotten some additional data from the state.
  - We are trying to build in items like maternal depression and look at how that is a risk indicator for the baby and the mother. We are partnering with medical and nursing schools and working to create a life course perspective of these individuals.
- With systems like this, every time you ask a question you don't have to build a new data set. That creates efficiencies.
- **Question:** when you looked at cumulative risk, was that risk assigned a weight or no?
  - Risks were not weighted.
- **Question:** What quality thresholds did you use to include data elements and exclude them?

- In addition to the usual standards (e.g. out of range variables, nonsensical values) we used a lot of human conversations with our partners who best knew the data to know what was good and what wasn't. Based on their responses, we'd see if we can resolve some of the issues or not for bad data elements. This is why relationships are so important with the local players who best know the data.

- **Summary of Presentation #3: Kelly Maxwell**

- Notes from Kelly
  - This session showed fantastic cases of integrating different data. What are you'll working on next?
    - Looking to combine zip code data with the ECE density to look at school readiness indicators. This will help policy makers know where they should be investing resources. For example, we could identify where more training is needed or how households are making decisions on quality. Helps us also give policy makers a new way of looking at issues through maps to bring in more data driven decision making.
  - Trying to look at Part C data and how that links with health care and Income
    - This is a form of multisource linking. The multisector integrations could look at infant diagnoses related to development delays and disability indicators. Could also be looked at by race.
    - We could learn more about how much cost is needed for these children when they are born, to better understand what Medicaid pays and the gaps.
  - Simple data source linked across sectors could drive many of the conversations.

#### 4. Brief Summary of Discussion

**Facilitator comment:** What tips and advice do folks have to offer researches when working with linking ECE and non-ECE data?

- It is important to build good relationships, especially with data owners. Linking data is very hard and scary. In our work we have been most successful when trying to understand the other people's needs.
- Having some flexibility with what we want to do or get at. Not to be rigid in the approach with specific expectations.
- If we are working with multiple types of data, we need to also establish the gold standard of quality for one of the data sources. For example one data set will have to be comprehensive, in that if there is a certain discrepancy in multiple datasets you need to be able to go back to or fall back on a certain gold standard data set to know what is correct.
- We set up an advisory committee with the Early Learning Council. We had data sharing agreements with them and ensured we kept a loop for recurring feedback from the locals. Every community is going to be different, and we need to account for that.

**Question:** In your presentation, you mentioned a two-generation approach to families. Please elaborate.

Katherine's Response: from a data perspective it is easier to look at "baby" and "mom". These are the two generations. This is because there is a trickledown effect from mom to baby (e.g. maternal depression, smoking, alcohol use, working non-traditional hours). We started with a literature review of all indicators that can capture the things that could happen to the "mom". Soon, we would like to start looking at "dad" too.

**Question:** are there sources of data that we should consider as supportive factors?

Response: a different perspective to look at is strengths data (e.g. geospatial measures of density of parks, areas free from violence). We should focus on receiving more input on these items from the community level.

**Question:** Do you have access to child support data?

Response: Katherine – yes; Roland - No

**Question:** What are some challenges with using integrated data systems?

Challenges:

- Data governance issues – you need trusting partners
- Legal framework surrounding the data – you need lawyers to solve these data problems
- Technical issues - there are expensive ways and cheaper ways to fix these issues. There are also federated and non-federated models to consider.
  - Federated: This is where all partners keep their own data, and everyone can hold their data and that data are then integrated. It can be faster since there is less new infrastructure needed and the partners can retain more control
  - Non-federated: this what happened in Philly for our use case. It is one building with one data center, requires infrastructure but is more centrally managed.
- Building the big repository is what takes the most resources.
- Establishing data sharing agreements.
- Issues that come up with HIPPA and FERPA. Sharing allowances can cause some stopping points.

**Question:** have you'll worked with researchers to use integrated data systems to link administrative data with survey data?

Response: Not exactly. We haven't had such a research request. Our work model is quality assurance for the state. Even though we are building this repository, we've done some work on subsets of data for research that we've developed for state agencies.

**Question:** What is next for Katie and Roland? Will it be a new policy question or is it new data you are after?

Katie:

I'd like to continue line of inquiry for Pre-K and who gets into which programs. What are their outcomes? Would like to follow them longitudinally to understand how the landscape is changing.

Roland:

We are looking at data with the current opioid epidemic. Pharmacy prescribed opioids given in pregnancies, specifically the last trimester, and their linkages to the health of mother and baby with the lens of opioid exposure is what we are looking at. We would have to look at the birth records, to know if baby is addicted to opioids and understand what happens to these kids after early childhood? We are working with multiple state agencies for this. The data is very disparate and the diagnosis of Neonatal Abstinence Syndrome (NAS) is still not standardized – for example, for a child - is it really NAS or is it a heart condition.

## 5. Summary of Key Issues Raised

- The field needs to work against always using a deficit perspective for data. A lot of the data available are gathered from a deficit perspective - we collect a lot of data on deficits. We are combatting that through working with medical partners etc. we would like to bring in more community members and include more protective factors that they think are very important in the realm of available data.
- A few people in the room had worked with an actual integrated data system; most link the data themselves.
- There is always a fear of data being misinterpreted or misreported. Something could get taken and translated incorrectly. Therefore, it is always good to have trusted partners with institutional knowledge who can help guide the understanding of that data. This also speaks to the importance of good relationships so that the research can be done in the context of a partnership.
- There tends to be a concern with taking admin data and combining it with your own survey data. Drawing a sample – using that for research etc. is not as easy or permitted at times. For example, you can't just take a state's admin data and do what you want with it. States would have an interest in knowing how that data are being used on top of all the data sharing agreements and privacy concerns etc.
- Sometimes, states don't have sufficient resources to work with data the way research sometimes can. This is where the importance of robust research teams come to play.

- For creating integrated data systems, it is easier when you are dealing with 1 or 2 data sources. But when you deal with a lot of different data it gets very complex.
- Always have an advisory board. You need someone who knows the programming and the data contributors.