Health Survey for England (HSE) data harmonization for cardiovascular disease (CVD) risk trend analyses

Shaun Scholes

Explaining Population Trends in Cardiovascular Risk: A Research Dissemination and Knowledge Exchange Event

University of Greenwich

September 4 2023

National Centre for Social Research

s.scholes@ucl.ac.uk

Background

- ExPoSE: employs data from nationally-representative repeated cross-sectional surveys conducted in England since 1998 with the goal to:

  - Describe recent trends in CVD risk;
  - Identify demographic, behavioural, social and environmental factors that may explain the observed trends and to what extent;
  - Determine the amount of population change in CVD risk that is unaccounted for by these factors

Background

- Creation of a harmonized dataset for merging and consolidation of the annual Health Surveys for England (HSE)

- Analysis of the data by the ExPoSE team

- The dataset, code and documentation will be useful for others to use in their own future research

Creating a harmonized dataset using HSE

- HSE is an annual cross-sectional survey, but with an important focus on monitoring trends….
- **But** very often this does not require a harmonized dataset
- Harmonized datasets of HSE data (combining data across years) have been created for various projects, including the recent report on ethnicity and health https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-england-additional-analyses/ethnicity-and-health-2011-2019-experimental-statistics

European Journal of Public Health, 1–9

https://doi.org/10.1093/eurpub/ckad146

Ethnic differences in multimorbidity after accounting for social-economic factors, findings from The Health Survey for England

Linda Ng Fat [1], Priyanka Patil [1,2], Jennifer S. Mindell [1], Logan Manikam[1,2], Shaun Scholes[1]

1 Health and Social Surveys Group, Research Department of Epidemiology and Public Health, University College London (UCL), London, UK
2 Aceso Global Health Consultants Pte Limited, Singapore, Singapore

Correspondence: Linda Ng Fat, Health and Social Surveys Group, Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 7HB, UK, Tel: +442076798235, e-mail l.ngfat@ucl.ac.uk

Background: Social-economic factors and health behaviours may be driving variation in ethnic health inequalities in multimorbidity including among distinct ethnic groups. **Methods:** Using the cross-sectional nationally representative Health Surveys for England 2011–18 ($N = 54\ 438$, aged 16+), we performed multivariable logistic regression on the odds of having general multimorbidity (≥2 longstanding conditions) by ethnicity [British White (reference group), White Irish, Other White, Indian, Pakistani, Bangladeshi, Chinese, African, Caribbean, White mixed, Other Mixed], adjusting for age, sex, education, area deprivation, obesity, smoking status and survey year. This was repeated for cardiovascular multimorbidity ($N = 37\ 148$, aged 40+: having ≥2 of the following: self-reported diabetes, hypertension, heart attack or stroke) and multiple cardiometabolic risk biomarkers (HbA_{1c} ≥6.5%, raised blood pressure, total cholesterol ≥5mmol/L). **Results:** Twenty percent of adults had general multimorbidity. In fully adjusted models, compared with the White British majority, Other White [odds ratio (OR) = 0.63; 95% confidence interval (CI) 0.53–0.74], Chinese (OR = 0.58, 95% CI 0.36–0.93) and African adults (OR = 0.54, 95% CI 0.42–0.69), had lower odds of general multimorbidity. Among adults aged 40+, Pakistani (OR = 1.27, 95% CI 0.97–1.66; $P = 0.080$) and Bangladeshi (OR = 1.75, 95% CI 1.16–2.65) had increased odds, and African adults had decreased odds (OR = 0.63, 95% CI 0.47–0.83) of general multimorbidity. Risk of cardiovascular multimorbidity was higher among Indian (OR = 3.31, 95% CI 2.56–4.28), Pakistani (OR = 3.48, 95% CI 2.52–4.80), Bangladeshi (OR = 3.67, 95% CI 1.98–6.78), African (OR = 1.61, 95% CI 1.05–2.47), Caribbean (OR = 2.18, 95% CI 1.59–2.99) and White mixed (OR = 1.98, 95% CI 1.14–3.44) adults. Indian adults were also at risk of having multiple cardiometabolic risk biomarkers. **Conclusion:** Ethnic inequalities in multimorbidity are independent of social-economic factors. Ethnic minority groups are particularly at risk of cardiovascular multimorbidity, which may be exacerbated by poorer management of cardiometabolic risk requiring further investigation.

Creating a harmonized dataset using HSE

- Perhaps the main challenge in creating harmonized datasets using HSE is to ensure that the combined data can be used to **replicate** the published trend estimates
- Extensive documentation, plus experience with HSE, was very useful in creating the dataset
- I compiled the dataset with valuable input from the ExPoSE team, with a central focus on CVD among adults between 1998 and 2017

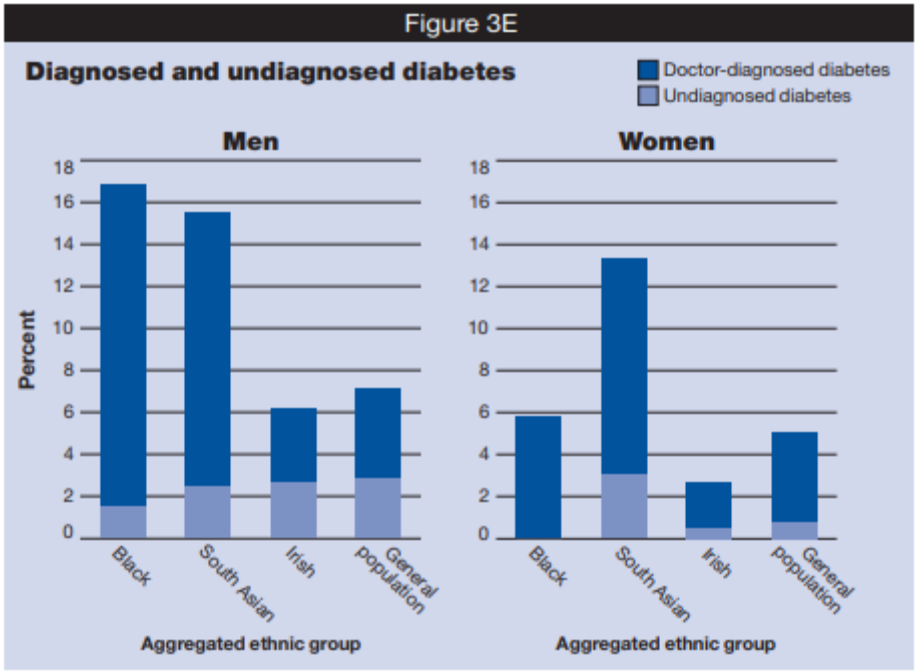
Creating a harmonized dataset using HSE

- Challenges in compiling this dataset to examine CVD risk over time included:
  - Boost samples of population subgroups (minority ethnic groups; older persons)
  - Gaps in survey years (core / additional content)
  - Random allocation of questionnaire modules
  - Changes in question wording (shift in focus; harmonization with other surveys)
  - Changes in variable derivation / names
  - Changes in measuring equipment
  - Changes in methods (such as non-response weighting)
  - Stricter guidelines regarding data release

Content of presentation

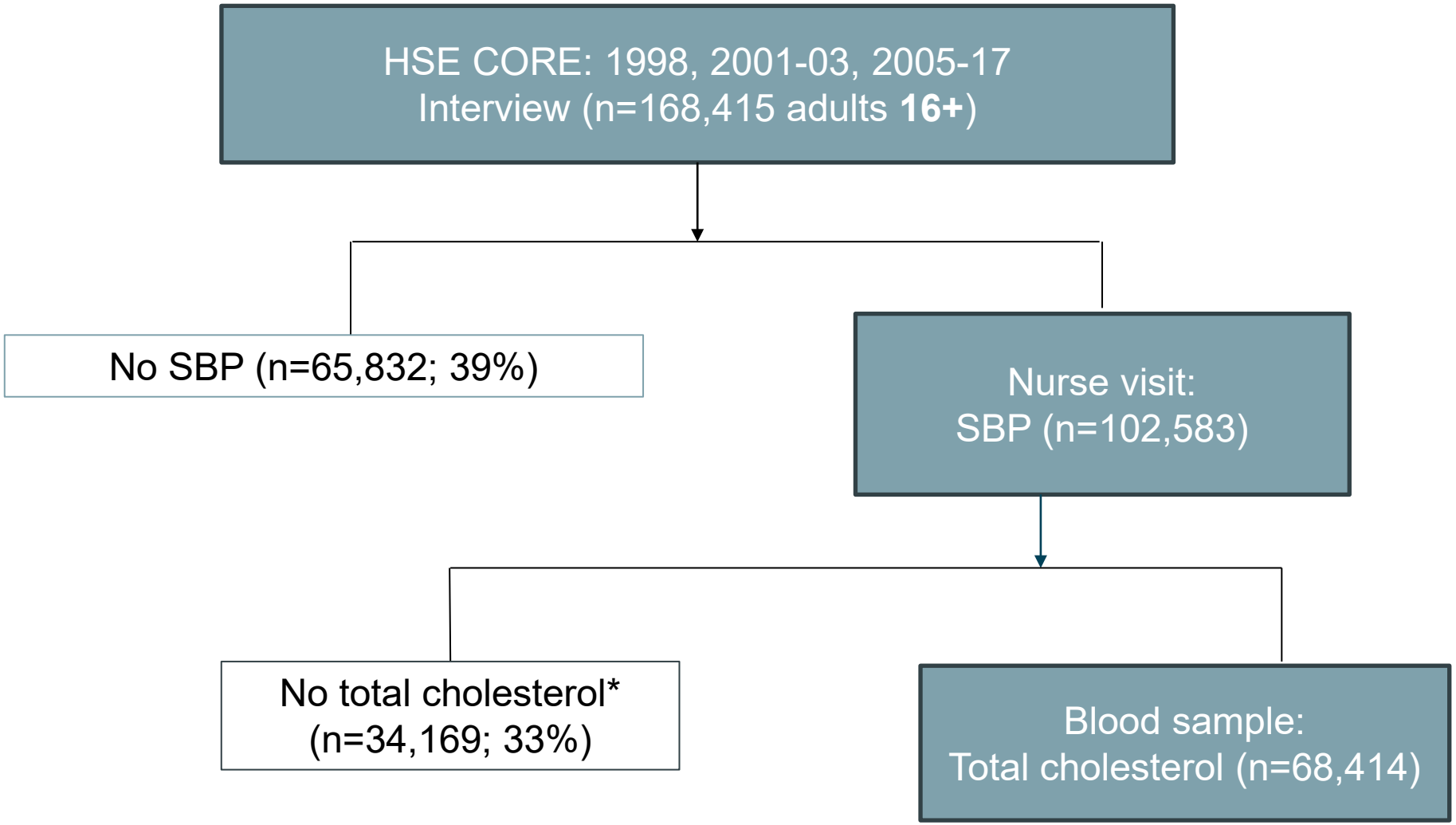
- Describe some the key variables in the dataset, whilst illustrating the challenges in harmonization

Boost samples of minority ethnic groups…excluded from harmonized dataset



Boost sample of minority ethnic groups in HSE 2004: a CVD focus but no measurements of blood pressure in the general population sample (CVD risk scores could not be calculated)

Inclusion in the harmonized dataset



*gaps in survey years

CVD RISK SCORE & COMPONENTS

Equation and survey years	
Laboratory: 1998, 2003, 2006*, 2009-17	• Age • Sex • Current smoking • Doctor-diagnosed diabetes • SBP • Total cholesterol
Non-laboratory: 1998, 2001-03, 2005-17	• Age • Sex • Current smoking • BMI • SBP

*Random allocation of questionnaire module:

In 2006, participants aged 65 and over were allocated at random to one of two questionnaire versions. This included either: (1) the CVD (including diabetes) and short physical activity modules, or (ii) the long physical activity module but not the CVD module. Adults aged 16-64 completed both the CVD and long physical activity modules.

CVD risk components

- ***Age (continuous)*:** (stricter guidelines regarding data release)
  - 2014 / 2015 (SL): age 90+ coded to 92
  - 2016 / 2017: mid-point of **age-band**

16-17	16.5	**50-54**	52
18-19	18.5	**55-59**	57
20-24	22	**60-64**	62
25-29	27	**65-69**	67
30-34	32	**70-74**	72
35-39	37	**75-79**	77
40-44	42	**80-84**	82
45-49	47	**85-89**	87
		90+	92

CVD risk components

- ***Smoking status*:** never; former; current cigarette smoker
- ***SBP*:** mean of 2nd and 3rd readings: Dinamap readings prior to 2003 converted to predicted Omron values via a calibration equation to allow comparison over years (changes in measuring equipment)

For systolic blood pressure:

Predicted Omron 8.90 (SE 2.94)
+ 0.91 (SE 0.02) × *Dinamap (1)

For diastolic blood pressure:

Predicted Omron 19.78 (SE 1.86)
+ 0.73 (SE 0.03) × *Dinamap (2)

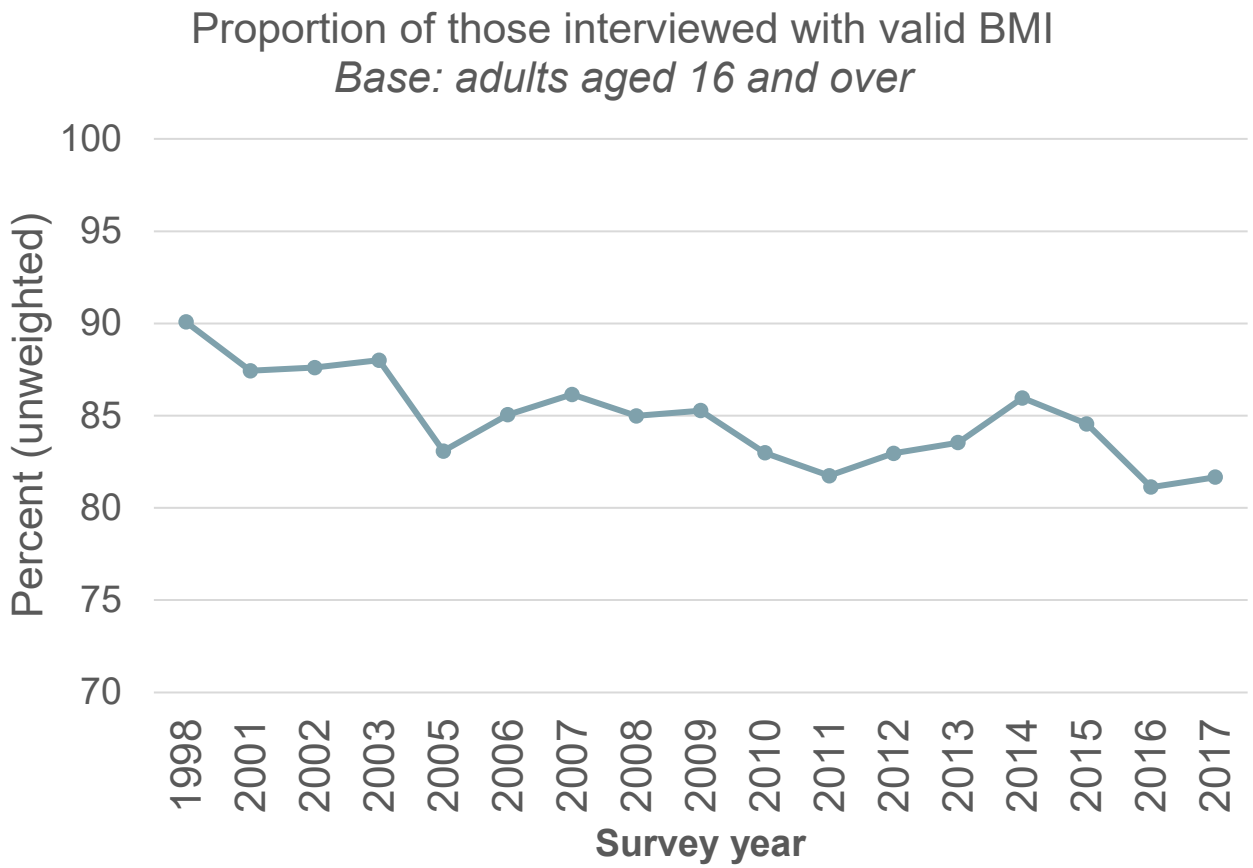
Improvement in hypertension management in England: results from the Health Survey for England 2003
Paola Primatesta[a] and Neil R. Poulter[b]

CVD risk components

***BMI (continuous)*: interviewer-measured height and weight**

- Changes in measuring equipment: More accurate Class III Seca scales from 2011 onwards: measure up to a maximum of 200kg (compared with maximum of 130kg in earlier models): use respondent values if estimated weight above maximum values
- Changes in variable name (wtval2 and bmival2 since 2012 reflecting change in derivation)

- Increasing proportion of missing data on BMI over time


Proportion of those interviewed with valid BMI
Base: adults aged 16 and over
Percent (unweighted)
100
95
90
85
80
75
70
1998
2001
2002
2003
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
Survey year

CVD risk components

- ***Doctor-diagnosed diabetes:*** *"Do you now have, or have you ever had diabetes?", "Were you told by a doctor that you had diabetes?"*
- *Gaps in survey years:* included in 1998; 2003; 2006*; 2009-17

Random allocation of questionnaire module

CVD risk components

- ***Total cholesterol:*** non-fasting blood samples
- Estimates of raised total cholesterol (≥5mmol/L) include those taking lipid-lowering medicines
- *Gaps in survey years:* included in 1998; 2003; 2006; 2008-17
- *Changes in measuring equipment and variable name:* adjustment (-0.1mmol/L) between April 2010 and June 2015 to allow comparison with earlier and later years

SOCIO-ECONOMIC FACTORS, MEASURES OF HEALTH, AND INDICATORS OF HEALTHY LIFESTYLES

Demographics

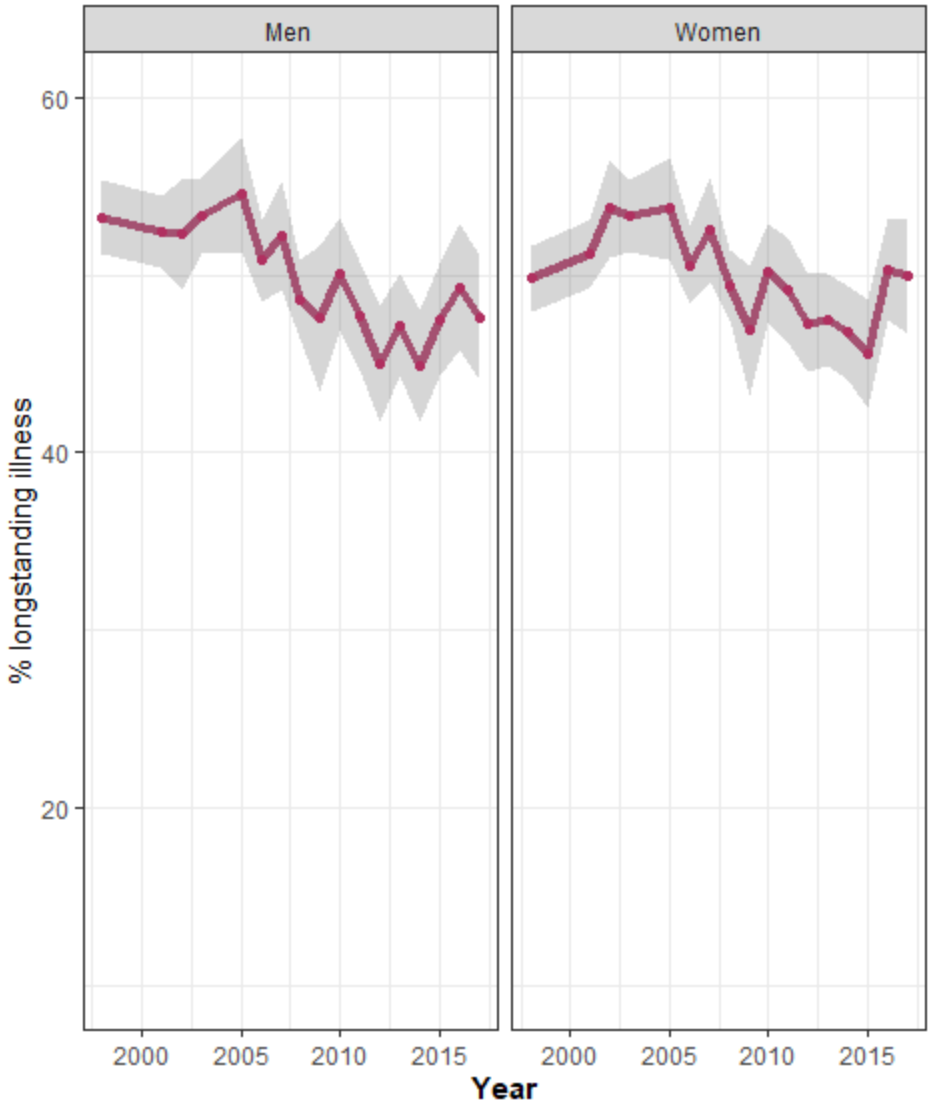
Marital status	• **Single** • **Married** • **Separated** • **Divorced** • **Widowed** • **Cohabitee (from 2002)**
Ethnicity	• White • Black • Asian • Mixed • Any other
Region	Government Office Region
Urban/rural	• Derived: Address Record Form (interviewer observation) • Sample: England & Wales Urban/Rural indicators (not archived in later years due to confidentiality)

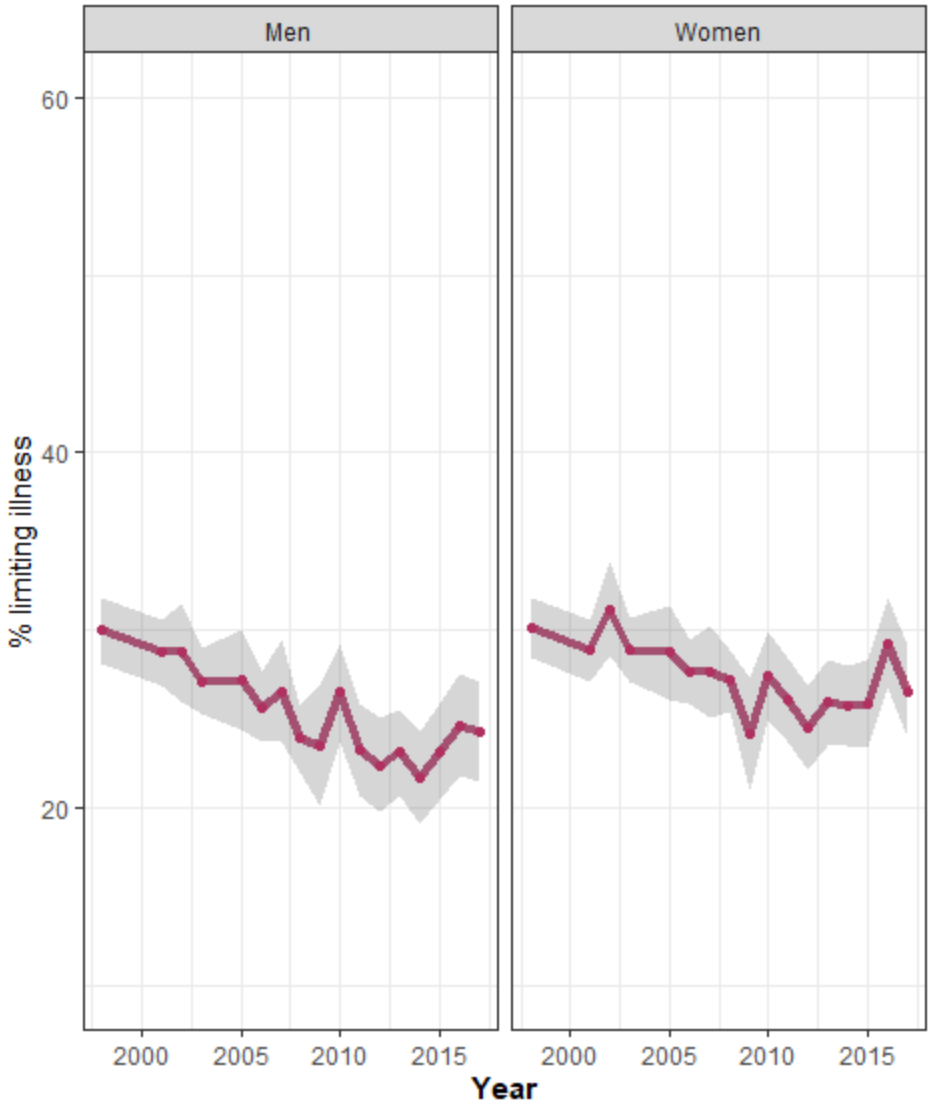
Measures of socioeconomic status

- **Area:**
  - (National) Index of Multiple Deprivation (IMD) quintiles (different releases over time but consistent geography; from 2001 onwards)
- **Household:**
  - Tenure
  - Car ownership
  - Income quintiles (equivalised)
  - Sources of income (e.g. state benefits)
- **Individual-level:**
  - Highest educational qualification
  - Social class (Registrar General) via SOC codes

Measures of health

- Long-standing illness (change in question):
  - Prior to 2012: ‘an illness, disability or infirmity…that has troubled you over a period of time or that is likely to affect you over a period of time’;
  - 2012 onwards: ‘any physical or mental health condition or illness lasting or expected to last 12 months or more’
- Long-standing **limiting** illness:
  - ‘Does this illness….limit your activities in any way’


Men
Women
60
40
20
% longstanding illness
2000
2005
2010
2015
Year

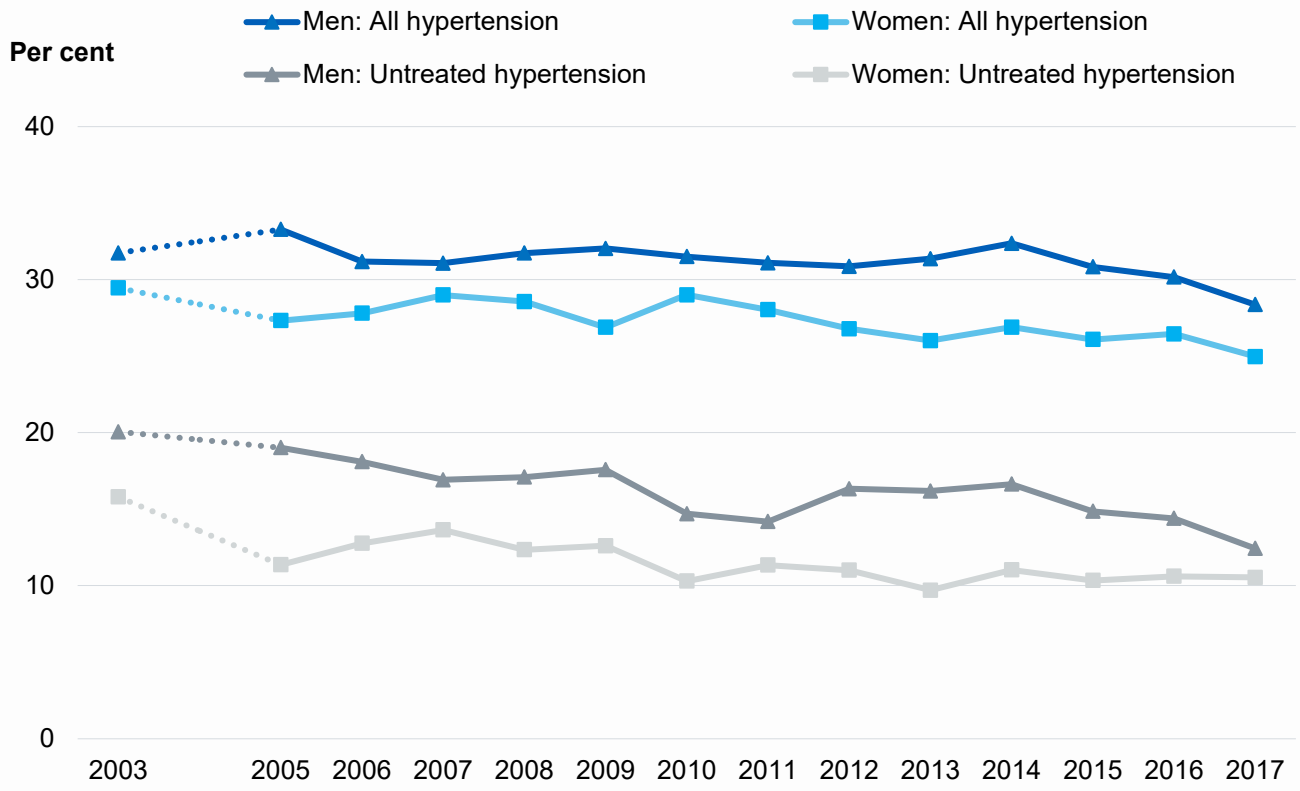

Men
Women
60
40
20
% limiting illness
2000
2005
2010
2015
Year

Measures of health

- Survey defined hypertension:
  - Normotensive untreated: SBP below 140mmHg and DBP below 90mmHg, not currently taking medication for blood pressure.
  - **Hypertensive controlled**: SBP below 140mmHg and DBP below 90mmHg, currently taking medication for blood pressure.
  - **Hypertensive uncontrolled**: SBP at or greater than 140mmHg and/or DBP at or greater than 90mmHg, currently taking medication for blood pressure
  - **Hypertensive untreated**: SBP at or greater than 140mmHg and/or DBP at or greater than 90mmHg, not currently taking medication for blood pressure.

Total and untreated hypertension by sex, 2003-2017

Base: Aged 16 and over with a nurse visit and valid blood pressure measurements



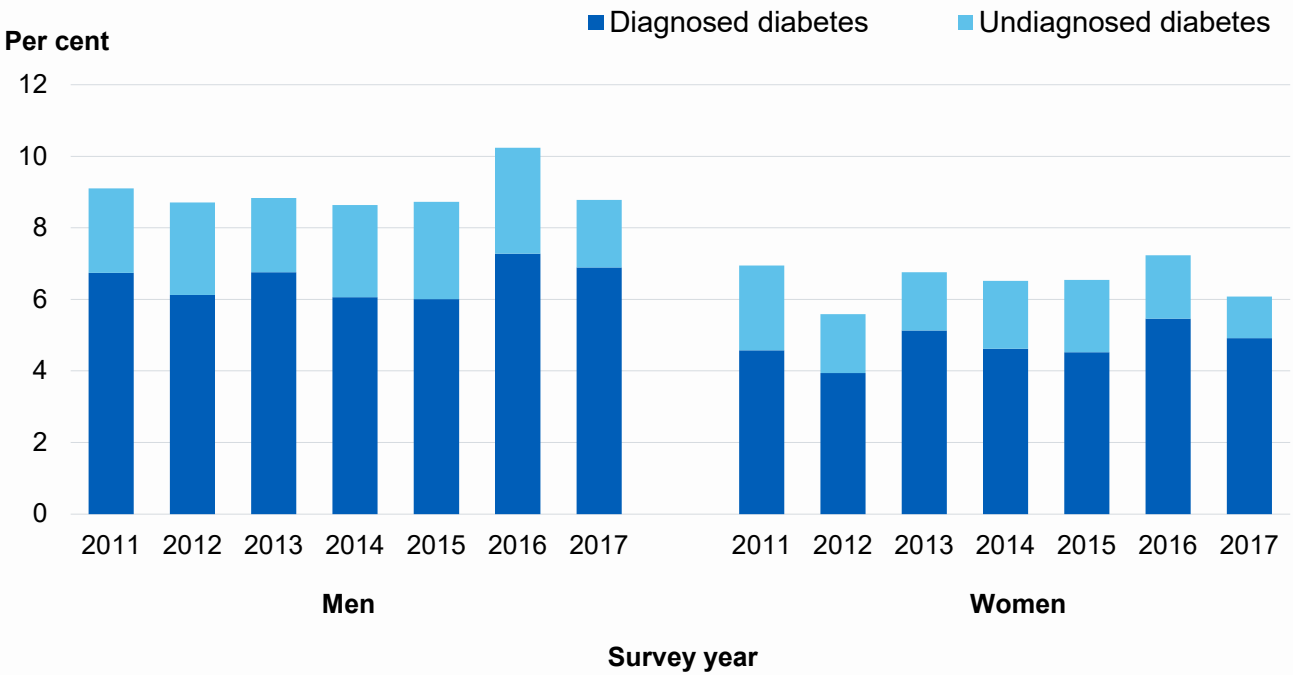
Measures of health

- **Total diabetes** includes adults who reported having doctor-diagnosed diabetes, as well as those with a blood sample measured as having an HbA1c level of 48mmol/mol or above, diagnostic of diabetes
- Among those with total diabetes, participants with a raised HbA1c who did not report having doctor-diagnosed diabetes are defined as having undiagnosed diabetes

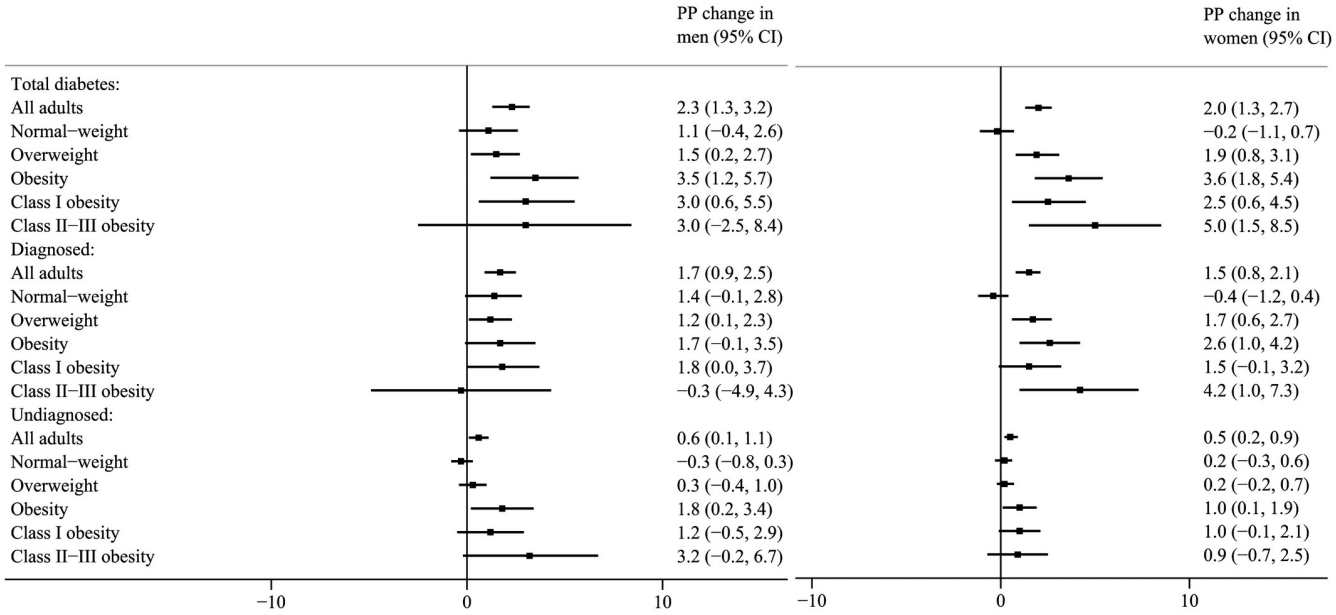
Challenges in harmonization: change in units of reporting for the published trends (% to 2011; mmol/mmol from 2012); change in measuring equipment (4th quarter of 2013)

Proportion of adults with diagnosed and undiagnosed diabetes, by sex, 2011-2017

Base: Aged 16 and over with a nurse visit and valid glycated haemoglobin measurement



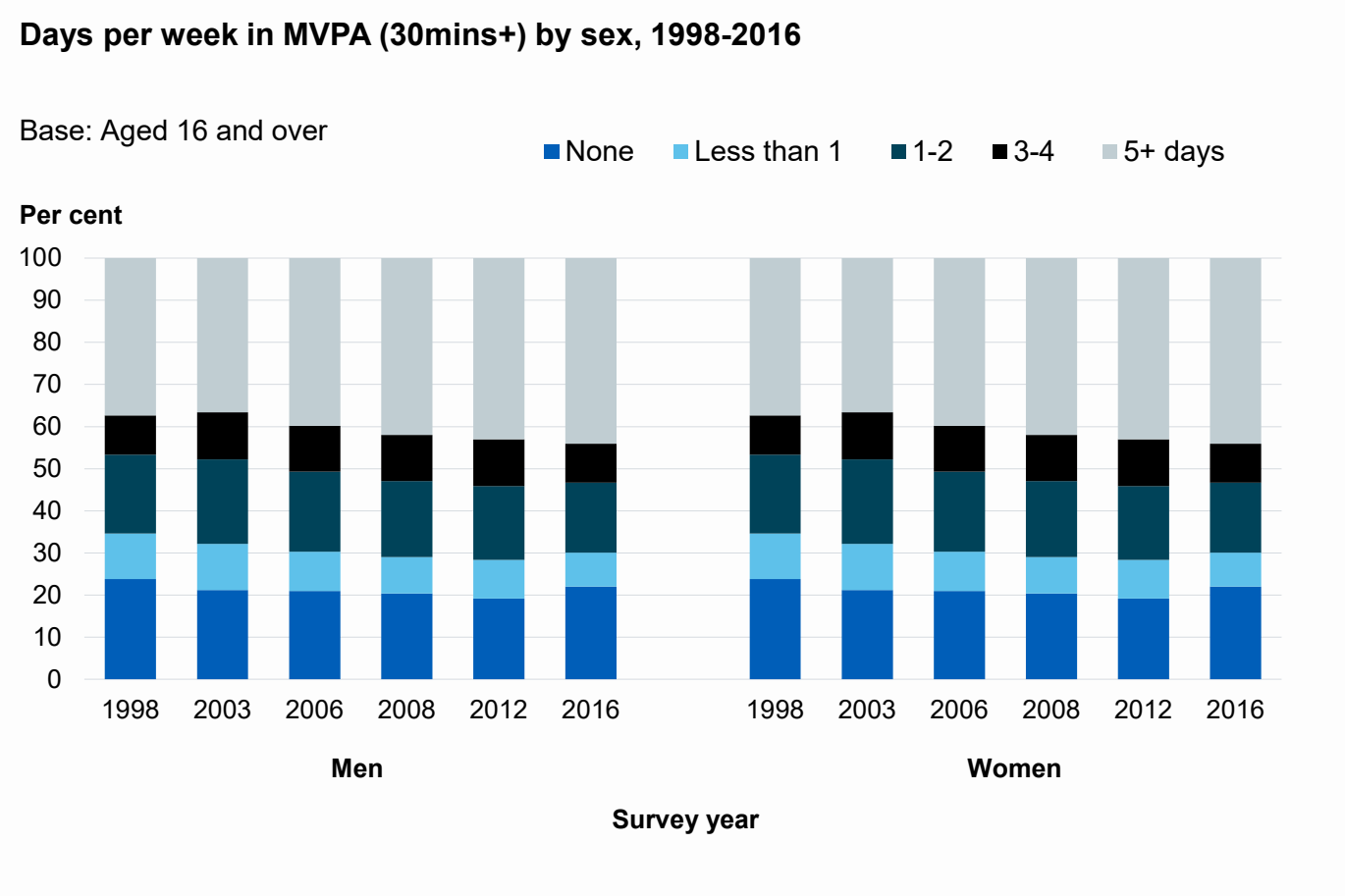
Trends in cardiovascular disease risk factors by BMI category among adults in England, 2003-2018



Obesity, Volume: 29, Issue: 8, Pages: 1347-1362, First published: 21 June 2021, DOI: (10.1002/oby.23184)

Physical activity

- Questions concerned participation during the last four weeks in housework, manual work, gardening and DIY activities, walking, and sports/exercise. For each activity, participants were asked:
  - how many separate days in the past four weeks they did that activity for at least 10 minutes at a time
  - the amount of time they usually spent doing that activity
  - (for sports/exercise and for walking) the intensity of the activity.
  - The questionnaire also asked about physical activity while at work.
- Challenges in harmonization:
  - gaps in survey years;
  - changes in questionnaire (long/short versions) / focus (minimum bout length)



Days/week in MVPA for at least 30 mins. For those participants who reported being 'very or fairly active' at work, arbitrary estimates of 3 or 5 days/week respectively were used, depending on whether the participant worked part- or full-time, to assess levels of physical activity whilst at work (Published: Table 2.5 HSE 2012)

Alcohol consumption

- Challenges in harmonization:
  - Changes in questionnaire / focus
  - Changes in conversion factors
- Dataset includes:
  - Whether drinks alcohol nowadays, occasionally or never (all years)
  - Consumption on the heaviest drinking day in the last week (all years)
  - Average weekly consumption (2011-17)

Complex survey design variables

Year	Survey year
Primary sampling unit	Combination of year and PSU
Stratification	Recommend region
Weighting variables	• Interview weight (wt_int + additional weights in 2006) • Nurse visit (wt_nurse) • Blood sample (wt_blood)

- Non-response weights introduced in 2003
- HSE weights are scaled to the unweighted sample size (average = 1)
- Participants in earlier years assigned a weight of 1

Dissemination

- Manuscript in preparation describing the dataset; will be published as Open Access
- The SPSS code to create the dataset will be made available through the ExPoSE Project website and other public repositories such as the UK Data Service. Users must access HSE datasets via the UKDS
- Dataset will follow HSE conventions (such as negative values for missing data)
- Relevant Stata code for using the dataset (including treatment of outlying values and calculation of CVD risk scores) will also be made available

Acknowledgements

- Staff of the Joint Health Surveys Unit of UCL and NatCen
  - Health and Social Surveys Research Group, Research Dept of Epidemiology & Public Health, UCL
  - NatCen Social Research
- Interviewers and nurses
- Participants in the health surveys
- NHS England – who fund the HSE