Using multiple longitudinal datasets to inform a micro-simulation model of the early life-course

COMPASS Colloquium
August 2013

Barry Milne
COMPASS Research Centre
University of Auckland
New Zealand

www.compass.auckland.ac.nz
To build a realistic simulation model of the early life course (0-13) for policy purposes, we are:

1. Combining information across four longitudinal studies into a unified (more robust) data set.
   - To analyse to get rules for transitioning people from one state to the next
2. Weighting the combined dataset by ethnicity
   - To analyse a sample that has a representative ethnic balance
3. Preparing a synthetic birth cohort from 2006 Census
   - So that our simulation represents NZ today

I will talk about 1 and 2 now, and 3 later
Four Studies

- Christchurch Health & Development Study (CHDS)
  - 1265 children born in ChCh 1977. Followed since

- Dunedin Multidisciplinary Health & Development Study (DMHDS)
  - 1037 children born in Dunedin 1972/3. Followed since

- Pacific Islands Families Study (PIFS)
  - 1398 children born at Middlemore, 2000, with at least one parent of Pacific Islands ethnicity. Followed since

- Te Hoe Nuku Roa Study (THNR)
  - Longitudinal study of Māori households (beginning 1995)
    - Auckland, Wellington, Manawatu, Gisborne, Northland, Southland, Nelson
  - 568 children (0-12) assessed at least twice in four waves
1. Data integration

- Original model based on CHDS
  - Use data from DMHDS & PIFS on those constructs used in CHDS-based model; ignore other constructs

- Issues around
  - Different times
  - Same constructs measured differently
  - Missing data
  - Ensuring combined is representative of NZ

- Solutions
1. Data integration - Different times

- Associations between X & Y assessed using longitudinal GEE analyses
  - Utilises data from all the ages available from the three studies (THNR not used)

<table>
<thead>
<tr>
<th>Age</th>
<th>$Y_{CHDS}$</th>
<th>$Y_{DMHDS}$</th>
<th>$Y_{PIFS}$</th>
<th>$X_{CHDS}$</th>
<th>$X_{DMHDS}$</th>
<th>$X_{PIFS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>1</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Data integration
- Construct measurement

- 30/36 constructs measured identically between CHDS & DMHDS; 24/26 between CHDS & PIFS

- 4 DMHDS constructs measured otherwise identically but cover different timeframe (e.g., past 2 years in DMHDS; past 1 year in CHDS)
  - Random imputation to subset to one year (r~0.65)

- 2 DMHDS & 2 PIFS constructs measured using different scales
  - Conduct disorder, Harsh punishment
  - Align to same metric using min/max points
1. Data integration

- Missing data

- ‘Holes’ in data in each study filled in
  - 60% vars have <10% missing; 14% vars have 20-30%
  - Model-based multiple imputation using within-study models, imputing vars with least error first (following SGP)

- Constructs in DMHDS/PIFS with missing ages
  - 15% constructs
  - Model-based multiple imputation using within-study models (or another study if time trends important)

- Missing constructs in DMHDS/PIFS
  - 4/40 constructs in DMHDS, 14/40 constructs in PIFS
  - Model-based multiple imputation using CHDS study models
2. Weighting by ethnicity

Combined CHDS, DMHDS & PIFS not representative of NZ’s ethnic distribution currently

Weight by ethnicity:

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>DMHDS</th>
<th>CHDS</th>
<th>PIFS</th>
<th>Combined</th>
<th>Census</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European</td>
<td>90.1%</td>
<td>86.1%</td>
<td>2.8%</td>
<td>55.9%</td>
<td>58.2%</td>
<td>58.2/55.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>= 1.04</td>
</tr>
<tr>
<td>Maori</td>
<td>8.4%</td>
<td>10.7%</td>
<td>6.2%</td>
<td>8.4%</td>
<td>24.2%</td>
<td>24.2/8.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>= 2.88</td>
</tr>
<tr>
<td>Pacific</td>
<td>1.5%</td>
<td>3.2%</td>
<td>91.0%</td>
<td>35.7%</td>
<td>9.2%</td>
<td>9.2/35.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>= 0.26</td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.5%</td>
<td></td>
</tr>
</tbody>
</table>
2. Weighting by ethnicity
- Cultural affiliation

- Likely that CHDS & DMHDS Māori not representative of Māori nationally

- Solution?
  - Use cultural affiliation as ‘representativeness’ indicator
  - Compare cultural affiliation between CHDS & DMHDS Māori and THNR Māori, and weight CHDS & DMHDS distributions to look like THNR
  - CHDS, DMHDS & THNR each have items on
    - Marae visit, Tangi attendance, involvement in Māori groups, language understanding, Māori language TV/radio
    - NB, No Māori cultural affiliation items in PIFS
  - Draw principal component from these items and compare CHDS & DMHDS against THNR quintiles
## 2. Weighting by ethnicity
- Cultural affiliation distributions

<table>
<thead>
<tr>
<th>Quintiles</th>
<th>THNR (%)</th>
<th>CHDS (%)</th>
<th>DMHDS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - low</td>
<td>20.0</td>
<td>53.7</td>
<td>66.7</td>
</tr>
<tr>
<td>2</td>
<td>20.0</td>
<td>22.3</td>
<td>12.3</td>
</tr>
<tr>
<td>3</td>
<td>20.0</td>
<td>8.3</td>
<td>7.0</td>
</tr>
<tr>
<td>4</td>
<td>20.0</td>
<td>12.4</td>
<td>5.3</td>
</tr>
<tr>
<td>5 - high</td>
<td>20.0</td>
<td>3.3</td>
<td>8.8</td>
</tr>
</tbody>
</table>
2. Weighting by ethnicity
- Cultural affiliation weights

<table>
<thead>
<tr>
<th>Quintiles</th>
<th>THNR (%)</th>
<th>CHDS (%)</th>
<th>Weight</th>
<th>DMHDS (%)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - low</td>
<td>20.0</td>
<td>53.7</td>
<td>20/53.7 =0.37</td>
<td>66.7</td>
<td>20/66.7 =0.30</td>
</tr>
<tr>
<td>2</td>
<td>20.0</td>
<td>22.3</td>
<td>20/22.3 =0.90</td>
<td>12.3</td>
<td>20/12.3 =1.63</td>
</tr>
<tr>
<td>3</td>
<td>20.0</td>
<td>8.3</td>
<td>20/8.3 =2.41</td>
<td>7.0</td>
<td>20/7.0 =2.86</td>
</tr>
<tr>
<td>4</td>
<td>20.0</td>
<td>12.4</td>
<td>20/12.4 =1.61</td>
<td>5.3</td>
<td>20/5.3 =3.77</td>
</tr>
<tr>
<td>5 - high</td>
<td>20.0</td>
<td>3.3</td>
<td>20/3.3 =6.06</td>
<td>8.8</td>
<td>20/8.8 =2.27</td>
</tr>
</tbody>
</table>
2. Weighting by ethnicity  
- Cultural affiliation assumptions

- A Māori sample representative on cultural affiliation will be a representative Māori sample
  - Perhaps. Geographic differences??

- THNR is a representative Māori sample
  - Probably for regions sampled.
  - Te Kupenga (Māori Social Survey) another option?

- Cultural affiliation is measured well by the items we used
  - Probably. Cultural affiliation items load on one factor.

- Cultural affiliation is stable across the life-course
  - Possibly. Items measured longitudinally (THNR) correlated moderately - strongly
Summary and Next Steps

- Integration of data from datasets feasible
  - Bit of work, similarity of constructs has helped
- Method to make analysis sample ethnically representative
  - Weighting; including weighting to attempt to get a representative sample of Māori
- Analyses about to be undertaken
  - Can compare results from one vs. three studies
  - Can compare results for weighted vs. unweighted analyses