Exploiting record linkage to quantify non-response bias and improve population estimates in health surveys

Emma Gorman

MRC/CSO Social and Public Health Sciences Unit, University of Glasgow

Health Surveys User Meeting
11 July 2013, London
Outline

• Motivations and research objectives
• Data and results
• A novel methodology
• Discussion
Motivations (I)

- The hazardous and harmful use of alcohol is a leading risk factor contributing to global death, disease and injury –

- Paired with the wider economic and social costs associated with alcohol consumption, this prompts substantial policy interest in mitigating alcohol-related harms

- Reliable measurement of alcohol consumption is required to inform policy planning and evaluation.

Figure: Time trends in age-standardised mortality for liver cirrhosis per 100,000: 1950-2000*

Motivations (II)

- Reliable population trends in alcohol consumption can be derived from **retail sales data** -
  - Available only at an aggregate (per capita) level

- National health **surveys** are an important resource for exploring **patterns** of drinking –
  - Distortions due to non-response, self-report bias, *etc.*

- Inconsistencies between estimated level and trends in population alcohol consumption

- **Declining** response levels may be a key factor.
Response levels and alcohol consumption

**Table:** Response levels and alcohol consumption estimates in men in the Scottish Health Surveys, retail-based consumption estimates and population male alcohol-related mortality in Scotland 1995-2011

<table>
<thead>
<tr>
<th>Survey year</th>
<th>Household response level (%)</th>
<th>Adult response level (%)</th>
<th>Achieved adult sample</th>
<th>Consent to linkage (%)</th>
<th>Mean alcohol units per week in men</th>
<th>National retail data</th>
<th>National mortality data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>81</td>
<td>84</td>
<td>7932</td>
<td>93</td>
<td>20.1†</td>
<td>41712</td>
<td>531</td>
</tr>
<tr>
<td>1998</td>
<td>77</td>
<td>76</td>
<td>9047</td>
<td>92</td>
<td>19.8†</td>
<td>43770§</td>
<td>755</td>
</tr>
<tr>
<td>2003</td>
<td>67</td>
<td>54</td>
<td>8148</td>
<td>91</td>
<td>19.8†</td>
<td>47175</td>
<td>1056</td>
</tr>
<tr>
<td>2008</td>
<td>61</td>
<td>54</td>
<td>6465</td>
<td>86</td>
<td>18.0‖</td>
<td>50346</td>
<td>971</td>
</tr>
<tr>
<td>2009</td>
<td>64</td>
<td>56</td>
<td>7531</td>
<td>85</td>
<td>17.5‖</td>
<td>50842</td>
<td>837</td>
</tr>
<tr>
<td>2010</td>
<td>63</td>
<td>55</td>
<td>7245</td>
<td>86</td>
<td>16.0‖</td>
<td>50524</td>
<td>909</td>
</tr>
<tr>
<td>2011</td>
<td>66</td>
<td>56</td>
<td>7544</td>
<td>86</td>
<td>15.0‖</td>
<td>48746</td>
<td>815</td>
</tr>
</tbody>
</table>

*Nielsen/CGA Strategy sales in Scotland dataset (off-trade sales in 2011 adjusted to account for the loss of discount retailers).21
†General Register Office for Scotland figures for 2011.51
‡The 1995 and 1998 surveys were prior to the significant change in the way in which alcohol consumption estimates were derived and are for men aged 16–64 only; thus, they are not comparable with those for 2003 onwards.
§Data not available for 1998—estimate interpolated from available figures for 1995 and 2000;
‖The estimates for the surveys from 2003 onwards are for men aged 16 and over.

Motivations (III)

- Falling response levels are problematic when accompanied by a loss of representativeness.

- Exploring non-response bias and making post-survey adjustments requires auxiliary information:
  - Information on non-responders, or target population

- This is commonly limited to socio-demographic information – eg, survey weights

- A lack of health-related information limits ability to correct for health-related non-response bias.
Research objectives

- Assess representativeness of Scottish Health Surveys in terms of subsequent **alcohol-related harms**
  - Compare rates of alcohol-related harm subsequent to interview in the SHeS respondents with the general population

- Develop a methodology to **address** any bias by harnessing information from record-linkage

- Obtain **improved survey-based estimates** of aggregate population consumption and **patterns** of drinking.
Outline

- Motivations and research objectives
- **Data and results**
- A novel methodology
- Discussion
Data – Linked Scottish Health Surveys

- Repeated cross-section - designed to inform on health of Scottish population living in private households

- Consenting SHeS respondents confidentially linked at an individual level to **morbidity** and **mortality** data (99% accurate; 90% complete); linkage performed until end of 2011

- **Alcohol-related** deaths and hospitalisations

- Despite falling response rates – linkage consent rates remain relatively high [93% in 1995; 86% in 2011]
Data – General population

- Contemporaneous comparison datasets representing the general population in each survey year are constructed from:
  - Census data - population counts in each survey year
  - Morbidity and mortality data for the general population of Scotland
  - Aggregated by age group, sex, region, area deprivation.
Results

Directly age-standardised, survey weighted incidence rates (a) and rate ratios (b) of subsequent alcohol-related harm in the 2003 SHeS respondents compared with the general population*

*Preliminary.
Outline

• Motivations and research objectives
• Data and results
• A novel methodology
• Discussion
A novel solution (I)

- **Compare** alcohol-related hospitalisations/deaths in SHeS-SMR responders with general population by age, sex, area deprivation & health board region

- Deviations from representativeness allow us to identify **who we are missing** in each group

- “**Simulate**” observations within these categories for unit non-responders to create a “representative sample”

- **Alcohol intake** remains **missing** for the unit non-responders –
  - **Multiply impute** alcohol consumption
A novel solution (II)

- Missingness mechanisms –
  - Missing completely at random (MCAR)
  - **Missing at random** (MAR)
  - Missing not at random (MNAR)

- **Missing at random** is a useful starting point
  - However underlying assumptions may be untenable

- We will **test the sensitivity** of our results by investigating various departures from a missing-at-random assumption.*

Outline

- Motivations and research objectives
- Data and results
- A novel methodology
- Discussion
Discussion

- Key strength is in the use of **reliable record-linkage** to inform exploration of bias and imputation model.

- Possible distortion from non-consenters to linkage:
  - Intend to explore any possible bias.

- Indeterminate outcomes in those who **emigrate**:
  - Acquiring further data identify and adjust for potential emigrants whose follow-up may be incomplete.
Conclusions

• We establish rates of subsequent alcohol-related harms are lower in the SHeS respondents relative to the general population of Scotland –
  o Implies socio-demographic corrections may not be sufficient to adjust for health-related differences

• Plan to generate enhanced population estimates of alcohol consumption – further adjusted for non-response bias

• We offer a methodology to increase the validity of survey data in a larger context of declining response rates; with potential general application.
Acknowledgements

Collaborators:

Alastair H Leyland\textsuperscript{1}, Gerry McCartney\textsuperscript{2}, Ian R White\textsuperscript{3}, S Vittal Katikireddi \textsuperscript{1}, Lisa Rutherford\textsuperscript{4}, Lesley Graham\textsuperscript{5}, Linsay Gray\textsuperscript{1}

\textsuperscript{1} MRC/CSO SPHSU, University of Glasgow, Glasgow
\textsuperscript{2} NHS Health Scotland, Glasgow
\textsuperscript{3} MRC Biostatistics Unit, Cambridge
\textsuperscript{4} Scotcen Social Research, Edinburgh
\textsuperscript{5} ISD, NHS NSS, Edinburgh

- Data were provided by -
  - Information Services Division Scotland and the Scottish Government.

- Funding was provided by the MRC

- Contact - emma.gorman@glasgow.ac.uk