Additional Challenges and Solutions: The Post-Censal Denominator Small-Area Population Estimates and the Massachusetts Approach

Robert Knorr
Director of Environmental Epidemiology
Bureau of Environmental Health
Massachusetts Department of Public Health

2018 CSTE Annual Meeting, West Palm Beach, FL
June 10, 2018
Presentation Outline

1. Why do we need sub-county population estimates and what were our options?
2. UMass Donahue Institute method
   1. What is it?
   2. Error associated with the estimates
   3. Evaluation
      1. Childhood lead poisoning screening
      2. Life Expectancy
3. Feasibility for use by other geographies
In Massachusetts, Sub-County Estimates are Essential to Stakeholders

• Disparities at larger geographies can be smoothed over and opportunities to affect improvements in health missed

• Population growth and changing population characteristics
  • Community Needs Assessment
  • Health Impact Assessment
  • Emergency Preparedness
  • Identifying Health Disparities and Vulnerable Populations

• Interest in census-tract level life expectancy estimates only adds to the need for reliable population data
Post-Censal Population Estimation Options

• Simple linear extrapolation of U.S. Census data
  • Inherently biased as population change is not constant
  • Assumes a constant rate of growth or decline, producing larger error as the decade progresses

• Commercial products (e.g. Geolytics)
  • MA found large unexplained flaws with previously purchased estimates
  • Methods descriptions are vague at best

• The American Community Survey
  • 5-year running sample
  • Rate calculation using ACS small areas is not recommended due to large margins of error
  • Does not offer single-year age groups
Factors to Consider in Choosing an Option

• Accuracy of the estimate
• How uncertainty is accounted for
• Necessary detail
• Costs of production
• Ease of application and explanation
And the Method Massachusetts has gone with …

• The Modified Hamilton-Perry method
• Developed by the University of Massachusetts Donahue Institute
• specifically account for the three components of population change, births, deaths, and migration.
• the population at the beginning of the time period (commonly the most recent census), to which the number of births and in-migrants that occurred over the time period are added, and from which the number of deaths and out-migrants that occurred over the time period are subtracted.
UMDI – Strengths of the Methodology

• Commonly employed projection method (Hamilton-Perry)
• Requires Census data products only
• Annually adjusted for improved performance
• Capable of producing race, ethnicity, sex, and single-year age stratum
UMDI Methodology - Assigning Error

- Projected 2010 estimates from 10 and 20-year Cohort Change Ratios compared to 2010 Census
- Population estimate error was generally best explained by age and cohort size
- Population estimate errors were assigned to estimates based on base population cohort size and age
- As with estimates, errors were projected to 2020 and applied across intervening years. Error increases as you move forward in time.
Comparing Error
UMDI method vs. ACS method

Comparing the coefficient of variation (CV) for UMDI and ACS population estimates for 5-year age groups, stratified by community-based cohort size.

<table>
<thead>
<tr>
<th>Cohort Size</th>
<th>n</th>
<th>Under 5</th>
<th>5 to 9</th>
<th>10 to 14</th>
<th>15 to 19</th>
<th>20 to 24</th>
<th>25 to 29</th>
<th>30 to 34</th>
<th>35 to 39</th>
<th>40 to 44</th>
<th>45 to 49</th>
<th>50 to 54</th>
<th>55 to 59</th>
<th>60 to 64</th>
<th>65 to 69</th>
<th>70 to 74</th>
<th>75 to 79</th>
<th>80 to 84</th>
<th>85 and up</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 99</td>
<td>3556</td>
<td>17</td>
<td>20</td>
<td>16</td>
<td>15</td>
<td>19</td>
<td>23</td>
<td>17</td>
<td>18</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>13</td>
<td>15</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>100 to 199</td>
<td>1810</td>
<td>11</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>200 to 499</td>
<td>3468</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>13</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>500 to 999</td>
<td>2183</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>12</td>
<td>13</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>999 to 1999</td>
<td>1130</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2000 to 3999</td>
<td>367</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>4000 to 9999</td>
<td>92</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>NA</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>10000 to 199999</td>
<td>19</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>NA</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>20000 and up</td>
<td>12</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
MA Evaluation and Implementation Work

1. Review of impacts on childhood lead screening rates – UMDI vs linear extrapolation estimates

Lead Screening Rates (age 9-47 months) 
UMDI method vs. Linear method 

- A-town 
- B-land 
- C-ville 
- D-boro 

Calculated Using: 
- Linear Method 
- UMASS Method
Lead Screening Rates >100%
UMDI method vs. Linear method

<table>
<thead>
<tr>
<th>Town population</th>
<th>Number of Towns</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000-74,999</td>
<td>35</td>
</tr>
<tr>
<td>&lt;10,000</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMDI</td>
<td>3.1%</td>
</tr>
<tr>
<td>Linear</td>
<td>10.3%</td>
</tr>
</tbody>
</table>
Point Estimates of Life Expectancy Using ACS Population Data as the Denominator
Point Estimates of Life Expectancy Using UMDI Population Data as the Denominator
Standard Error Using ACS Population Data as the Denominator

ACS Denominator

Standard Error
- 0-1
- 1.1-2
- 2.1-3
- 3.1+
- Incomplete Data
Standard Error Using UMDI Population Data as the Denominator
Denominator Point Estimates of Life Expectancy

The graph shows the frequency distribution of life expectancy across different age groups. The x-axis represents the life expectancy ranges, and the y-axis shows the frequency. Two categories are compared:

- ACS
- UMDI
Standard Error of Denominator

![Histogram showing frequency of standard error values for ACS and UMDI datasets.](chart.png)
Census Tract Estimates
Census Tract Estimates

Standard Error

ACS
UMDI
Most easily applied where:

• Geographies remain fairly stable from Census to Census and/or
• Have excellent correspondence files across time 1990-2010.

Difficulty presented when:

• Boundaries have changed or annexations have taken place from Census to Census
  • Would require efforts to document changes
  • Helpful if records available (e.g. correspondence files)
• Sub-county population sizes are small
  • A higher geographic level may be needed (i.e. tract aggregation)
  • More uncertainty in the estimates may occur
QUESTIONS?

Contact Information

Robert Knorr
Director
Environmental Epidemiology Program
Bureau of Environmental Health
MA Department of Public Health
250 Washington Street, 7th floor
Boston, MA 02108

(617) 624-5757
robert.knorr@state.ma.us

Susan Strate
Senior Manager Population Estimates Program
UMass Donahue Institute
Economic & Public Policy Research
100 Venture Way, Suite 9
Hadley, MA 01035

(413) 577-0753
sstrate@donahue.umassp.edu
www.donahue.umassp.edu
Supplemental Slides
UMDI Methodology – Aging Forward

• Standard Hamilton-Perry Methodology

• Cohort Change Ratio – “Ages” age-sex cohorts forward from one census to the next

• Ratio is applied to base population to create future estimate.

<table>
<thead>
<tr>
<th>Town</th>
<th>AGE</th>
<th>SEX</th>
<th>1990 Census</th>
<th>2000 Census</th>
<th>10-year CCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABINGTON</td>
<td>0-4</td>
<td>F</td>
<td>482</td>
<td>479</td>
<td>0.15</td>
</tr>
<tr>
<td>ABINGTON</td>
<td>5-9</td>
<td>F</td>
<td>495</td>
<td>500</td>
<td>0.15</td>
</tr>
<tr>
<td>ABINGTON</td>
<td>10-14</td>
<td>F</td>
<td>451</td>
<td>530</td>
<td>1.10</td>
</tr>
<tr>
<td>ABINGTON</td>
<td>15-19</td>
<td>F</td>
<td>450</td>
<td>462</td>
<td>0.93</td>
</tr>
<tr>
<td>ABINGTON</td>
<td>20-24</td>
<td>F</td>
<td>558</td>
<td><strong>364</strong></td>
<td><strong>0.81</strong></td>
</tr>
<tr>
<td>ABINGTON</td>
<td>25-29</td>
<td>F</td>
<td>608</td>
<td>458</td>
<td>1.02</td>
</tr>
<tr>
<td>ABINGTON</td>
<td>30-34</td>
<td>F</td>
<td>633</td>
<td>577</td>
<td>1.03</td>
</tr>
<tr>
<td>ABINGTON</td>
<td>35-39</td>
<td>F</td>
<td>569</td>
<td>689</td>
<td>1.13</td>
</tr>
<tr>
<td>ABINGTON</td>
<td>40-44</td>
<td>F</td>
<td>516</td>
<td>631</td>
<td>1.00</td>
</tr>
</tbody>
</table>
UMDI Methodology – Reflects local Shift in CWRs

Birth Rate by Maternal Age Group, U.S. and Massachusetts

<table>
<thead>
<tr>
<th>Maternal Age Group</th>
<th>United States</th>
<th>Massachusetts</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 to 50 years</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>20 to 34 years</td>
<td>95</td>
<td>75</td>
</tr>
<tr>
<td>15 to 19 years</td>
<td>21</td>
<td>10</td>
</tr>
</tbody>
</table>

Birth rate per 1,000 women

U.S. Census Bureau American Community Survey 5-year dataset 2010-2014
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0-99</td>
<td>3874</td>
<td>35%</td>
<td>59%</td>
<td>38%</td>
<td>31%</td>
<td>32%</td>
<td>51%</td>
<td>48%</td>
<td>45%</td>
<td>27%</td>
<td>22%</td>
<td>25%</td>
<td>22%</td>
<td>20%</td>
<td>22%</td>
<td>24%</td>
<td>24%</td>
</tr>
<tr>
<td>100-199</td>
<td>1915</td>
<td>25%</td>
<td>19%</td>
<td>14%</td>
<td>15%</td>
<td>14%</td>
<td>22%</td>
<td>27%</td>
<td>20%</td>
<td>16%</td>
<td>14%</td>
<td>11%</td>
<td>13%</td>
<td>11%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td>200-499</td>
<td>3228</td>
<td>19%</td>
<td>14%</td>
<td>13%</td>
<td>10%</td>
<td>15%</td>
<td>19%</td>
<td>22%</td>
<td>17%</td>
<td>14%</td>
<td>11%</td>
<td>11%</td>
<td>9%</td>
<td>11%</td>
<td>11%</td>
<td>9%</td>
<td>9%</td>
</tr>
<tr>
<td>500-999</td>
<td>2115</td>
<td>17%</td>
<td>13%</td>
<td>10%</td>
<td>9%</td>
<td>14%</td>
<td>13%</td>
<td>21%</td>
<td>17%</td>
<td>9%</td>
<td>6%</td>
<td>6%</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>1,000-1,999</td>
<td>1067</td>
<td>14%</td>
<td>13%</td>
<td>9%</td>
<td>9%</td>
<td>14%</td>
<td>10%</td>
<td>16%</td>
<td>12%</td>
<td>8%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>7%</td>
<td>8%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>2,000-3,999</td>
<td>328</td>
<td>7%</td>
<td>15%</td>
<td>7%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>13%</td>
<td>9%</td>
<td>7%</td>
<td>6%</td>
<td>4%</td>
<td>3%</td>
<td>2%</td>
<td>2%</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>4,000-9,999</td>
<td>85</td>
<td>3%</td>
<td>15%</td>
<td>5%</td>
<td>5%</td>
<td>8%</td>
<td>8%</td>
<td>10%</td>
<td>9%</td>
<td>9%</td>
<td>4%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>10,000-19,999</td>
<td>14</td>
<td>5%</td>
<td>33%</td>
<td>7%</td>
<td>4%</td>
<td>3%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>20,000+</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12636</td>
<td>22%</td>
<td>24%</td>
<td>17%</td>
<td>15%</td>
<td>20%</td>
<td>28%</td>
<td>27%</td>
<td>21%</td>
<td>14%</td>
<td>11%</td>
<td>12%</td>
<td>12%</td>
<td>13%</td>
<td>14%</td>
<td>15%</td>
<td>16%</td>
<td>18%</td>
</tr>
</tbody>
</table>
UMDI Projected Population Estimate Error, by Geography

Absolute Percent Error by Town for 10-Year Cohort Change Ratio Estimates

Average Absolute Percent Error for all MCDs: 8.6%
Challenges with Implementing Estimates

1. Race categories
   - Excludes “Other” category
   - “Other” re-distributed proportionally to other race categories based on census expertise and local information.
     - “Other” captures different group of people at each census due to changes in wording of question through decades.
     - “Other” is also inconsistently applied across datasets, so these numbers are difficult to interpret regardless of denominator definition.
Challenges with Implementing Estimates

2. Totals

• Estimates were modeled at the smallest cell (though “all race” and “all ethnicities” were modeled as well)
• Race re-distribution and rounding create inconsistencies depending on aggregation parameters
  • Geographies: County aggregation inconsistent between census tract and Municipality
    • Creates some problems for “Special Geographies” used by the state
• Age: single year vs. 5-year
• Individual races/ethnicity vs. all races/ethnicity categories
Challenges with Implementing Estimates

3. Moving Estimates

- With each vintage of County Population census data, estimates are updated for all years 2010 – most current vintage.
- Creates changes in denominators over time, and altered rates as a result.
  - Most datasets lag enough that this shouldn’t impact resulting rates
  - Not the case for Lead in MA