

Overcoming Common Barriers to Data Linkage

ASTHO, with support from CDC, launched the first cohort of the Linking Pregnancy Risk Assessment Monitoring System (PRAMS) and Clinical Outcomes Data Multi-Jurisdiction Learning Community in October 2021. The Learning Community provides technical support and capacity building expertise to states as they link their PRAMS data to their own set of clinical or social data, such as vital records, birth certificates, or home visiting.

On October 13, 2021, four state teams (**Alaska, New Mexico, Texas, and Washington state**) and members of an interagency workgroup made up of federal, academic, and industry partners attended a three-hour kickoff event. Individuals shared challenges they encountered during previous data linking activities and discussed lessons they learned in the process. Although state teams were the main contributors to these conversations, interagency workgroup members also offered advice based on their expertise. This brief examines the takeaways from these discussions as categorized into the following themes:

- Leadership buy-in and internal capacity.
- Partnerships and data sharing agreements.
- Matching and linking data.
- Dissemination and sustainability.

Challenges and Lessons Learned with Data Linkage

Leadership Buy-In and Internal Capacity

It is critical to garner leadership buy-in early in the data-linkage process to minimize delays in activities. However, this proves difficult when a health department finds itself too removed from its key decision makers. Having allies in other departments can be useful, but indirect communication with decision makers through advocates can result in miscommunication around project intent. Leadership may have organization-level concerns for the project team to consider, so forging internal partnerships is crucial. Additionally, leadership may have different priorities in their day-to-day responsibilities and by viewing the project through their lens, teams will be able to highlight the most important pieces of their work.

It is also important to be aware of the project team's ability to continue their work with the loss of a critical subject matter expert. With the possibility of job turnover, retirements, and staff reassignments, it is important to have a contingency plan. To support this project, Texas reallocated staff from a different part of the division to conduct data linkage activities. In response to losing critical staff at the start of the project period, another team committed to prioritizing that each team member was cross-trained and fully capable of filling gaps in capacity.

Partnerships and Data Sharing Agreements

When the desired dataset is identified, determining ownership and potential barriers to sharing must be considered. Developing the relationships and documents necessary to access the dataset can take months, as the politics surrounding data linkage can be more complex than the actual linkage itself. Preemptively researching potential barriers and motivators for data sharing, as well as engaging the support of legal experts, can accelerate the process. Because these partners can be external to the

health department, data use agreements may be required and the data steward may be difficult to identify. Once the appropriate stakeholders are engaged, the use of clear, specific language in data sharing agreements can help departments avoid miscommunication.

Through its work in the Learning Community, New Mexico experienced unique challenges securing home visiting data. The team pursued partnerships and entered into complex data sharing agreements with the Early Childhood, Education, and Care Department (ECECD) to obtain the appropriate data sets, which were housed in a university setting but needed to be routed through the separate agency. While the process to establish connections was lengthy, New Mexico leveraged existing relationships with the Office of the Secretary at ECECD to secure the necessary data.

Matching and Linking Data

Once the dataset has been secured, it is important that the project team allows enough time to complete the data linkage process. Each dataset contains unique variables. Determining which are critical for answering the research question must be done before matching. By carefully considering the research question, variable selection can ensure the question is answered. Linkage should occur in multiple iterative passes, with each series documented, so ample time to assess and clean data needs to be built into the process. To ensure equity throughout their linked dataset, Alaska conducted equity reviews after each step in the linkage process.

Familiarity with the population being examined can help the researcher avoid bias through data and sampling errors. The project team needs to have a strong understanding of what they're linking and what potential sampling errors could occur. New Mexico and Texas knew they needed to be aware of challenges surrounding name homogeneity. Name homogeneity can lead to duplication of patients and thus cause additional challenges related to case matching within the linked dataset.

Dissemination and Sustainability

Having a plan for sharing the results of data linkage is critical for the planning phase, as it often must be incorporated into any data sharing agreements. The outcomes and interested stakeholders will also drive research questions and variable selection. For the Washington state team, the ability to publicly share the linked PRAMS dataset was important. Early in their planning processes, they explored the possibility of dissemination through a newsletter. Texas set goals around creating a dissemination plan to ensure they were exhausting all opportunities to share their data.

With so many changes occurring in the public health workforce, knowledge documentation and transfer is vital to ensure that the work can continue even if a staff member leaves. However, building infrastructure and training is a time- and resource-intensive activity. Setting up these processes early in the data linkage process will save time and allow team members to be more focused on the technical aspects of data linkage.