
Article:
Stephenson, J., Smith, C.M., Goyder, E.C. orcid.org/0000-0003-3691-1888 et al. (7 more authors) (2021) Cohort profile update : The Yorkshire Health Study. International Journal of Epidemiology, 50 (4). 1070-1070d. ISSN 0300-5771

https://doi.org/10.1093/ije/dyaa272

Cohort Profile Update: The Yorkshire Health Study

Authors

Stephenson J[1*], Smith CM[1], Goyder EC[2], Holding E[2], Haywood A[2], Crum A[2], Blackburn J[1], Cooke J[2], Relton, C[2,3], Bissell P[1]

[1] School of Human and Health Sciences, University of Huddersfield

[2]School of Health and Related Research, University of Sheffield

[3] Centre for Primary Care and Public Health, Queen Mary University of London

*Corresponding author: Dr J Stephenson, School of Human and Health Sciences, University of Huddersfield, Queensgate, Huddersfield, United Kingdom HD1 3DH. Email: j.stephenson@hud.ac.uk

Keywords: research platform, public health, long term conditions, health, lifestyle, Healthcare resource usage

Wordcount:1987, not including references

Key Features

- The Yorkshire Health Study originally known as the South Yorkshire Cohort, is a longitudinal observational study of health and lifestyle acting as a platform for health research and to facilitate the running of cohort multiple randomized controlled trials.
- The first phase involved data collected from 27,813 individuals visiting GP surgeries between 2010 and 2012. A second phase involved data collection from 43,023 individuals between 2013 and 2015 through online recruitment and via NHS Trusts, supported by the NIHR Clinical Research Network to augment existing data with additional data more representative of the underlying population. This resulted in an augmented data set comprising in total 70,836 individuals aged 16-85 years.
- The dataset includes information on demographics, lifestyle and health-related variables including long-term conditions, health service utilisation and medication use.
- For anonymised data and details regarding using the resource for recruiting participants to studies, contact Professor Elizabeth Goyder, University of Sheffield (e.goyder@sheffield.ac.uk).

The original cohort

The Yorkshire Health Study (YHS), originally known as the South Yorkshire Cohort, was introduced in 2014 (1). Its aims were to be a longitudinal observational study of health and lifestyle in Yorkshire and the Humber and to act as a platform for health research, in response to the Healthy Weight, Healthy Lives (2008) government report (2) which had proposed five key aims for future research to tackle obesity levels in England:

- a longitudinal approach to research (incorporating the short, medium and long-term);

- exploring the association of obesity to other key health issues;
- analysing the impact of policies, with a focus on reducing social inequalities;
- evaluating the effectiveness of long-term policies and interventions;
- targeting of population groups and incorporation of 'natural experiments.

Green et al. (1) summarised the sampling strategy for the first wave of the data collection. Data were collected between 2010 and 2012 from patients aged between 16 and 85 years, registered at 43 participating GP surgeries (50% acceptance rate among surgeries). 27,813 were responses received from 156,833 questionnaires mailed (15.9% response rate). Initially data was collected exclusively from boroughs within South Yorkshire (Barnsley, Doncaster, Rotherham and Sheffield), plus a small amount of data from Derbyshire.

A list of current projects can be viewed at: http://www.yorkshirehealthstudy.org/#!your-research-studies/c17r5. The aim of this paper is to introduce an expanded YHS dataset to potential collaborators.

What is the reason for the new data collection?

The scope of data collection was subsequently expanded to cover the Yorkshire & Humberside Government Office Region, including boroughs in North Yorkshire, West Yorkshire, East Riding of Yorkshire and North Lincolnshire. This second wave of data collection from 2013 to 2015 was via online recruitment to increase coverage and sample size, and to generate a sample which was more representative of the underlying population. Information about the study was promulgated through a regional media campaign that utilised leaflets, posters and a variety of other methods such as the YHS Health Calculator (https://www.yorkshirehealthstudy.org/about1-cyg9), an engagement tool that compares individual data with the health of people of the same age and sex in Yorkshire; Facebook adverts to target underrepresented groups within the cohort and campaigns with various local organisations, such as local councils; Sheffield International Venues, and NHS organisations

with emails to their staff. All residents in the Yorkshire and the Humber region were invited to participate via a questionnaire and online website (http://www.yorkshirehealthstudy.org/).

What will be the new areas of research?

The Academy of Medical Sciences report 'Improving the Health of the Public by 2040' (3) postulated that the complexity of interlinked factors influencing health lie beyond the scope of the current paradigm of biomedical research. We must refocus our research efforts on prevention and early interventions at scale and involve disciplines that would not usually be considered to be within the public health field; an approach integrating aspects of natural, social and health sciences, alongside the arts and humanities, which directly or indirectly influence the health of the public. The YHS is ideally placed to underpin this 'health of the public' approach. Initial studies planned link obesity with quality of life, and the impact of non-genomic factors on the development of Type 2 diabetes, cardiovascular disease and mental illness in black and ethnic minority populations. For certain studies a Trials within Cohorts (TWiCs) methodology is proposed which offers a potential solution to the challenges of recruitment to time and target in health research (4, 5).

Who is in the cohort?

Table 1 presents the baseline characteristics of participants in both waves of the cohort by key factors recorded in both waves of the data collection.

TABLE 1 HERE

Participants in the cohort are older than in the total Yorkshire and the Humber region; and the proportion of females in the sample is also somewhat higher than the population proportion.

Some differences in categorical proportions of demographic factors were observed across the two phases of the study. The second phase included a substantially higher number of female

respondents (69.0%) than the first phase (56.4%). Respondents in the second phase were younger: in the first wave, 30.0% of respondents were in the oldest two age groups (i.e. were aged over 65 years), while 16.8% were in the youngest two groups (i.e. were aged under 35 years or under). In the second wave, 17.2% of respondents were in the oldest two age groups, while 30.3% were in the youngest two groups. A higher proportion of respondents were employed in Wave 2 (58.2%) than in Wave 1 (46.8%). Take-up of health service use was higher in Wave 2 (64.4% of respondents using one or more health services in the last 3 months) than in Wave 1 (49.6% of respondents using one or more health services in the last 3 months). Proportions of respondents with one or more longstanding conditions were similar in the two waves (61.7% in Wave 1; 60.3% in Wave 2).

In order to further facilitate the expansion of the study across the region the study was adopted on the National Clinical Research Network (CRN) portfolio of studies in 2016. This meant that the study was eligible for support from the CRN Y&H for project set up, recruitment and delivery of the study within NHS sites across the region. Research teams within the participating sites printed paper copies of the YHS questionnaire and distributed these to staff and patients through various recruitment and advertising methods (including emails, adverts, pop up events). GCP trained staff would then input the paper questionnaires into the PROSPECT online data capture system.

A total of 43 023 valid responses were received during this second phase of data collection; thus 70 836 responses in total from the two waves of data collection. This represents 1.34% of the total population of 5 283 733 of the Yorkshire and the Humberside Government Office Region, as measured in the 2011 Census.

This cohort has been followed up once, but baseline data only is presented here.

What has been measured?

The key variables collected in the two waves of data are summarised in Table 2. The exact formulation of some questions (for example: those eliciting information on alcohol intake and weight loss methods utilised) varied between the two waves of data collection.

TABLE 2 HERE

58 766 participants (83% of all participants) consented for their health records to be accessed; comprising 25 662 participants from Wave 1 (92%) and 33 104 participants from Wave 2 (77%). 64 902 participants (91.7% of all participants) consented to future contact; comprising 26 375 participants from Wave 1 (95%) and 38 527 participants from Wave 2 (90%).

Green et al. (1) list specific benefits of the inclusion of key variables; observing that:

- The EuroQoL EQ-5D health-related quality of life (HRQoL) questionnaire (6) included is used within the NHS and is the National Institute for Health and Care Excellence's preferred measure of HRQoL in economic evaluations.
- Postcode information facilitates collection of additional spatial information.
- Inclusion of long-standing conditions reflects the growing interest in the impact of multimorbidity (7, 8).
- GP surgery information can also be linked to participants, providing information relating to the characteristics of the primary health care provider used by each individual.
- Data linkage of individual records in the cohort to GP clinical records (e.g. morbidity, health care and medication usage) and other sources (e.g. Hospital Episode Statistics) is possible.

What has it found? Key findings and publications

The data has been used to evaluate quality of life (QoL) in obese (defined to be 30 $kg/m^2 \leq BMI < 40$ kg/m^2) and morbidly obese individuals (defined to be BMI $\geq$ 40 kg/m2) and compared against those

in non-obese sectors. A study currently under review related BMI and QoL data, adjusting for demographic- and health-related factors, from 64 000 respondents of two waves of the YHS. The study found substantively lower QoL scores (as measured by the EQ-5D summary index) in both BMI groups compared to the non-obese group; with reductions of QoL of 4.5 percentage points (95% confidence interval 4.1 to 4.9) in obese individuals and 12.2 percentage points (95% confidence interval 11.3 to 13.2) in morbidly obese individuals. Both effects were significant at the 5% significant level ($p<0.001$ in both cases). However, raised BMI was a less important predictor of QoL than other factors, notably living with long-term conditions: the presence of one or more of these conditions was associated with a reduction of 15.2 percentage points (95% confidence interval 14.9 to 15.5; $p<0.001$). These results are consistent with findings of Wu et al. (9) and Kearns et al. (10), who also found long-term conditions and raised BMI respectively to have substantive negative effects on quality of life.

What are the main strengths and weaknesses?

Strengths include: a large sample size covering a wide range of population sub-groups; longitudinal design for data collection; data collected on a broad range of personal, lifestyle and health-related variables; a 'research platform' designed to facilitate the running of 'cohort multiple randomised controlled trials' and facilitate participant identification and recruitment to health-related research; consent to link the majority of individuals to their NHS health records; and questions designed to be policy relevant, to encourage the collaboration between academia, local authorities and the NHS.

Weaknesses include: some demographic bias with a higher proportion of elderly, female, White ethnicities and individuals from less deprived areas; a low response rate; some areas of the Yorkshire and the Humber region disproportionately represented; an extended period of data collection and variation in questionnaire design leading to inconsistencies in information elicited in the two waves of data collection, with some variables collected in one wave only.

Can I get hold of the data? Where can I find out more?

Anonymised data and details regarding using the resource for recruiting participants to studies can be gathered by contacting Professor Elizabeth Goyder (e.goyder@sheffield.ac.uk). Multi-disciplinary collaboration is strongly encouraged.

Funding This work was supported by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) South Yorkshire and the University of Sheffield.

Ethics Ethical approval for the cohort was obtained from the Leeds East NHS Research Ethics Committee (ref: 09/H1306/97).

Acknowledgements We are grateful to all the individuals who have enrolled in the cohort. We also acknowledge the GP practice staff for their contribution in the recruiting process. We would also like to acknowledge the participation and resources of the members of the ARC YH ACORN group and our collaborating organisations: Barnsley Hospital NHS Foundation Trust (NFT), Bradford District Care NFT, Leeds York Partnership NFT, Sheffield Health and Social Care NFT, Sheffield Teaching Hospitals NFT, Mid Yorkshire Hospitals NFT, Bradford Teaching Hospitals NFT, Leeds Teaching Hospitals NFT, North Lincolnshire and Goole NFT, Yorkshire Ambulance Service Trust, Airedale NFT, Harrogate & District NFT, Humber NFT, Rotherham NFT.

This report is independent research funded by the National Institute for Health Research Applied Research Collaboration Yorkshire and Humber. The views expressed in this publication are those of the authors and not necessarily those of the Yorkshire Health Study Management Team or Steering Committee, National Institute for Health Research or the Department of Health and Social Care.

Conflict of interest: None declared.

References

1. Green MA, Li J, Relton C et al. Cohort profile: the Yorkshire health study. International journal of epidemiology. 2016 Jun 1;45(3):707-12.

2. Department of Health. Healthy Weight, Healthy Lives: A Cross Government Research and Surveillance Plan. London: DoH, 2008.

3. Academy of Medical Sciences. (2016). Improving the health of the public by 2040: Optimising the research environment for a healthier, fairer future. https://acmedsci.ac.uk/file-download/41399-5807581429f81.pdf (last accessed 3.2.20)

4. Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. BMJ 2010;340:c1066.

5. Clegg A, Relton C, Young J, Witham M. Improving recruitment of older people to clinical trials: use of the cohort multiple randomised controlled trial design. Age and ageing 2015;44(4):547-50.

6. Rabin R, Charro FD. EQ-SD: a measure of health status from the EuroQol Group. Annals of medicine 2001;33(5):337-43.

7. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet. 2012 Jul 7;380(9836):37-43.

8. Mangin D, Heath I, Jamoulle M. Beyond diagnosis: rising to the multimorbidity challenge. BMJ 2012;344:e3526.

9. Wu M, Brazier JE, Kearns B, Relton C, Smith C, Cooper CL. Examining the impact of 11 long-standing health conditions on health-related quality of life using the EQ-5D in a general population sample. The European Journal of Health Economics 2015;16(2):141-51.

10. Kearns B, Ara R, Young T, Relton C. Association between body mass index and health-related quality of life, and the impact of self-reported long-term conditions–cross-sectional study from the south Yorkshire cohort dataset. BMC Public Health 2013;13(1):1009.

Title: Cohort Profile Update: The Yorkshire Health Study

Authors: Stephenson J[1], Smith CM[1], Goyder EC[2], Holding E[2], Haywood A[2], Crum A[2], Blackburn J[1], Cooke J[2], Relton, C[2,3], Bissell P[1]

Cite this as: The full version of this profile is available at *IJE* online and should be used when citing this profile.

Corresponding author: Dr J Stephenson, School of Human and Health Sciences, University of Huddersfield, Queensgate, Huddersfield, United Kingdom HD1 3DH. Email: j.stephenson@hud.ac.uk

Keywords: research platform, public health, long term conditions, health, lifestyle, Healthcare resource usage

The original cohort: The Yorkshire Health Study (YHS) was set up to be a longitudinal observational study of health and lifestyle in Yorkshire and Humber and to act as a platform for health research.

The first wave of baseline data was collected through GP surgeries between 2010 and 2012 (n = 27,813).

The new focus: A second wave of baseline data collection was undertaken between 2013 and 2015 (n = 43,023), through online recruitment and via NHS Trusts, supported by the NIHR Clinical Research Network, bringing the total number of participants to 70,836 individuals aged 16-85 years. The majority of participants gave consent for recontact and for data linkage. This additional data collection sought to increase the geographical coverage and sample size, and to generate a sample which was more representative of the underlying population.

Who is left: There are currently no plans to collect follow up data. The entirety of the existing dataset may be considered to represent the participants remaining in the study.

New measures: Personal health and lifestyle questions relating to family membership and employment status, household income and dietary factors

Unique features: Data was collected on a broad range of personal, lifestyle and health-related variables designed to be policy relevant, to encourage the collaboration between academia, local authorities and the NHS. The YHS is a 'research platform' designed to facilitate the running of 'cohort multiple randomised controlled trials' and facilitate participant identification and recruitment to health-related research. The majority of participants gave permission for future contact and for data linkage.

Reasons to be cautious: There is some demographic bias in this dataset, with a higher proportion of elderly, female, white ethnicities and individuals from less deprived areas.

Collaboration and data access: Anonymised data and details regarding using the resource for recruiting participants to studies can be gathered by contacting Professor Elizabeth Goyder (e.goyder@sheffield.ac.uk). Multi-disciplinary collaboration is strongly encouraged.

Funding and competing interests: This work was supported by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) for South Yorkshire (Obesity Theme) and the University of Sheffield. There are no competing interests.

Author affiliations: [1] School of Human and Health Sciences, University of Huddersfield; [2] School of Health and Related Research, University of Sheffield; [3] Centre for Primary Care and Public Health, Queen Mary University of London

Table 1: Characteristics of the participants of second-wave data collection from 2013 to 2015

Variable	Frequency (valid %)
Age (*n*=69008)	
16-25 years	7552 (10.9%)
26-35 years	10 139 (14.7%)
36-45 years	10 307 (14.9%)
46-55 years	12 737 (18.5%)
56-65 years	12 505 (18.1%)
66-75 years	10 434 (15.1%)
76+ years	5334 (7.7%)
Sex (*n*=70828)	
Male	25 826 (36.5%)
Female	45 002 (63.5%)
BMI category (*n*=64673)	
Underweight (BMI below 18 kg/m^2)	535 (0.8%)
Normal weight (18 kg/m^2 ≤ BMI < 25 kg/m^2)	27 755 (42.9%)
Pre-obesity (25 kg/m^2 ≤ BMI < 30 kg/m^2)	21 994 (34.0%)
Obesity Class I & II (30 kg/m^2 ≤ BMI < 40 kg/m^2)	12 676 (19.6%)
Obesity Class III (40 kg/m^2 ≤ BMI < 50 kg/m^2	1713 (2.7%)
BMI ≥ 50 kg/m^2	235 (0.4%)
Employment status (*n*=70836)	
Not in employment	32 805 (46.1%)
In employment	38 031 (53.9%)
Longstanding health conditions (*n*=68073)	
No conditions	26 636 (39.1%)
One or more conditions (any)	41 437 (60.9%)
Conditions reported (*n*=41347):	

Tiredness/fatigue	12 730 (18.7%)
Pain	14 472 (21.3%)
Insomnia	5137 (7.5%)
Anxiety/nervousness	8957 (13.2%)
Depression	7323 (10.8%)
Memory problems	2608 (3.8%)
Diabetes	5569 (8.2%)
Breathing problems	7820 (11.5%)
Hypertension	11 394 (16.7%)
Heart disease	3599 (5.3%)
Osteoarthritis	5967 (8.8%)
Stroke	1209 (1.8%)
Cancer	2496 (3.7%)
Other	14 008 (20.6%)
Take-up of health services in last 3 months (*n*=70836)	
No take-up of any service	29 341 (41.4%)
Take-up of 1 or more services	41 495 (58.6%)

[1] More than one condition could be reported

Table 2: Variables collected during second-wave data collection from 2013 to 2015

Group	Variable	Included in Wave 1	Included in Wave 2
Personal	Age	✓	✓
	Sex	✓	✓
	Ethnicity	✓	✓
	Height[1]	✓	✓
	Weight[1]	✓	✓
	Education level	✓	
Health	Longstanding conditions (LSCs)[4]	✓	✓
	Level of trouble due to LSCs		✓
	Health-related quality of life[3]	✓	✓
	Medication use[5]	✓	✓
	Use of health services	✓	✓
	Days lost to illness	✓	✓
	Alcohol intake	✓	✓
	Smoking habits	✓	✓
Lifestyle	Number in household		✓
	Living with partner		✓
	Number of children	✓	✓
	Living with carer		✓
	Employment status	✓	
	Type of employment		✓

	Income		✓
	Alcohol consumption	✓	✓
	Dietary habits		✓
	Use of weight loss methods	✓	
	First aid experience		✓

[1]Used to derive body mass index (BMI)

[2]Used to derive socio-economic deprivation level

[3]Using the EQ-5D instrument

[4]Including: Tiredness/fatigue, Pain, Insomnia, Anxiety/nervousness, Depression, Diabetes, Breathing problems, Hypertension, Heart disease, Osteoarthritis, Stroke, Cancer, Memory problems (wave 2 only)

[5]Including: Hospital, General practitioner, Social/welfare, Mental health, Alternative therapies, Dentist, (wave 2 only) Podiatrist (wave 2 only), Optician (Wave 2 only)