Evaluation of a New Edit Methodology for the Common Core of Data (CCD) Nonfiscal Surveys

2013 Federal Committee on Statistical Methodology Research Conference

Elizabeth Goldberg, U.S. Census Bureau
Robert Stillwell, National Center for Education Statistics
Jeffrey Little, U.S. Census Bureau

Disclaimer: The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau or the National Center for Education Statistics.
Agenda

• CCD Overview
• Editing Process
  – Prior Method
  – New Method
• Results
• Conclusions and Next Steps
CCD OVERVIEW
CCD Overview

• Set of annual surveys sponsored by the National Center for Education Statistics (NCES)

• Data submissions by State Education Agencies (SEAs) from:
  – All public schools and local education agencies (LEAs)
  – 50 states, District of Columbia, territories and dependent agencies
CCD Overview

1. Universe surveys of various staff counts and student enrollment numbers at the state, agency, and school level

2. Universe counts of dropouts and high school completers at the state and school district level

3. Universe enumeration of finance data for each state (NPEFS) and school district (F-33)
CCD Overview

• Conducted since the 1950s
• In its current form since 1986
• Only annually updated comprehensive directory of U.S. public elementary and secondary schools and LEAs
• Data used to allocate funds from many federal grant programs
• Sampling frame for federal and non-federal surveys
## CCD Universe Characteristics

### CCD Universe Frequencies for 2010-11

<table>
<thead>
<tr>
<th>States</th>
<th>School Districts (LEAs)</th>
<th>Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>58 states and territories</td>
<td>18,478 LEAs</td>
<td>103,813 schools</td>
</tr>
</tbody>
</table>

### CCD Numeric Data Items for 2010-11

<table>
<thead>
<tr>
<th>States</th>
<th>School Districts (LEAs)</th>
<th>Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>240</td>
<td>244</td>
<td>253</td>
</tr>
</tbody>
</table>

CCD Universe Characteristics

• Detailed information published for every school and agency, not summary statistics

• Information generally publicly available from other sources

• Information of great interest to general public and others
CCD EDITING PROCESS
Editing Process

• Directory match
  – Comprehensive annual reconciliation of current year to prior year for all LEAs and schools

• Cross File Checks (Membership and Staff)
  – State total < LEA totals, State totals < School totals, LEA totals < School totals
  – Differences in State vs. LEA, State vs. School, or LEA vs. School by more than 5%

• Current Year vs. Prior Year Checks
CCD Numeric Data Edits

- Many items have associated sub-items

<table>
<thead>
<tr>
<th>School</th>
<th>Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Student Membership</td>
<td>Total Student Membership</td>
</tr>
<tr>
<td>Membership by Grade for grades PK-12</td>
<td>Membership by Grade for grades PK-12</td>
</tr>
<tr>
<td>Membership by Race</td>
<td>Membership by Race</td>
</tr>
<tr>
<td>Membership by Race/Grade/Gender</td>
<td>Membership by Race/Grade/Gender</td>
</tr>
</tbody>
</table>

- Each detailed data item has a CY/PY data check associated with it
  - For example, compare 2011 AM01F vs. 2010 AM01F
Prior Method

\[
\text{Ratio} = \frac{|\text{current year} - \text{prior year}|}{\text{prior year}}
\]

or

\[
\text{Difference} = |\text{current year} - \text{prior year}|
\]

where the data item fails the edit based on comparison of Ratio and/or Difference to a predetermined cutoff
Prior Edit Method

• Compares 2 years at a time
• Does not factor in historic variability of a data item
• Looks at each data item in isolation
• Subject to issues with prior year data as well as current year data
New Multi-Year Edit Method

• Designed to detect unusual variability in a school or agency vs. a broader historical context
• Based on classic statistical process control techniques
• Looks at data items in pairs vs. one data element at a time
• Reduces respondent and analyst burden
Multi-Year Edit Variables

• Agency
  – Membership and Pupil Teacher Ratio
  – Total classroom teachers and Pupil Teacher Ratio

• School
  – Membership and Pupil Teacher Ratio
  – Total classroom teachers and Pupil Teacher Ratio
  – Free and reduced lunch edit
  – Gender breakdown edit
  – Race breakdown edit
  – Elementary / high school membership edit
## Agency Variables

<table>
<thead>
<tr>
<th>Edit Type</th>
<th>Edited Items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agency Membership</strong></td>
<td>Total student membership  &lt;br&gt; Pupil Teacher Ratio = MEMBER / TOTTCH</td>
</tr>
<tr>
<td><strong>Total Teachers</strong></td>
<td>Total classroom teachers  &lt;br&gt; Pupil Teacher Ratio = MEMBER / TOTTCH</td>
</tr>
</tbody>
</table>
### School Variables

<table>
<thead>
<tr>
<th>Edit Type</th>
<th>Edited Items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School Membership</strong></td>
<td>Total student membership (MEMBER)</td>
</tr>
<tr>
<td></td>
<td>Pupil Teacher Ratio = MEMBER / FTE</td>
</tr>
<tr>
<td><strong>Total Teachers</strong></td>
<td>Total classroom teachers (FTE)</td>
</tr>
<tr>
<td></td>
<td>Pupil Teacher Ratio = MEMBER / FTE</td>
</tr>
<tr>
<td>**Free/Reduced Lunch</td>
<td>Total free and reduced lunch membership (TOTFRL)</td>
</tr>
<tr>
<td>Membership</td>
<td>(TOTFRL/MEMBER) * 100</td>
</tr>
</tbody>
</table>
New Method

For 2011-12, the multiyear method compares 2011 to 2010-2007

\[ \text{VARIABLE Y1} \]
\[ = \text{mean}(\text{all } |\text{VARIABLE } Y_i - \text{VARIABLE } Y_j|) \text{ for all years } i \text{ and } j \]
\[ \neq \text{to current year} \]


\[ \text{VARIABLE Y2} \]
\[ = \text{mean}(|\text{VARIABLE } Y_{\text{current year}} - \text{VARIABLE } Y_j|) \text{ for all years } j \]
\[ \neq \text{to current year} \]

When comparing current year to 4 other years, \text{VARIABLE Y2} is the mean of 4 differences (so 2011 to 2010, 2011 to 2009, 2011 to 2008, and 2011 to 2007)
New Method

A school or agency is flagged as an edit failure when both variables in each edit “pair” fail both criteria below:

\[ \text{VARIABLE Y2} \geq \text{MINDIF} \]

\[ \text{VARIABLE Y2} \geq \text{VARIABLE Y1} \times \text{LIMIT} \]

Where MINDIF and LIMIT are determined and set by Census staff for each variable of interest.
# Example Scenarios

<table>
<thead>
<tr>
<th>Total Student Membership</th>
<th>2007-08</th>
<th>2008-09</th>
<th>2009-10</th>
<th>2010-11</th>
<th>2011-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 1</td>
<td>81</td>
<td>54</td>
<td>43</td>
<td>65</td>
<td>85</td>
</tr>
<tr>
<td>School 2</td>
<td>5</td>
<td>8</td>
<td>15</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>School 3</td>
<td>1100</td>
<td>1300</td>
<td>1250</td>
<td>1000</td>
<td>1400</td>
</tr>
<tr>
<td>School 4</td>
<td>245</td>
<td>300</td>
<td>275</td>
<td>28</td>
<td>295</td>
</tr>
</tbody>
</table>
RESULTS
Method Comparison

• 2011-12 Comparison of Agency Failures by Method

<table>
<thead>
<tr>
<th></th>
<th>Frequency of Failures</th>
<th>Percentage Failed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old Method</td>
<td>New Method</td>
</tr>
<tr>
<td>Agency Membership</td>
<td>980</td>
<td>75</td>
</tr>
<tr>
<td>Teachers</td>
<td>477</td>
<td>102</td>
</tr>
</tbody>
</table>

• 2011-12 Comparison of School Failures by Method

<table>
<thead>
<tr>
<th></th>
<th>Frequency of Failures</th>
<th>Percentage Failed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old Method</td>
<td>New Method</td>
</tr>
<tr>
<td>School Membership</td>
<td>4,585</td>
<td>216</td>
</tr>
<tr>
<td>Teachers</td>
<td>14,571</td>
<td>836</td>
</tr>
<tr>
<td>Free/Reduced Lunch</td>
<td>3,977</td>
<td>3571</td>
</tr>
</tbody>
</table>
School Membership Edit

![Image of scatter plots showing variables Y1 and Y2 for school membership edit against NCES School ID Number.]
Agency Membership Edit

Variables Y1 and Y2 for Agency Membership Edit

- Member Y1
- Member Y2
- PUPITCH Y1
- PUPITCH Y2

NCES Agency ID Number
Suppression

- Prior to 2011-12 CCD survey cycle, data not corrected by respondent were published as reported.
- For 2011-12, data flagged using the new method not given an assignable cause is suppressed.
- Suppression decision based on results analyst research and respondent follow-up.
- Suppression of top level data item and also associated lower level detail items.
Suppression Results

- 2011-12 Agency Suppression Results

<table>
<thead>
<tr>
<th></th>
<th>Identified</th>
<th>Suppressed</th>
<th>% Suppressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membership</td>
<td>75</td>
<td>38</td>
<td>50.7%</td>
</tr>
<tr>
<td>Teachers</td>
<td>102</td>
<td>34</td>
<td>33.3%</td>
</tr>
</tbody>
</table>

- 2011-12 School Suppression Results

<table>
<thead>
<tr>
<th></th>
<th>Identified</th>
<th>Suppressed</th>
<th>% Suppressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membership</td>
<td>216</td>
<td>93</td>
<td>43.1%</td>
</tr>
<tr>
<td>Teachers</td>
<td>836</td>
<td>212</td>
<td>25.4%</td>
</tr>
<tr>
<td>Free/Reduced Lunch</td>
<td>3571</td>
<td>3527</td>
<td>98.8%</td>
</tr>
</tbody>
</table>
Conclusions

• CCD Staff and respondents consider the new method a success:
  – Reduces analyst and respondent burden
  – Streamlines editing process with fewer edit referrals
  – Requires more complex method to program
  – Requires larger historical database to be maintained concurrently
  – Requires at least 2 years of data or alternative edit methodology
Looking Ahead

• Adding new data groups to agency for 2012-13: administrative staff counts, support staff counts, ELL student membership, and special education membership

• Alternative methodologies being considered

• Addition of selective criteria for editing
  • For example, editing using school and agency size as a classification variable for edit limits

• Imputation vs. suppression
References


Contact Information

• General CCD Information
  patrick.keaton@ed.gov

• Edit Methodology
  elizabeth.j.goldberg@census.gov
  robert.stillwell@ed.gov