MRC/CSO Social and Public Health Sciences Unit




Health survey non-representativeness bias methodology and validation

Linsay Gray[1]

Megan McMinn1 (née Yates), Tommi Härkänen[2], Oarabile Molaodi[1], Hanna Tolonen[2], Alastair H. Leyland[1] and Pekka Martikainen[3]

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

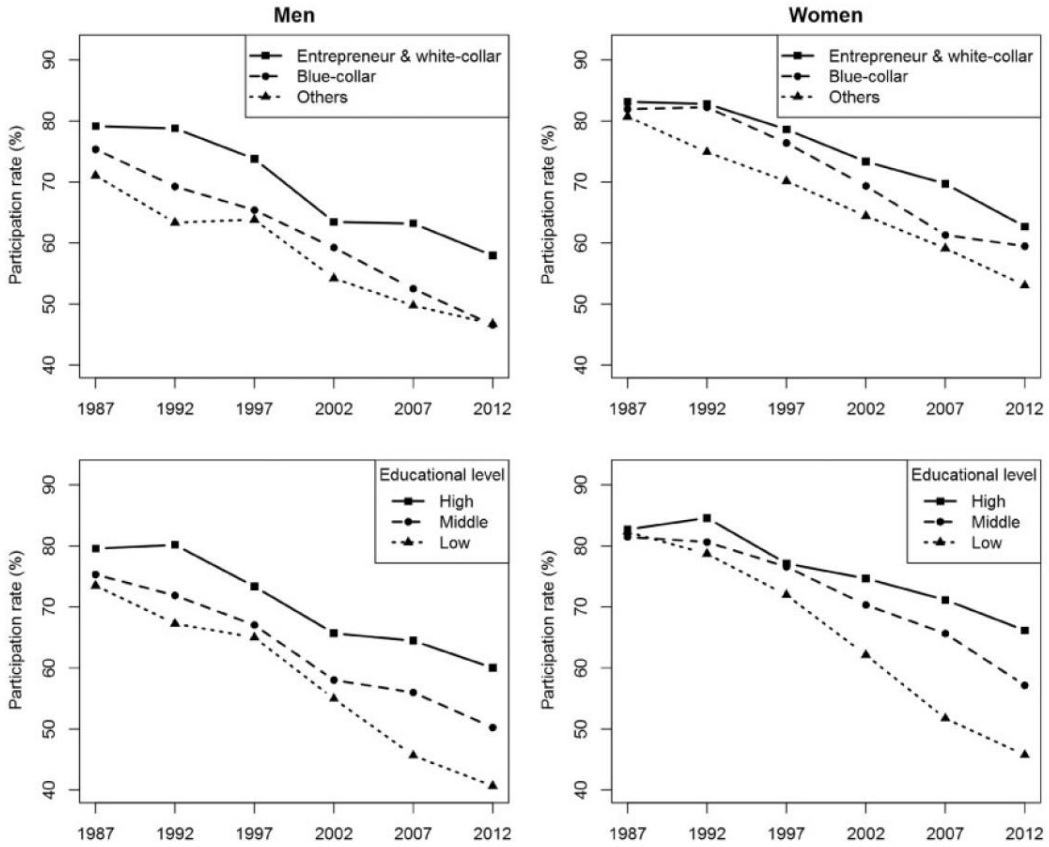
BigSurv18 Barcelona 25-27 October 2018

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

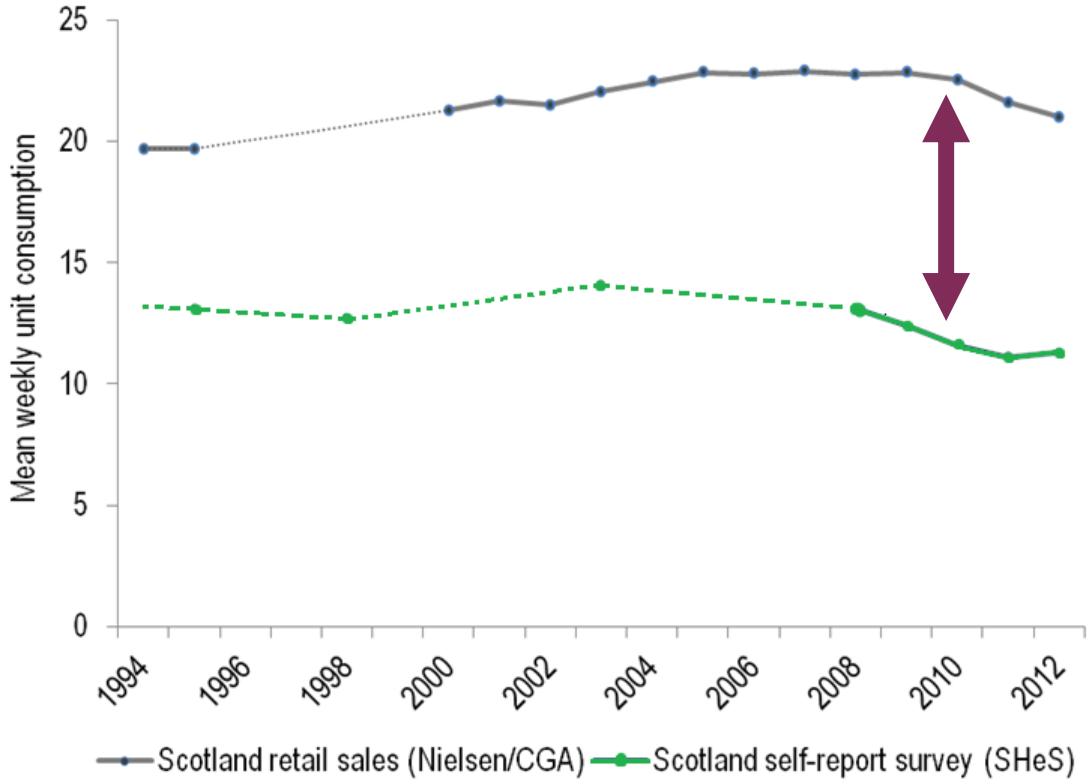


REINIKAINEN, J., et. al. 2018. Participation rates by educational levels have diverged during 25 years in Finnish health examination surveys. *European Journal of Public Health,* 28**,** 237-243.

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


25
20
15
10
5
0
Mean weekly unit consumption
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Scotland retail sales (Nielsen/CGA)
Scotland self-report survey (SHeS)

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

* Gorman E, Leyland AH, McCartney G*, et al.* Adjustment for survey non-representativeness using record-linkage: refined estimates of alcohol consumption by deprivation in Scotland. *Addiction* 2017;112:1270-80.

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

- SHeS 1995, 1998, 2003 and 2008-10
  - 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

***** Gorman E, Leyland AH, McCartney G*, et al.* Adjustment for survey non-representativeness using record-linkage: refined estimates of alcohol consumption by deprivation in Scotland. *Addiction* 2017;112:1270-80.

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

Participants	Synthetic observations on Non-participants
Age group	Age group
Sex	Sex
Area level deprivation	Area level deprivation
Alcohol-related hospitalisations	Alcohol-related hospitalisations
All-cause deaths	All-cause deaths
Alcohol consumption	**Missing**

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

2003	Males		Females	
	Mean	*(95 % CI)/SD*	*Mean*	*(95 % CI)/SD*
Participants	21.8	(20.5–23.1)	10.8	(10.1–11.6)
MAR	22.4	(20.3–24.4)	10.8	(9.8–11.7)
$MNAR^{CR}$	24.9	(22.8–27.0)	11.5	(10.5–12.4)
$MNAR^{a}$	24.6	(22.4–26.7)	11.0	(10.0–12.0)
$MNAR^{b}$	28.9	(26.4–31.5)	11.5	(10.5–12.5)
$MNAR^{c}$	33.3	(30.1–36.5)	11.9	(10.8–13.0)

Gorman E, Leyland AH, McCartney G*, et al.* Adjustment for survey non-representativeness using record-linkage: refined estimates of alcohol consumption by deprivation in Scotland. *Addiction* 2017;112:1270-80.

Alternative method – results for men by deprivation

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

Gorman E, Leyland AH, McCartney G*, et al.* Adjustment for survey non-representativeness using record-linkage: refined estimates of alcohol consumption by deprivation in Scotland. *Addiction* 2017;112:1270-80.

Validation

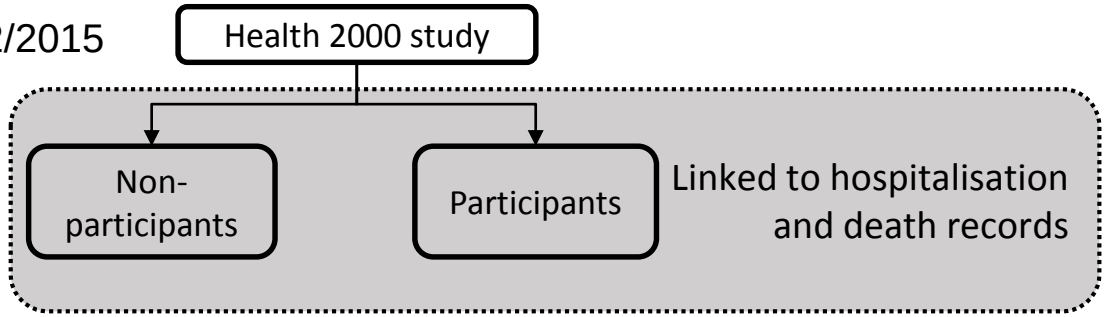
- Need information on non-participants
- Nordic countries have individual level population registers which can be linked to many different registers, such as socioeconomic, hospital and death records

- 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

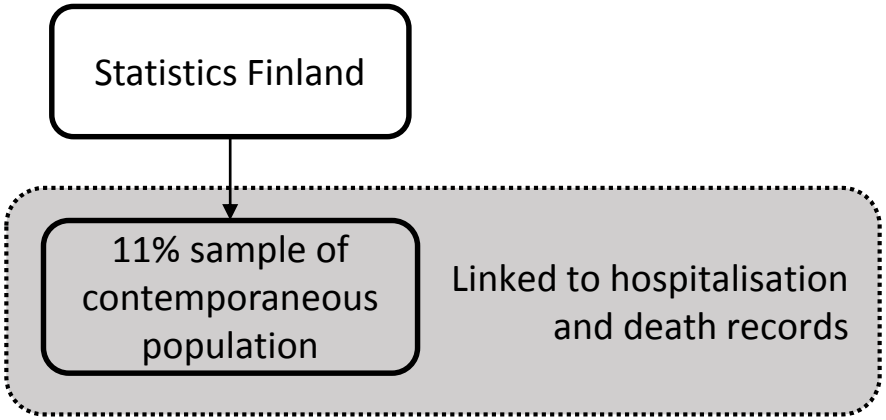
- Representative sample of 8,028 persons aged 30+ in 2000/01 in Finland
- Data collection methods included home interviews, health examinations and 3 questionnaires
  - Questionnaire 1 included questions on alcohol consumption
    - » *Average weekly alcohol consumption (g/week)*

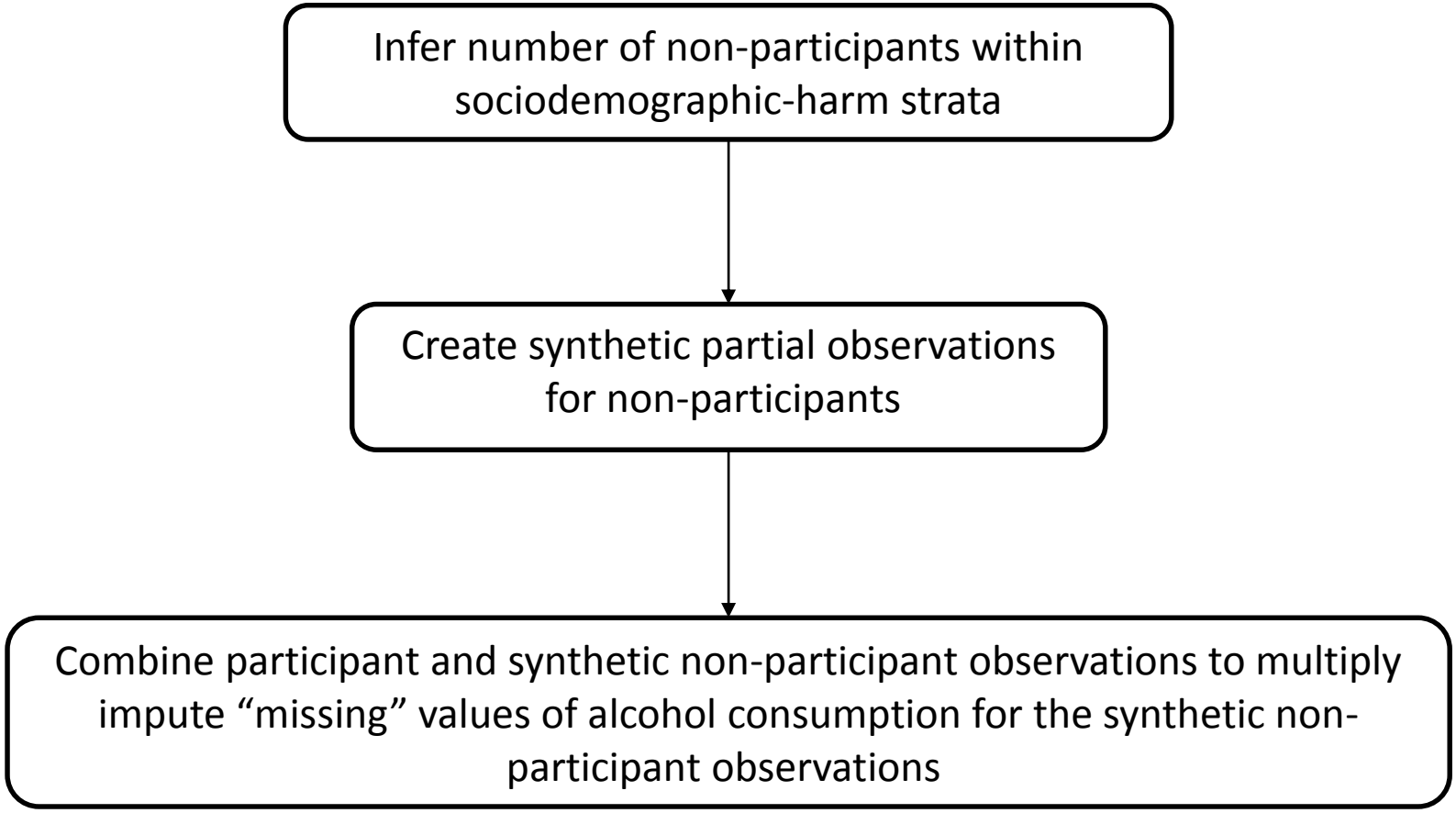
- Linked to morbidity and mortality records at the individual level
  - Participants **and** non-participants
  - Follow-up available until 31/12/2015

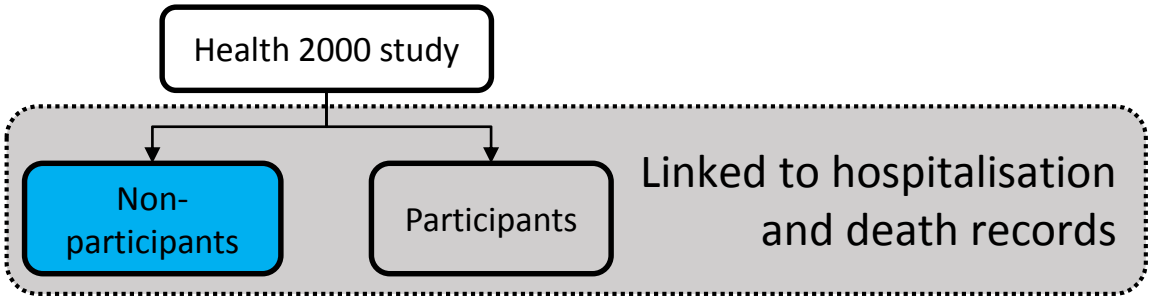


Population

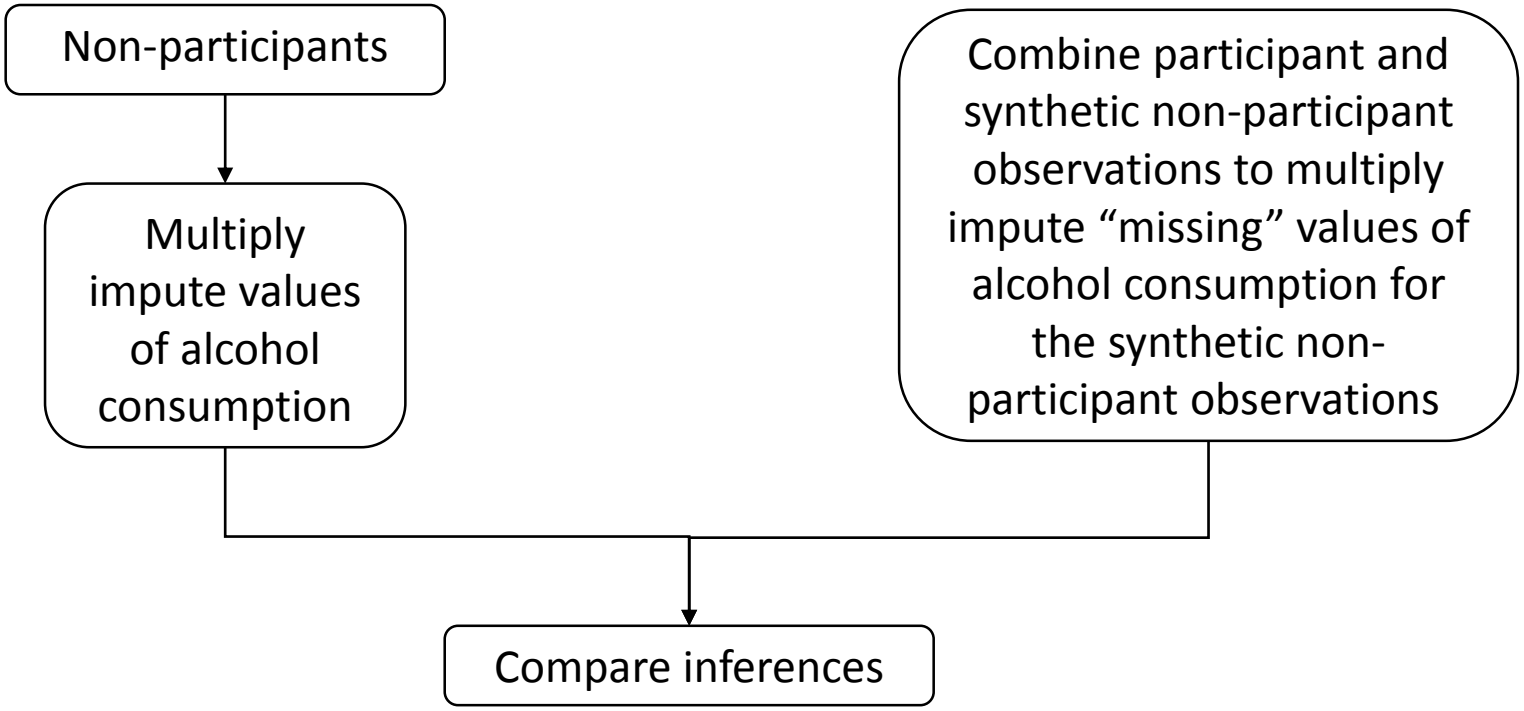
- Population wide register with unique identifiers
  - Able to link socioeconomic data with hospitalisation and mortality records at the individual level
  - 11% representative sample
  - Follow-up to end of 2012




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



Participants	Non-participants
Age group	Age group
Sex	Sex
Socioeconomic group	Socioeconomic group
Alcohol-related hospitalisations	Alcohol-related hospitalisations
All-cause deaths	All-cause deaths
Alcohol consumption	**Missing**


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