

Slide 1 of 17

Title Slide: Considerations for Analysis of SASS and NTPS Data

Slide 2 of 17

This module describes analytic considerations that should be kept in mind when using NTPS data. Since the data in NTPS are inter-related and are stored in separate files, users may want to merge data files. For example, when a school is sampled, the principal and one or more teachers from that school are also sampled. Since these data are stored in separate data files, a school file, a principal file and a teacher file. Researchers can merge these files.

There are some cautions about how to merge the files and about nonresponse across data files that will also be described in this module. In addition, this module will explain the data file considerations including the importance of unit of analysis when merging files. This module also presents information on low response rates and nonresponse bias analysis for NTPS. Lastly, this module will describe considerations for analysis of NTPS and SASS data.

Throughout this module, underlined blue screen text indicates a link to additional resources.

Slide 3 of 17

Redesigned from the Schools and Staffing Survey (SASS) with a focus on flexibility, timeliness, and integration with other ED data, the NTPS system allows for school, principal, and teacher characteristics to be analyzed in relation to one another. NTPS maintains SASS's long-standing role as the primary source of data on teacher and principal labor markets, and on the state of K-12 school staffing. As a data analyst, you have decisions to make regarding how to use multiple data files in your analysis. That is, each data file can be analyzed separately, or two or more files can be merged.

The school control number "CNTLNUMS " is common to [all 3 data files per sector \(or in each sector\)](#) and serves as the primary variable for merging. Files are merged by way of CNTLNUMS when using one-to-one or one-to-many matching.

To analyze public and private sector data together, you will concatenate or stack data files. Merges are only used with data files within the same sector (public, private). Concatenation is used to combine data for the same type of respondent (e.g., all schools) across sectors. Concatenation never changes the unit of analysis, but merging does. For example, it is not possible to merge private teachers to the public school file by CNTLNUMS.

Slide 4 of 17

With one-to-one matching, each data record is matched by a unique identifying variable. For NTPS, the unique identifying variable is the school control number. You can match principals to schools in a one-to-one merge using the school control number. When

Considerations for Analysis of SASS and NTPS Data

there is no match on school control number, the separate data records from the second data file are added to the first data file in the order of the matching variable.

The newly merged data file would include information from both datasets, listed in school control number order. When there is a match on the control number, the variables from the second data file (as listed in the merge statement) that are not on the first data file, are added onto the matched record from the first data file. For NTPS, there are no questionnaire variables common across surveys since variables begin with a different letter on each of the school, principal, and teacher questionnaires. So, when merging the public school and the principal data files, for example, the additional principal variables that are not in the school data file are added at the end of the merged file.

Slide 5 of 17

Data from NTPS can also be combined by using a merging referred to as “one-to-many” matching. With one-to-many matching, the identifying variable is unique in one data file, but is repeated across many data records in another file. In other words, the variable that is unique in file 1 can be matched with multiple records or rows from file 2. For example, information from a given school data file can be merged with a data file containing information for all teachers in the sample. You can match teachers to schools or teachers to principals using a one-to-many merge. The illustration shows how teachers’ base salary information contained in the teacher data file, or file 1 can be compared to the number of days per school year for enrolled students from the school data file, or file 2.

Slide 6 of 17

When using Stata, a default merge variable is created during the merging of data files. The default name of this variable is “underscore merge” (*_merge*). The variable *_merge* identifies the various categories of data in a one-to-one match and can be used to specify a unit of analysis. The NTPS data files can be merged by using the school control number (CNTLNUMS). Both data files being merged must be sorted by the school control number prior to performing the merge. In the Stata code shown, words in italics are meant to be replaced by file or variable names that the user specifies. When merging any of the school, principal, or teacher files together for a given school, the school’s control number, CNTLNUMS, is used to merge data files.

Slide 7 of 17

In merging files within NTPS using SPSS, both data files being merged must be sorted by the variable listed in the “by” statement prior to performing the merge. In SPSS, value labels are attached automatically during the extraction process. For the syntax code shown here, words in italics are meant to be replaced by the file or variable names that the user specifies. When merging any of the NTPS files together, the school’s control number, CNTLNUMS, is used to merge the files.

Slide 8 of 17

The syntax on screen first specifies the two data files, with each one sorted by the school control number CNTLNUMS. The file specified as dataset1 is the unit of analysis.

So if the analysis is combining principal data with school data, and principals are the unit of analysis, dataset1 is the principal file and dataset2 is the school file. In SPSS, the command for merging is called match files. The table statement that follows, a forward slash in the match files command specifies the second data file, and the by statement with a forward slash provides the variable on which to perform the merge.

Slide 9 of 17

In merging data files within NTPS with SAS statistical software, please note that both data files being merged must be sorted by the matching variable listed in the “by” statement prior to performing the merge.

Here is an example of using SAS to perform a matching data file merge. Portions of this code have been ‘commented out’ to provide users with an explanation of what each line of code does when run. ‘Commented out’ code is contained within forward slashes and asterisks. It is also important to note that words shown in this example in italics are meant to be replaced by the file or variable names relevant to your analysis.

In SAS syntax, the convention is to identify the data file that is the unit of analysis with the statement “in equals a,” and that statement has to appear within parentheses. It can be used in a variety of ways in one-to-one and one-to-many matches.

For more information on different types of merges and using the “in equals a” convention, users should refer to the SAS statistical procedures manual.

Slide 10 of 17

As mentioned in “NTPS Sample Design, Weights and Variance,” each data file in NTPS has a different final weight. If you are only using one data file, such as the public teacher data file, then TFNLWGT is the weight you need. But if you have followed any of the above steps to merge together more than one component, your merged dataset will have more than one final weight. While a merged school and teacher data file might require the teacher weight for one analysis, another analysis might require the school weight.

In some cases it may not be obvious which weight is the proper one to use.

It is extremely important to identify the appropriate unit of analysis and use the correct weight, or else your analytic approach may be misguided and have misleading results. Usually, considering the target population is the best way to select which weight variable to use. In other words, considering which group the analysis is being generalized to represent will help you identify the unit of analysis for your research.

Considerations for Analysis of SASS and NTPS Data

If the research question is “Do rural schools have a higher proportion of experienced teachers than those in other community types?” the unit of analysis is the school, although the data on the experience level of teachers is drawn from the teacher file, the appropriate final weight is the school weight. But if the research question is “In rural schools, are experienced teachers more likely to be dissatisfied with school safety policies?” the unit of analysis is the teacher, since teachers are the ones voicing their opinions. Note that while rural schools might seem to be the target population, it is actually the teachers in those schools whose opinions are being represented. Consequently, in this second scenario, the teacher final weight should be applied.

Slide 11 of 17

Sample surveys rely on achieving an acceptable level of response. NTPS makes decisions on acceptable levels of response by using the unit response rate. The unit response rate indicates the percentage of sampled cases that met the definition of a complete interview. Nonresponse bias analysis is conducted for respondent characteristics falling below the reporting threshold level. NCES Statistical Standards require a nonresponse bias analysis when the level of response falls below the threshold rate of 85 percent. NCES will not publish estimates where the overall response rates fall below 50 percent until additional bias analyses are done and more information about potential bias is known.

Unit response rates fell below 50 percent for city school teachers in the 2015-16 NTPS. After conducting supplemental analyses, NCES has published data on teachers in city schools, but these data should be interpreted with caution. After nonresponse adjustments, the nonresponse bias for this category is greater than for other characteristics. In 2017-18, unit response rates fell below 50 percent for public schools, principals, and teachers in the state of Maryland. Additional details on unit response rates and nonresponse bias for public and private (2017-18 only) surveys are available for restricted-use data users within Chapter 4 of the respective User’s Manuals for the 2015–16 and 2017-18 National Teacher and Principal Survey Volume 1: Overview, which can only be accessed from the restricted-use DVD for each survey administration.

Slide 12 of 17

Biased data are responses that may not be representative of all of the target population. Nonresponse bias analysis is conducted for respondent characteristics falling below the reporting threshold level much like the statistical standards for the unit response rate. NCES requires a nonresponse bias analysis when item response rates fall below the threshold rate of 85 percent as survey respondents who skip some items may differ in key characteristics from survey respondents who supplied data to those items. When some respondents complete an item and others do not results may be biased. For a more detailed explanation of response rates and nonresponse bias analyses, the Restricted-Use NTPS User’s Manual for each survey is the best resource.

Slide 13 of 17

NTPS is designed to support explicit sampling strata estimates that were discussed in the module titled “NTPS Sample Design, Weights, and Variance,” and in Chapter 6 of the NTPS User’s Manuals, Volume 1. The NTPS User’s Manuals are only available via the restricted-use DVD. For any data analysis, it is imperative to compute the correct standard errors for all estimates computed. Analysts who do not do this may commit either a Type I or Type II error, finding data significant that are not statistically different, or failing to identify results as statistically significant.

The variance section of the “NTPS Sample Design, Weights, and Variance” module covers how to correctly compute standard errors in three of the most commonly used statistical software packages.

Potential survey respondents to NTPS are informed that all survey results will be published only in aggregate statistics. While there are no names recorded in the NTPS data files, some data elements could be used to potentially identify specific survey respondents. As a data analyst, you have an obligation not to identify specific respondents, including schools, teachers, and districts, in any research results or papers you publish.

You must formally agree to this under penalty of law when signing into PowerStats or as part of the restricted-use data licensing process.

Slide 14 of 17

Redesigned from the Schools and Staffing Survey (SASS) with a focus on flexibility, timeliness, and integration with other ED data, the NTPS system allows for school, principal, and teacher characteristics to be analyzed in relation to one another. The NTPS maintains many of the SASS teacher, school and principal items through core content and its series of rotating modules. “Core content” refers to the same items that are collected during each data collection.

Other items traditionally collected in SASS will be replaced with data from other NCES sources. Please check the NTPS website for more information about NTPS methodology, content, and data collection schedule.

NTPS modules rotate every other cycle. “Rotating modules” refers to administering modules or sets of related items not in every data collection but rotated into the questionnaires on a regular schedule.

Slide 15 of 17

These modules describe how to use the NTPS data files. NTPS content and procedures differ from SASS, which also underwent changes over the course of more than 20 years. Earlier versions of SASS may have sampling changes, content changes, or processing changes not covered here, including important changes to sampling and weighting procedures.

Considerations for Analysis of SASS and NTPS Data

Like SASS, each round of NTPS will have its own Data File User's Manual and Survey Documentation applicable to that data collection. Crosswalks do exist for comparing item content over time but each user is well advised to study questionnaire wording directly from one data collection to another.

A small proportion of the same schools may be included in SASS and NTPS samples from one data collection to the next but is not sufficient for any form of longitudinal or pseudo-longitudinal analysis. Samples of principals and teachers are independent in each data collection of NTPS. Background variables such as locale codes may change over time.

Slide 16 of 17

This module described analytic considerations that should be kept in mind when using NTPS data. Since the data in NTPS are inter-related, users may want to merge data files. Some types of data file merges are relatively simple to perform with NTPS data; others require some planning. However, there are some cautions about how to merge the files and about nonresponse across data files that were also described in this module.

This module explained the data file considerations including the importance of unit of analysis when merging files. The module also presented information on response rates, nonresponse bias, and other considerations for analyses of SASS and NTPS data.

Slide 17 of 17

Important resources that have been provided throughout the module are summarized in this slide for your reference.

This concludes the NTPS dataset training. You may now click the exit button to return to the landing page