Health survey non-representativeness bias methodology and validation

Linsay Gray¹
Megan McMinn¹ (née Yates), Tommi Härkänen², Oarabile Molaodi¹, Hanna Tolonen², Alastair H. Leyland¹ and Pekka Martikainen³

1. MRC/CSO Social and Public Health Sciences Unit, University of Glasgow
2. National Institute for Health and Welfare (THL), Finland
3. Department of Sociology, University of Helsinki, Finland

BigSurv18 Barcelona 25-27 October 2018
Background I

- National health surveys are used to estimate the prevalence of various health behaviours in a population (e.g. smoking, alcohol consumption, obesity)
  - Development, implementation and evaluation of social and public health policy
  - Important that the sample is representative of the population
Background II

- Declining levels of participation
  - Loss of information
  - Validity threatened if the participants and the non-participants differ systematically
  - Difficult in many settings as we don’t know who the NP are
  - Evidence on – employment, middle vs younger age, female, higher SES/income and higher education being associated with increased participation

Motivations

- Non-participants tend to be in poorer health compared with participants
  - Greater hospitalisation rates
  - More likely to be hazardous drinkers

- Reasons for not participating vary
  - Hard to contact groups
  - Healthy volunteer effect
Methods for correcting for non-participation

- Post-survey corrections (weighting/MI) typically based on limited socio-demographic information,
  - Weights calculated based on age, sex, region, etc.
  - Not adequate if participants and non-participants with the same socio-demographic characteristics differ in other ways (health related behaviours/health status)
Alternative method

• Developed and applied in Scotland using the Scottish Health Surveys*
• Proportion participating
  • 1995 – 81%
  • 2010 – 63%
• Test case – weekly alcohol consumption

Alternative method – Application to the Scottish Health Survey (SHeS)

  - Consenting participants (89%) aged 20-64 years
  - Linked to hospitalisation and mortality records
    - Alcohol related harms (hospitalisations + deaths)
    - All-cause mortality
- Population
  - Linked hospitalisation and mortality records
  - Aggregated by sex, quintile of multiple deprivation and 5-yr age group

Alternative method – Scottish Health Survey

- Generate synthetic observations for non-participants by comparing the proportions of participants within each age-sex-deprivation group, with the same groups in the population.
  - Any deviations identify where there are non-participants

<table>
<thead>
<tr>
<th>Participants</th>
<th>Synthetic observations on Non-participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td>Age group</td>
</tr>
<tr>
<td>Sex</td>
<td>Sex</td>
</tr>
<tr>
<td>Area level deprivation</td>
<td>Area level deprivation</td>
</tr>
<tr>
<td>Alcohol-related hospitalisations</td>
<td>Alcohol-related hospitalisations</td>
</tr>
<tr>
<td>All-cause deaths</td>
<td>All-cause deaths</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>Missing</td>
</tr>
</tbody>
</table>
Missing data mechanisms

- Missing At Random (MAR)
  - Missingness depends on observed data

- Missing Not At Random (MNAR)
  - Missingness depends on unobserved data
  - Assumes alcohol-related harms are greater in non-participants, than in participants
    1. Continuum of resistance – late participants are similar to non-participants
    2. Subgroup of very heavy drinkers
       - Sex-specific mean consumption
       - 2x, 4x and 6x greater than observed mean
### Alternative method - results

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (95 % CI)/SD</td>
<td>Mean (95 % CI)/SD</td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td></td>
<td>Mean (95 % CI)/SD</td>
<td>Mean (95 % CI)/SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>21.8 (20.5–23.1)</td>
<td>10.8 (10.1–11.6)</td>
</tr>
<tr>
<td>MAR</td>
<td>22.4 (20.3–24.4)</td>
<td>10.8 (9.8–11.7)</td>
</tr>
<tr>
<td>MNAR&lt;sup&gt;CR&lt;/sup&gt;</td>
<td>24.9 (22.8–27.0)</td>
<td>11.5 (10.5–12.4)</td>
</tr>
<tr>
<td>MNAR&lt;sup&gt;a&lt;/sup&gt;</td>
<td>24.6 (22.4–26.7)</td>
<td>11.0 (10.0–12.0)</td>
</tr>
<tr>
<td>MNAR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>28.9 (26.4–31.5)</td>
<td>11.5 (10.5–12.5)</td>
</tr>
<tr>
<td>MNAR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>33.3 (30.1–36.5)</td>
<td>11.9 (10.8–13.0)</td>
</tr>
</tbody>
</table>

Alternative method – results for men by deprivation

<table>
<thead>
<tr>
<th>Quintile of deprivation</th>
<th>Survey-weighted estimates among respondents$^a$</th>
<th>MNAR$^d$ estimates in adjusted sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (95 % CI)</td>
<td>Mean (95 % CI)</td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least deprived</td>
<td>23.1 (20.9–25.3)</td>
<td>26.1 (22.0–30.2)</td>
</tr>
<tr>
<td>2</td>
<td>21.4 (19.2–23.6)</td>
<td>26.8 (21.6–31.9)</td>
</tr>
<tr>
<td>3</td>
<td>21.9 (18.8–25.0)</td>
<td>30.3 (23.9–36.7)</td>
</tr>
<tr>
<td>4</td>
<td>20.0 (17.6–22.5)</td>
<td>33.1 (27.1–39.2)</td>
</tr>
<tr>
<td>Most deprived</td>
<td>22.5 (17.7–27.3)</td>
<td>52.1 (40.5–63.8)</td>
</tr>
<tr>
<td>All quintiles</td>
<td>21.8 (20.5–23.1)</td>
<td>33.3 (30.1–36.5)</td>
</tr>
</tbody>
</table>

Validation

- Need information on non-participants
- Nordic countries have individual level population registers which can be linked to many different registers, such as socioeconomic, hospital and death records

- Validating using alcohol consumption, but could equally be applied to any other survey measured outcome, such as smoking, use of e-cigarettes, physical activity etc.
Health 2000

- Representative sample of 8,028 persons aged 30+ in 2000/01 in Finland
- Data collection methods included home interviews, health examinations and 3 questionnaires
  - Questionnaire 1 included questions on alcohol consumption
    » Average weekly alcohol consumption (g/week)
- Linked to morbidity and mortality records at the individual level
  - Participants and non-participants
  - Follow-up available until 31/12/2015

Health 2000 study

Non-participants

Participants

Linked to hospitalisation and death records
Population

- Population wide register with unique identifiers
  - Able to link socioeconomic data with hospitalisation and mortality records at the individual level
  - 11% representative sample
  - Follow-up to end of 2012
Infer number of non-participants within sociodemographic-harm strata

Create synthetic partial observations for non-participants

Combine participant and synthetic non-participant observations to multiply impute “missing” values of alcohol consumption for the synthetic non-participant observations
### Health 2000 study

- Linked to hospitalisation and death records

<table>
<thead>
<tr>
<th>Participants</th>
<th>Non-participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td>Age group</td>
</tr>
<tr>
<td>Sex</td>
<td>Sex</td>
</tr>
<tr>
<td>Socioeconomic group</td>
<td>Socioeconomic group</td>
</tr>
<tr>
<td>Alcohol-related hospitalisations</td>
<td>Alcohol-related hospitalisations</td>
</tr>
<tr>
<td>All-cause deaths</td>
<td>All-cause deaths</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td><strong>Missing</strong></td>
</tr>
</tbody>
</table>
Non-participants

- Multiply impute values of alcohol consumption

Combine participant and synthetic non-participant observations to multiply impute “missing” values of alcohol consumption for the synthetic non-participant observations

Compare inferences
Potential results

- Similar
  - methodology is valid
  - Approach can be used to improve estimates

- Dissimilar
  - Further investigation into differences between true and synthetic non-participants
Strengths and limitations

- **Strengths**
  - Individual level, record-linked data for total sample
  - 85.5% 30-79yr olds completed Questionnaire

- **Limitations**
  - Sample of population, rather than total population
  - Individual vs area level SES
Thank you

We would like to thank the participants of the Health 2000 study, National Institute for Health and Welfare (THL) and Statistics Finland for the provision of the sociodemographic, hospitalisation and death data.

Linsay.Gray@Glasgow.ac.uk