Mortality Trends and Variation in China

Yong Cai caiyong@unc.edu
University of North Carolina, Chapel Hill

May 15, 2019
Harnack Haus, Berlin

Satellite meeting
Mortality data and methodological approaches in estimating mortality in developing countries

Sponsored by Max Planck Institute for Demographic Research
Overview

- Background
- Data sources & challenges
- Mortality trends and variation
Socioeconomic background

- Economic growth
  - Changes in nutrition and consumption
- Social transformation
  - Individual awareness and life style
- Establishment of “new” social welfare system
  - Healthcare and delivery
  - Public health investment
Data sources and challenges

• Main sources
  • Population census and surveys
  • Health surveillance systems
  • Death registration
  • Household registration (Hukou)
  • Social pension system

• Main challenges
  • Data quality
    • Underreporting
    • Migration
    • Government evaluation
  • Data availability
    • Segmentation
    • Undocumented adjustment
Chinese data: good and bad

FIGURE 2  Intercensal survival ratio by sex and birth cohort, China 2000–2010
Mortality trends: $e_0^t$

![Graph showing mortality trends with different data sources and time periods.](image-url)
Mortality trends: $q_{x}^{t}$
Mortality trends: $lgt(l_x^t)$ compared to 2000
Mortality trends: $lgt(l_x^t)$ compared to 2000
Mortality trends: gender gap $e^f_0 - e^m_0$
Morality trends: Shanghai $e_0$
Morality trends: Shanghai
Mortality trends: infant mortality rate

![Graph showing mortality trends over time, with lines for Surveillance Adjusted, Population Survey, and Surveillance Raw.](image-url)
Mortality trends: basic observations

- Rising life expectancy fits well with the general pattern of mortality decline as it was driven by multiple factors happening in China: socioeconomic development, investment in public health, change of life style...
- Mortality underreporting is a constant problem, and will remain to be.
- While overall mortality decline can be collaborated with other data sources, ascertain of mortality level requires careful cross-examination with other population data.
Mortality trends by rural/urban (unadjusted $e_0^t$)
Mortality trends: rural/urban gap (unadjusted $e_{0t}$) comparing to $e_{0t}$ for villages

![Graph showing mortality trends in towns and cities from 1998 to 2018. The graph displays two lines: one for towns and one for cities. The trend for towns shows a decrease over time, while the trend for cities shows fluctuations.](#)
Regional trends and variation, $e_0^t$
Mortality variation: $e_0(2000, \text{ unadjusted})$
## Mortality variation: county level $e_0$ (2000, raw)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>5</th>
<th>Med</th>
<th>95</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_0^m$</td>
<td>69.9</td>
<td>3.9</td>
<td>46.4</td>
<td>62.9</td>
<td>70.5</td>
<td>75.3</td>
<td>86.9</td>
</tr>
<tr>
<td>$e_0^f$</td>
<td>73.8</td>
<td>4.7</td>
<td>45.3</td>
<td>65.4</td>
<td>74.5</td>
<td>80.0</td>
<td>106.7</td>
</tr>
<tr>
<td>$1q_0^m$</td>
<td>25.7</td>
<td>23.4</td>
<td>0.0</td>
<td>4.6</td>
<td>22.9</td>
<td>91.1</td>
<td>376.7</td>
</tr>
<tr>
<td>$1q_0^f$</td>
<td>32.8</td>
<td>32.0</td>
<td>0.0</td>
<td>4.9</td>
<td>18.9</td>
<td>68.6</td>
<td>307.3</td>
</tr>
</tbody>
</table>
Brass Logit Model Smoothing

- Brass (1971) relational model: linear relationship between $l_x$'s of an observed population and of a standard population at logit scale

$$\text{logit}(l_x) = \alpha + \beta \cdot \text{logit}(l_x^s)$$

- Consistency of mortality across different age groups

- Choice of standard life table: “must be some kind of average Brass (1971)” We use provincial average.
Empirical Bayes Estimates


- Information from neighboring units serves as *priori* for the local estimate.

\[
\hat{\theta}_i = m_n + \frac{S^2 - \frac{m_n}{\bar{n}}}{S^2 - \frac{m_n}{\bar{n}} + \frac{m_n}{n_i}} (m_i - m_n)
\]
Empirical Bayes Estimates
## Mortality variation: county level $e_0$ (2000 smoothed)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>5</th>
<th>Med</th>
<th>95</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_0^m$</td>
<td>69.8</td>
<td>3.7</td>
<td>46.6</td>
<td>62.9</td>
<td>70.5</td>
<td>74.5</td>
<td>79.3</td>
</tr>
<tr>
<td>$e_0^f$</td>
<td>73.4</td>
<td>4.2</td>
<td>45.5</td>
<td>65.3</td>
<td>74.3</td>
<td>78.6</td>
<td>84.0</td>
</tr>
<tr>
<td>$q_0^m$</td>
<td>26.0</td>
<td>22.1</td>
<td>0.9</td>
<td>6.1</td>
<td>23.8</td>
<td>93.0</td>
<td>355.2</td>
</tr>
<tr>
<td>$q_0^f$</td>
<td>33.6</td>
<td>31.3</td>
<td>1.2</td>
<td>6.6</td>
<td>19.2</td>
<td>66.9</td>
<td>267.7</td>
</tr>
</tbody>
</table>
Mortality variation: county level $e_0^t$ (2000)
Mortality variation: county level clusters (2000)

- low, <.10
- low, <.05
- low, <.01
- high, <.01
- high, <.05
- high, <.10
- not sig.
### Mortality variation: $e_0$ 2000 regression

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Male Model</th>
<th></th>
<th></th>
<th>Female Model</th>
<th></th>
<th></th>
<th>Difference (Male - Female)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coef</td>
<td>std</td>
<td>beta</td>
<td>coef</td>
<td>std</td>
<td>beta</td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>61.526</td>
<td>1.9371</td>
<td>***</td>
<td>64.099</td>
<td>2.3035</td>
<td>***</td>
<td>-2.573</td>
</tr>
<tr>
<td>GDP</td>
<td>0.6292</td>
<td>0.1119</td>
<td>0.1205 ***</td>
<td>1.0133</td>
<td>0.1331</td>
<td>0.1676 ***</td>
<td>-0.384 *</td>
</tr>
<tr>
<td>Urbanization</td>
<td>0.0046</td>
<td>0.0042</td>
<td>0.0296</td>
<td>-0.0014</td>
<td>0.0050</td>
<td>-0.0080</td>
<td>0.006</td>
</tr>
<tr>
<td>Employment</td>
<td>0.0179</td>
<td>0.0089</td>
<td>0.0435 *</td>
<td>0.0312</td>
<td>0.0106</td>
<td>0.0654 **</td>
<td>-0.013</td>
</tr>
<tr>
<td>Health Care</td>
<td>-0.6532</td>
<td>0.2782</td>
<td>-0.0518 *</td>
<td>-0.4184</td>
<td>0.3308</td>
<td>-0.0286</td>
<td>-0.235</td>
</tr>
<tr>
<td>Water</td>
<td>0.0044</td>
<td>0.0026</td>
<td>0.0337</td>
<td>0.0040</td>
<td>0.0031</td>
<td>0.0265</td>
<td>0.000</td>
</tr>
<tr>
<td>Lavatory</td>
<td>0.0058</td>
<td>0.0022</td>
<td>0.0448 **</td>
<td>0.0069</td>
<td>0.0026</td>
<td>0.0458 **</td>
<td>-0.001</td>
</tr>
<tr>
<td>Bath Facility</td>
<td>0.0121</td>
<td>0.0042</td>
<td>0.0659 **</td>
<td>0.0162</td>
<td>0.0050</td>
<td>0.0761 **</td>
<td>-0.004</td>
</tr>
<tr>
<td>Education</td>
<td>0.9515</td>
<td>0.0982</td>
<td>0.3524 ***</td>
<td>1.0719</td>
<td>0.1167</td>
<td>0.3428 ***</td>
<td>-0.120</td>
</tr>
<tr>
<td>Discrimination</td>
<td>0.0465</td>
<td>0.1629</td>
<td>0.0060</td>
<td>0.0353</td>
<td>0.1937</td>
<td>0.0039</td>
<td>0.011</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>-0.0253</td>
<td>0.0029</td>
<td>-0.2152 ***</td>
<td>-0.0151</td>
<td>0.0035</td>
<td>-0.1107 ***</td>
<td>-0.010 *</td>
</tr>
<tr>
<td>Marriage</td>
<td>-0.0417</td>
<td>0.0133</td>
<td>-0.0677 **</td>
<td>-0.0705</td>
<td>0.0158</td>
<td>-0.0987 ***</td>
<td>0.029</td>
</tr>
<tr>
<td>Family Structure</td>
<td>0.0117</td>
<td>0.0095</td>
<td>0.0229</td>
<td>0.0106</td>
<td>0.0113</td>
<td>0.0180</td>
<td>0.001</td>
</tr>
<tr>
<td>Latitude</td>
<td>0.0324</td>
<td>0.0177</td>
<td>0.0580</td>
<td>-0.0023</td>
<td>0.0210</td>
<td>-0.0035</td>
<td>0.035</td>
</tr>
<tr>
<td>Longitude</td>
<td>0.0064</td>
<td>0.0153</td>
<td>0.0167</td>
<td>0.0168</td>
<td>0.0182</td>
<td>0.0379</td>
<td>-0.010</td>
</tr>
<tr>
<td>Elevation</td>
<td>-0.0003</td>
<td>0.0001</td>
<td>-0.0780</td>
<td>-0.0002</td>
<td>0.0002</td>
<td>-0.0415</td>
<td>0.000</td>
</tr>
<tr>
<td>Terrain</td>
<td>-0.0009</td>
<td>0.0003</td>
<td>-0.0544 **</td>
<td>-0.0011</td>
<td>0.0004</td>
<td>-0.0548 **</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*** denotes p<.001; ** denotes p< .01; * denotes p< .05
Mortality variation: county level $e_0^f$ (2000)
Mortality variation: observations

- Huge regional variation: China is a world of its own!
- Regional variation is relatively stable
- Regional differences are largely driven by developmental factors
Next Steps

- A model/system simultaneously estimating fertility, mortality, population age structure using all available survey/census data to consistently estimate demographic parameters.
- Subnational evaluation
- HMD standards/protocols
Thank you!