NSECE Downloadable Presentation
II: Data Files and Documentation
NSECE Purpose: Address Key Policy Needs

- Comprehensive national data
- Match parental needs and preferences and availability of providers at the local level

- Profile all families with age-eligible children in a location
- Document all care in that location
  - Center-based providers to children not yet in kindergarten
  - Home-based providers to children under age 13 years
- Characterize early care and education (ECE) teachers and caregivers
- Describe place-based aspects of child care provision/use
National Survey of Early Care and Education

- An integrated set of surveys conducted in 2012 with
  - Households with children under 13
  - Home-based providers regularly providing early care and education to children under 13
  - Center-based programs providing early care and education to children not yet in kindergarten
  - Center-based classroom-assigned staff

- Sponsored by Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services
The figure on the slide depicts the five categories of data files that the NSECE will offer:

- Quick Tabulation Public Use: user-friendly files designed to facilitate quick tabulations on key topics. Quick Tabulation variables are also included as part of the main public-use files for each survey.
- Main Public Use: more complex than the quick tab files; include survey response data and extensive created variables.
- L1 Restricted Use Questionnaire Data: survey response data not available in the public-use files due to limited disclosure risk. These files will be available from ICPSR beginning Fall 2015.
- L2 Restricted Use Geographic Data
- L3 Archive of directly identifying information
Quick Tabulation File Description

- Variables can be used with a single sampling weight
- Skip patterns/eligibility fully specified
- Variables ready for tabulation
- Generally one variable per construct
- Sample defined for ready analysis
- Many fewer variables than main files

Variables in the Quick Tabulation data files are ready for tabulation:
- no known quality issues
- do not require extensive data processing
- could be used in stand-alone tabulation (has at least minimal sample size and variation).
The NSECE offers the six quick tabulation files list above. They contain the following types of data:

- Some questionnaire response data
- Some created variables
- Categorical community characteristics variables from American Community Survey data
- No identifying information
Main Public Use Files

• Center-Based Provider
• Classroom-assigned Workforce
• Home-Based Provider
• Household
• Household Calendar
• All files currently available

Five data files that include survey response data and extensive created variables. These are more complex than the quick tabulation files and most suitable for individuals with programming expertise and data/statistical knowledge.

Type of Data:
• All questionnaire response data represented subject to disclosure considerations
• Extensive created variables
• Categorical community characteristics variables from ACS data
• No identifying information

The household survey has two separate main public-use data file. The household calendar file includes child-level data file with care schedule in 15-minute increments for an entire week along with the parental work schedule. The household main public-use file contains household-level data on the 11,629 households.
There are four Level 1 Restricted use data files. These contain questionnaire response data and some created variables with disclosure risk. Restricted use files do not include variables that appear in the Main Public Use file other than necessary for linking the two.
(Level 2) Restricted Use Geographic Files

- These data will be made available as needed to approved research projects via remote access
- Available variables for each of the four public use files will be:
  - County (or county cluster) name
  - State name
  - County-level sampling weight
  - Cluster ID (but not actual location)
  - Cluster-level sampling weight
(Level 3) Restricted Use Identifying Data - Archive

- NSECE captured highly identifying information that will be available for processing by authorized staff under stringent conditions:
  - Household data for merging admin data
  - Provider name/location
  - GIS location of household employers and providers
Overview: Geographic Data

• Quick Tab/Public Use: selected community characteristics [in categorical form]
• L1 Restricted Use Questionnaire: community characteristics [raw form]
• L2 Restricted Use Geographic: county and/or state names as required
• L3 Archive Identifying Data: cluster locations, respondent names/addresses
### 4 Surveys and 23 Data Files

<table>
<thead>
<tr>
<th>Survey</th>
<th>Quick-Tabulation Files</th>
<th>Main Public-Use Files</th>
<th>Level 1 Restricted-Use Questionnaire Files</th>
<th>Level 2 Restricted-Use Geographic Files</th>
<th>Level 3 Restricted-Use Identifying Files (Archive)</th>
<th>Files per survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household</td>
<td>2 files: Household-level and Child-level</td>
<td>2 files: Household (incl. child data and calendar)</td>
<td>Household (incl. child data and calendar)</td>
<td>Household</td>
<td>Household</td>
<td>7</td>
</tr>
<tr>
<td>Center-Based providers</td>
<td>Center-Based providers</td>
<td>Center-Based providers</td>
<td>Center-Based providers</td>
<td>Center-Based providers</td>
<td>Center-Based providers</td>
<td>5</td>
</tr>
<tr>
<td>Home-Based providers</td>
<td>2 files: Listed home-based providers and unlisted home-based providers</td>
<td>Home-based providers file (listed and unlisted)</td>
<td>Home-based providers file (listed and unlisted)</td>
<td>Home-based providers file (listed and unlisted)</td>
<td>Home-based providers file (listed and unlisted)</td>
<td>6</td>
</tr>
<tr>
<td>Workforce</td>
<td>Workforce</td>
<td>Workforce</td>
<td>Workforce</td>
<td>Workforce</td>
<td>Workforce</td>
<td>5</td>
</tr>
</tbody>
</table>
Different Types of Documentation

- Data Collection and Sampling Summary
- Quick Tabulation Manuals (one per QT file)
- User’s Guides (one per survey)
  - Variable-level documentation (covering QT, Main Public Use, Restricted Use Questionnaire (L1) variables)
  - Information on use of sampling weights
- Research Reports, Fact Sheets, etc.
Each of the six quick tabulation files has an associated manual. The manual contains the following:

- An overview of the study, sample design and four NSECE surveys
- An overview of the NSECE data files
- Key concepts for using the quick tabulation file
- A schematic describing the key features of the codebook entries
- A table of main constructs in the data file with the associated variables
- Codebook entries for each variable
There are four user’s guides (one for each survey), which include variable-level documentation and descriptions of the data file contents for the quick tabulation, main public-use file, and the level 1 restricted-use file.
Variable name: WF_CHAR_HHINCOME
Label: Household income
Description: Income level of WF respondents, categorized into $7,500 intervals up until $45,001.
Variable type: Constructed variable
Counterpart in public use file: WF_E15_INCOMERANGE_CV
Counterpart in Home-Based Quick Tab data file: HB_CHAR_HHINCOME (Analogue constructed from similar source variables in Home-Based Provider Data)
Source questions asked of respondents:
E14. (WF_E14_INCOME) Approximately what was your total household income in 2011, before taxes or deductions? Please include income from wages and salaries earned by you or other adults in your household, as well as income from government assistance, gifts, or other income sources.
1. Don't know/Refused/No answer
E15. (WF_E15_INCOMERANGE) It can be difficult to remember or report these numbers and an approximate range is fine. What was your total household income in 2011 before taxes or deductions?
1. 0 to $7,500
2. $7,501 to $15,000
3. $15,001 to $22,500
4. $22,501 to $30,000
5. $30,001 to $45,000
6. $45,001 or more
7. Don't know/Refused/No answer
Variable construction:
This variable categorizes the WF respondent’s household income for 2011 into one of six income ranges. WF respondents were asked to approximate their total household income in WF_E14_INCOME. If they refused or did not know, they were then asked to select from given income levels in WF_E15_INCOMERANGE. These two questions were combined to create this single categorical income variable.
Notes: Original response data is available in the NSECE WF restricted-use data file.

Response Codes
<table>
<thead>
<tr>
<th>Income Range</th>
<th>Unweighted Frequency</th>
<th>Weighted Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to $7,500</td>
<td>1</td>
<td>7.01%</td>
</tr>
<tr>
<td>$7,501 to $15,000</td>
<td>2</td>
<td>9.95%</td>
</tr>
<tr>
<td>$15,001 to $22,500</td>
<td>3</td>
<td>10.81%</td>
</tr>
<tr>
<td>$22,501 to $30,000</td>
<td>4</td>
<td>10.64%</td>
</tr>
<tr>
<td>$30,001 to $45,000</td>
<td>5</td>
<td>14.09%</td>
</tr>
<tr>
<td>$45,001 or more</td>
<td>6</td>
<td>32.49%</td>
</tr>
<tr>
<td>DK/REF</td>
<td>1</td>
<td>15.01%</td>
</tr>
</tbody>
</table>

This slide contains an example of a codebook entry from a quick tabulation manual.
Key Features of Codebook Entry

- Variable name, label, and short description
- Name of variable in the main public-use data file (if different) and other relevant quick tab files
- Source questions (the name, question text, and response categories in the main public-use file)
- Variable construction
- Notes
- Response tabulations: unweighted counts and weighted percentages

Variable name: Name of variable in the quick tabulation (QT) data file. These names were created by using nested categories that distinguish the questionnaire and the content of the variable.

Source questions asked of respondents: Section includes questions from the survey instrument that are relevant to the quick tabulation variable. Direct report variables include the text of the question asked of respondents; constructed variables include text for all questions used to create the new variable.

Variable construction: a longer explanation of the variable content and how it was constructed. This includes a description of the universe along with any other assumptions or definitions used in the creation of the variable if applicable.

Notes: Additional points users may find helpful including NSECE resources and materials beyond the quick tabulation manual that will provide further information on the study, the data files, and select variables.
Variable-level Documentation

WF_CHAR_MARITAL

What is your current marital status?

1. Never married, not living with a partner (n=1312)
2. Married or living with a partner (n=2688)
3. Separated (n=139)
4. Divorced (n=466)
5. Widowed (n=114)
6. Don’t know/Refused/No answer (n=113)

NOTES:
In order to minimize the risk of disclosure, categories 3 and 5 have been collapsed into one. Total frequencies for each of the original response categories are also presented above (denoted by n=) and represent workers who served ages 0-5, only. The presented frequencies are for the total number of respondents – 4,832.

Original response data is available in the NSECE WF restricted-use data file.

This slide shows an example of an entry from the variable level-documentation for the quick tabulation, main public-use, and restricted-use data files.
Key Features of Variable-level Documentation Entry

- Variable name(s)
- Question text and response categories (if survey response)
- Description of variable content (if created)
- Notes
- Skip instructions (where appropriate)
- Interviewer instructions

Variable names: variables can have different names across the different files. All variable names will be included at the start of the entry. The name of the variable in the main public-use file is in color, the restricted-use version of the variable is in black and will be preceded by the letter R in parentheses. If a variable name has a check mark, it means that variable is also in the quick tabulation file.

Notes: the notes field includes additional information for data users like an explanation of how the variable was treated for disclosure, references to other variables within the data file, and explanations of how other/specify coding was done.
<table>
<thead>
<tr>
<th>Information Covered in Different Types of Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quick Tabulation Manuals</strong></td>
</tr>
<tr>
<td>Question wording</td>
</tr>
<tr>
<td>Response options</td>
</tr>
<tr>
<td>Item eligibility and skip patterns</td>
</tr>
<tr>
<td>Variable construction</td>
</tr>
<tr>
<td>Disclosure protection</td>
</tr>
<tr>
<td>Distribution of data values</td>
</tr>
<tr>
<td>Construct interpretation/tabulation</td>
</tr>
<tr>
<td>Key survey definitions and issues</td>
</tr>
<tr>
<td>Sampling information</td>
</tr>
<tr>
<td>Data collection information</td>
</tr>
</tbody>
</table>
Additional Resources
NSECE Resources

• Visit Research Connections
  http://www.researchconnections.org/childcare/studies/35519
  for NSECE updates
Questions about using the NSECE data?

Please write NSECE@NORC.ORG