
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)

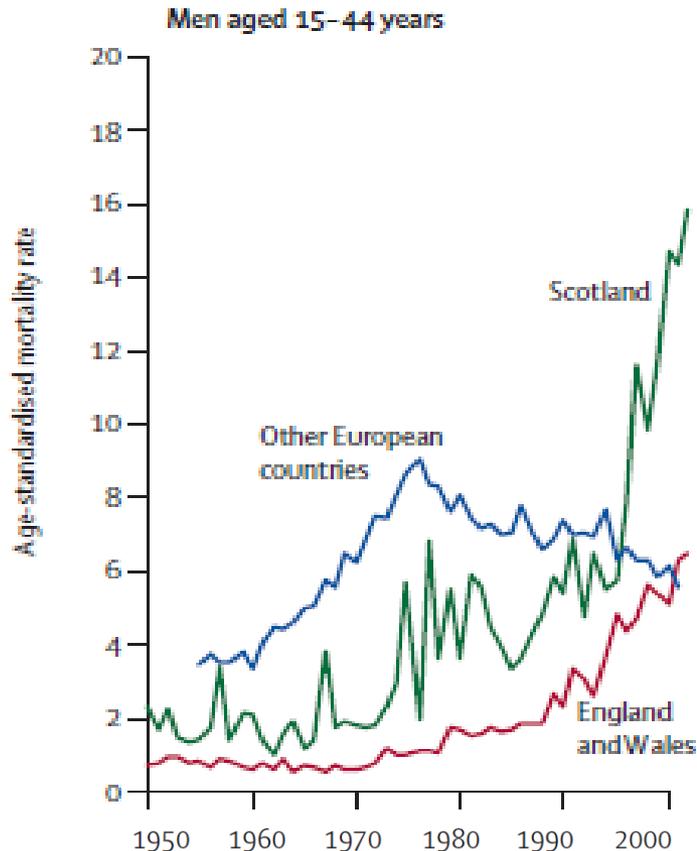


Figure: Time trends in age-standardised mortality for liver cirrhosis per 100,000: 1950-2000*

* Leon, D. A., & McCambridge, J. (2006). Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. *The Lancet*, 367(9504), 52-56.

- 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.

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

Survey data					National retail data	National mortality data	
Survey year	Household response level (%)	Adult response level (%)	Achieved adult sample	Consent to linkage (%)	Mean alcohol units per week in men	Total volume of pure alcohol sold (1000 l) [*]	Number of male alcohol-related deaths [†]
1995	81	84	7932	93	20.1 [‡]	41712	531
1998	77	76	9047	92	19.8 [‡]	43770 [§]	755
2003	67	54	8148	91	19.8 [¶]	47175	1056
2008	61	54	6465	86	18.0 [¶]	50346	971
2009	64	56	7531	85	17.5 [¶]	50842	837
2010	63	55	7245	86	16.0 [¶]	50524	909
2011	66	56	7544	86	15.0 [¶]	48746	815

^{*}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.

* Gray L, McCartney G, White IR, *et al* . Use of record-linkage to handle non-response and improve alcohol consumption estimates in health survey data: a study protocol. *BMJOpen* 2013; 3.

MRC/CSO Social and Public Health Sciences Unit

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

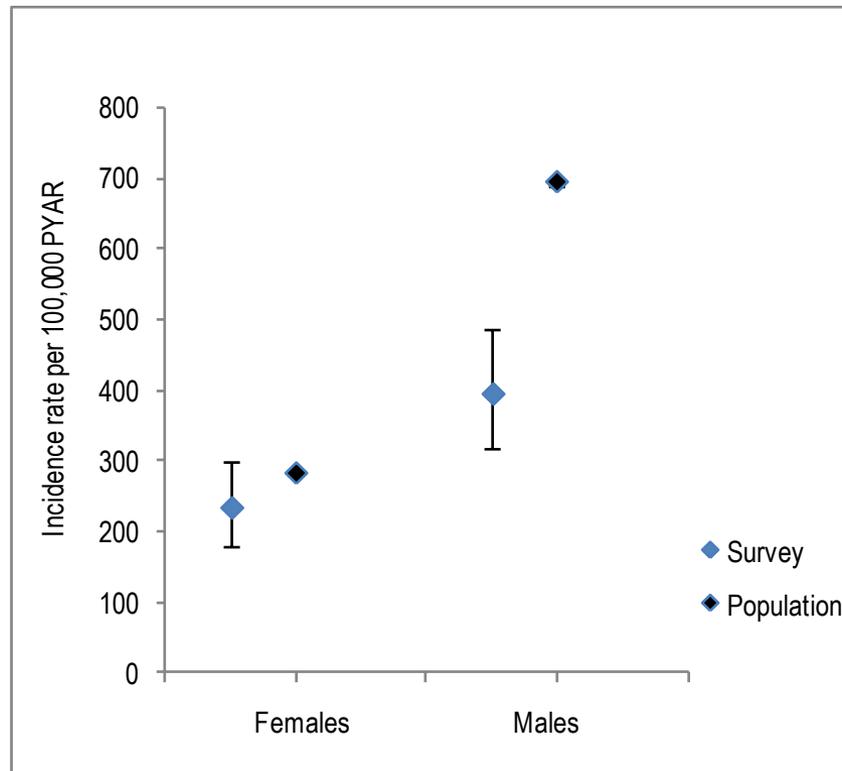
- Repeated cross-section - designed to inform on health of Scottish population living in private households
 - 1995, 1998, 2003, 2008, 2009 and 2010
- Consenting SHeS respondents confidentially linked at an individual level to **morbidity** and **mortality** data (99% accurate; 90% complete); linkage 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

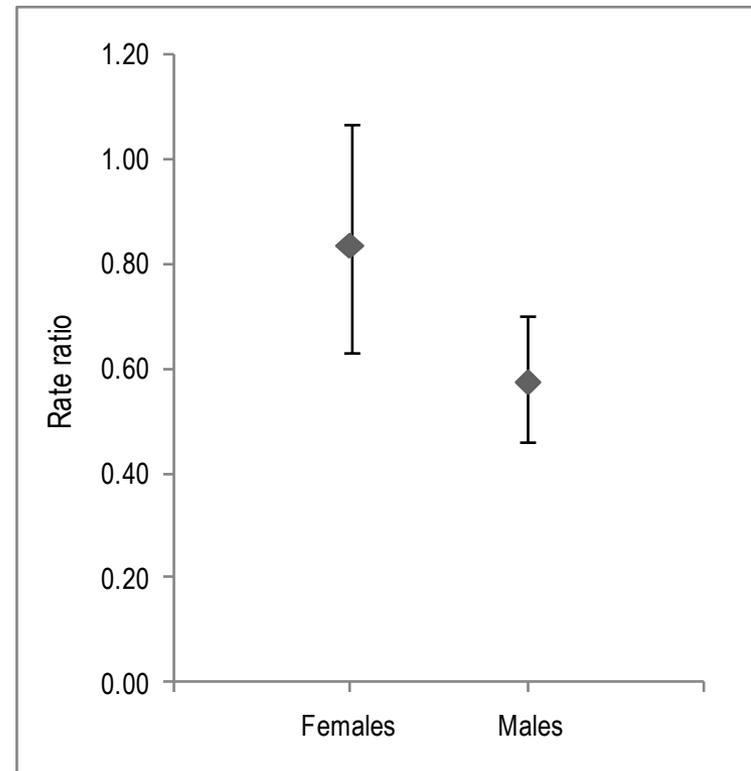
- Contemporaneous comparison datasets representing the general population in each survey year are constructed from:
- Census data - population counts in each survey year
- **Morbidity** and **mortality** data for the general population of Scotland
- Aggregated by age group, sex, region, area deprivation.

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*



(a)



(b)

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.*

* Gray L, McCartney G, White IR, *et al* . Use of record-linkage to handle non-response and improve alcohol consumption estimates in health survey data: a study protocol. *BMJOpen* 2013; 3.

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 –
 - 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 Leyland¹, Gerry McCartney², Ian R White³, S Vittal Katikireddi¹, Lisa Rutherford⁴, Lesley Graham⁵, Linsay Gray¹

¹ MRC/CSO SPHSU, University of Glasgow, Glasgow

² NHS Health Scotland, Glasgow

³ MRC Biostatistics Unit, Cambridge

⁴ Scotcen Social Research, Edinburgh

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