SRAENE
Sexual Risk Avoidance Education
National Evaluation

June 2023

Tip Sheet

OPRE Report No. #2023-156

Preparing to Link Survey Data

Survey responses can be anonymous, or they can be linked to the person who gave them. Linking data from the same person across different surveys can help you learn more about your program than you would learn from analyzing anonymous data. An example of how to use anonymous data is a standard analysis of performance measures. Linking is unnecessary if you want to do that kind of analysis (for example, to report average program attendance rates over time), or if your goal is to assess aggregate changes over time (such as the average change between outcomes you are measuring at baseline and follow-up). Linked data, in contrast, can help you track which individuals do and do not complete surveys in order to follow up with individuals who do not complete surveys to improve response rates and understand whether certain groups of individuals respond differently to programming. Linked data can and should preserve an individual's privacy. The accompanying video introduced a few key steps to follow to link surveys and associate responses with the correct person while maintaining that person's privacy. These steps are summarized in the list below. Key terms and their definitions are in Table 1.

Decide whether you need to link survey data. You need to link survey data if you plan to compare the program outcomes of groups of youth whose baseline characteristics are different from each other; if you want to account for differences in the baseline characteristics in your outcome analysis; or if you want to track which youth do and do not respond to surveys in order to follow up with individuals who do not respond to improve response rates.

Use unique ID numbers to link the same data to the correct person. Create a unique ID number for each youth, one that is known only to a few people overseeing data collection. You can use the same ID number for the same person every time they complete a survey. To maximize privacy, assign survey IDs randomly so it is not easy to tell which ID number belongs to which person.

Prepare and handle materials to accurately link data and maintain privacy. If you are administering a baseline and follow-up survey, label each survey with the assigned survey ID so you can be confident the responses to each linked survey are from the same person. If you are using paper surveys, print IDs on the surveys and generate a removable cover page with the youth's name on it. If you are using online surveys, assign an ID through the online platform. After the respondent completes the survey, remove any data or materials with their name or other identifying information on them. Retain the survey ID to link the responses to name, baseline data, or a key used to track survey completion.

Table 1. Key terms

Term	Definition
Linked data	Two sets of responses from the same person, matched using a common identifier.
Baseline data	Data collected before or at the start of program participation and include measures of individual characteristics.
Follow-up data	Data collected after program participation and include measures of individual characteristics.
Survey ID	A unique identification number used to link data collected from the same person at different data collection points, such as baseline and follow-up.
ID randomization	Taking survey IDs out of numerical order before assigning them to youth in alphabetical order by youth's name.
Survey ID key	File that lists survey IDs and youth names, also known as a crosswalk.
De-identification	The process of removing any personally identifying information from surveys, such as names and contact information of youth who provided data.

For more on reasons to link survey data:

This brief discusses the benefits of linking data over time: https://www.casey.org/longitudinal-data/

This article discusses the benefits of linking data and different strategies to use when linking data with IDs: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0260569

For more on creating unique identification numbers:

This web page describes how to create unique identification numbers using SAS, a common statistical software: https://stats.oarc.ucla.edu/sas/faq/how-do-i-make-uniqueanonymous-id-variables-for-my-data/

This video describes how to create a random list of identification numbers in Microsoft Excel: https://www.youtube.com/watch?v=5yFvRbOACw4

This article discusses the use of self-generated identification numbers: https://www.researchgate.net/publication/5373291_The_Use_of_Self-Generated_Identification_Codes_in_Longitudinal_Research

For more on handling materials to maintain privacy:

This web page discusses data de-identification: https://www.umassmed.edu/it/security/research-and-clinical-data-access/data-de-identification/

These survey administration guidelines discuss how to collect completed surveys and keep them safe: https://www.prepeval.com/DataCollection/Survey_Admin_Guidelines.pdf

About this series

This video series, and the accompanying tip sheets on understanding and collecting high-quality data, were created as part of the Sexual Risk Avoidance Education National Evaluation (SRAENE). The series covers a range of data-related topics to help grantees understand the importance of high quality data and provide guidance on how they can collect them in their program. Although some of the resources are drawn from topic areas that are not related to SRAE, the content on data is still relevant.

FYSB does not recommend any particular survey platform or data system that may be referenced in tip sheets.

For more information or questions, contact the SRAENE team at SRAETA@mathematica-mpr.com.

Suggested citation: Stapleton, T., Eddins, K. (2023). *SRAENE – Preparing to Link Survey Data Tip Sheet* (OPRE Report No. #2023-156). Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.