

Howcutt, S., Barnett, AL., Barbosa Boucas, S. and Smith, L. (2017) 'Patterns of response by sociodemographic characteristics and recruitment methods for women in UK population surveys and cohort studies', *Women and Health*

DOI: <https://doi.org/10.1080/03630242.2017.1310170>

This document is the authors' Accepted Manuscript.

License: <https://creativecommons.org/licenses/by-nc-nd/4.0>

Available from RADAR: <https://radar.brookes.ac.uk/radar/items/736ecdbb-93af-42d0-a3f8-049cf7ce98a8/1/>

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners unless otherwise waved in a license stated or linked to above. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

Manuscript title	<b>Patterns of response by sociodemographic characteristics and recruitment methods for women in UK population surveys and cohort studies.</b>
Running Title	Representativeness of women in UK health studies
Author's full names, affiliations and addresses	<p>Sarah J. Howcutt (Oxford Brookes University, Faculty of Health and Life Science, Jack Straw's Lane, Marston, OXON. OX3 0FL)</p> <p>Anna L. Barnett (Oxford Brookes University, Faculty of Health and Life Science, Gipsy Lane, Headington, OXON. OX3 0BP)</p> <p>Sofia Barbosa Boucas (Bucks New University, Queen Alexandra Road, High Wycombe, BUCKS. HP11 2JZ)</p> <p>Lesley A. Smith (Oxford Brookes University, Faculty of Health and Life Science, Jack Straw's Lane, Marston, OXON. OX3 0FL)</p>
Corresponding author's name, telephone number and e-mail address	
Acknowledgements	There is no financial grant or award associated with this work.
Keywords	Women's health; Representativeness; Recruitment; Cohort studies; Population surveys
Figures	Six Figures are submitted. They have been produced in colour format (1200dpi) for online publication only. Black and white versions can be provided for print.
Publication Statement	This work has not been published elsewhere nor submitted simultaneously for publication elsewhere.

## **Abstract**

Women are an important public health focus, because they are more likely to experience some social determinants of disease, and they influence family health. Little research has explored the sociodemographic representativeness of women in research studies. We examined the representativeness of female respondents across four sociodemographic factors in UK population surveys and cohort studies. Six UK population-based health surveys (from 2009 to 2013) and eight Medical Research Council cohort studies (from 1991 to 2014) were included. Percentages of women respondents by age, income/occupation, education status and ethnicity were compared against contemporary population estimates. Women aged <35 years were under-represented. The oldest women were under-represented in four of nine studies. Within income/occupation, at the highest deprivation level, the range was 4% under-representation to 43% over-representation; at the lowest level, it was 6% under-representation to 21% over-representation. Of nine studies reporting educational level, four under-represented women without school qualifications, and three under-represented women with degrees. One of five studies over-represented non-white groups and under-represented white women (by 9%). Response patterns varied by topic and recruitment and data collection methods. Future research should focus upon the methods used to identify, reach and engage women to improve representativeness in studies addressing health behaviors.

**Keywords** Women's health; Representativeness; Recruitment; Cohort studies; Population surveys

## **Introduction**

A focus of epidemiologic research is to describe patterns of health-related events (Centers for Disease Control and Prevention 2012). Cross-sectional surveys are frequently used to estimate the prevalence of health risks and events at particular time points (Coggon, Rose, and Barker 2016). Data can be collected relatively quickly and compared over time, but surveys rely on samples that represent the populations to which results are extrapolated.

The efficacy of survey recruitment is typically assessed by response rates (Mindell, Aresu, et al. 2012). These have declined in the UK. For example, the response rate of the Scottish Health Survey (SHeS) has fallen from 84% in 1995 (Gorman et al. 2014) to 56 % in 2013 (Corbett et al. 2014). No consensus exists on acceptable response rates, and overall response rates cannot quantify the extent to which samples represent the sampled population. Also, the danger of ecological fallacy exists whereby estimates, for the aggregated data, do not adequately reflect the risks and events of individuals due to under- and over-representation of different sectors of the source population.

Epidemiologic studies also aim to investigate determinants of disease. Cohort studies are useful for this because they allow incidence and etiology to be explored, although they do not necessarily aim for a representative sample. Rothman Gallacher and Hatch (2013) argue that representativeness is irrelevant for establishing causal explanations. Yet in health research, potential explanatory factors may be tied to sociodemographic attributes (e.g., access to information). The external validity of a study is, therefore, threatened if respondents differ, in terms of these determinants, from the population to whom conclusions will be generalized.

Approaches exist to limit non-response bias. These either try to increase response rates or use post-hoc statistical adjustments. Both strategies are problematic. Fieldwork strategies to

increase response rates, such as increased call attempts, are costly and may increase questionnaire item-nonresponse (Olson 2013). Another approach is to target under-represented groups through ‘boost’ samples, i.e., oversampling of certain communities. Yet, boost samples pose difficulties in identifying sampling frames and may introduce selection bias (Platt 2013).

Statistical adjustment is often used to standardize results to population norms, for example, weighting data to alter the influence of responses (Brick 2013). However, statistical adjustments are only useful for larger data samples with sufficient data to inform assumptions about missing information. They also rely on the flawed supposition that respondents will be similar to non-respondents. In studies investigating alcohol use, heavier drinkers are less likely to participate, so that adjustments result in underestimation of actual consumption (Catto 2008). A further concern is that statistical adjustments effectively increase the sample size, making summary statistics appear more accurate than they should.

Considerable research has already identified sociodemographic characteristics linked with lower response, suggesting that non-respondents are more likely to be male, young or elderly (if infirm), in lower income occupations, educated to lower levels and unmarried (Uhrig 2008). These findings present a challenge to representativeness and consequently to external validity. While concerns about the external validity are relevant to both genders, this paper focuses upon women. Considerable focus in the UK has been upon the methods to understand social determinants of health to reduce health inequalities (Scottish Public Health Observatory 2015). Gender is an important factor. Data suggest that women experience a greater number of years with ill-health and disability, despite longer life expectancy (Office for National Statistics 2015). Women are also marginally more vulnerable to deprivation, an important predictor of ill-health (Office for National Statistics UK 2015a). Moreover, women’s health risk behaviors are particularly influential over the later health behaviors of

their children (Griffith et al. 2007). Therefore, women are an important focus for harm prevention, yet little research has explored the extent to which study samples represent the population from which they are drawn.

The aim of these analyses was to describe patterns of response for women by sociodemographic attributes across a sample of health-related population-based surveys and cohort studies in the UK, to compare these against sociodemographic profiles of respective sampled populations.

## **Methods**

### **Study selection**

General population surveys were identified via the UK Data Service and included if they covered one or more UK nations and if their focus was health and/or health risk behaviors (e.g., tobacco use). Data for the last five years available at the time of writing were explored (2009-2013). Medical Research Council (MRC) cohort studies were identified from the *Cohort Directory* (Medical Research Council 2015) and included if they recruited women aged 16 to 65 years from non-treatment seeking populations. Data collection period was not an exclusion criterion, and data collection periods of the reviewed studies spanned 1991-2014. We selected MRC studies because they were considered to be of high methodological quality. Cohort studies concerned solely with individuals with a shared medical diagnosis or workplace were excluded because they were not recruiting a general population sample. This review focuses upon adult women's health, not ageing, and therefore studies which only recruited women >50 years were also excluded, as were birth cohort studies when their focus was solely on infant/child health.

### **Data sources**

For the population surveys, data for each sociodemographic factor were taken from unweighted datasets held by the *UK Data Service* (Department of Health, Social Services and Public Safety 2013; Field et al. 1995; Johnson et al. 2005; Johnson and Centre for Sexual and Reproductive Health Research 2015; National Centre for Social Research 2011; 2013; 2014; 2014a; 2015; National Centre for Social Research et al. 2005; 2013; 2014; 2015; 2015a; 2015b; School Scottish Centre for Social Research 2009; 2010; 2011; 2012; 2013), with the exception of the General Lifestyle Survey. Data for this survey series were taken from the survey's Technical Appendices (Dunstan 2012; Office for National Statistics 2011; 2011a). Annual population survey data were extracted for the years 2009-2013. One population survey - the National Survey for Sexual Attitudes and Lifestyles (Natsal) is conducted every 10 years, and data were therefore taken from the 1990-1991, 1999-2001 and 2010-2012 runs. Data for cohort studies came from published articles (Cade et al. 2015; Davey et al. 2003; Fone et al. 2013; Fraser et al. 2013; Hayat et al. 2014; Inskip et al. 2006; Robinson et al. 2004; Swerdlow et al. 2011; Wilkinson, Inskip, and Southampton Women's Survey Study Group 2006), with the exception of *Understanding Society* for which it was possible to obtain data from the technical summary (Buck, Laurie, and Nolan 2011) and the *UK Data Service* (University of Essex, 2015).

### **Sociodemographic attributes**

Age, socioeconomic status, defined by measures of occupation/income and education and ethnicity, were the selected sociodemographic attributes. Marital status was also explored, but this was not included due to the large variation in categories used by each study which limited meaningful comparison.

Age was summarized in 10-year age groups taken from the original data sources or created by collapsing data for single years of age for which raw data were available. Age data from cohort studies were taken as reported in published articles.

Proxy measures of socioeconomic status (SES) were selected which reflected economic status. Most often this was the National Statistics Socioeconomic Classification 8 (NS-SeC 8) to provide a good level of detail for subgroups (Office for National Statistics 2010).

Education was classified by level of the highest qualification earned to facilitate comparisons between systems and over time. English and Scottish levels were compared using the Scottish Credit and Qualifications Framework Partnership (2015). The National Careers Service (2015) and Health Education East Midlands (2013) allowed changes in English Qualifications over time to be reconciled.

Ethnicity has only recently been routinely collected in UK surveys, and a wide range of classification systems exist due to changes in the political and social landscape (Mathur, Grundy, and Smeeth 2013). To facilitate comparison between different classification systems used, ethnicity was simplified to four categories: 'white', 'black' (black African and black Caribbean), Asian (Indian, Bangladeshi, Pakistani and Chinese) and 'Other' (mixed race and other ethnicities).

### **Summary statistics**

The unit of analysis was the individual to facilitate comparison across studies because not all studies had household data. If full datasets were available, the denominator in percentage calculations was the percentage of the total number of female respondents classified within each category.

Population estimates were primarily taken from census data. These data were obtained from the UK Data Service. The InFuse portal (Office for National Statistics 2011c) was used to



locate data within the 2001 and 2011 censuses for England and Wales and the *Casweb* portal (Office for National Statistics 1991) for the 1991 iteration. The *Census for Scotland* was accessed via, *the Census Data Explorer* (National Records for Scotland 2015)

Estimates were as specific as possible to region. To compare region-specific cohort studies, we selected county-level data. These were taken either from the peer-reviewed publications or were derived from mid-year population estimates, using the Stats Wales tool (2015) or Office for National Statistics reference tables (Office for National Statistics 2015b).

Contemporaneous data were used for population estimates for comparison against the population surveys and cohort studies. In comparisons between population estimates and samples, we used data from the nearest available year. We were only able to explore response rates across multiple years for age due to insufficient data for other sociodemographic attributes. Aggregate census data was the main source of population estimates for SES, education and ethnicity, and therefore the year 2011 was selected for these attributes to allow comparison against the most recent census datasets.

To assess representativeness, we calculated the difference between the percentage of each sociodemographic attribute in each sample and in the population. Negative values denoted under-representation and positive values over-representation.

## **Results**

Six population surveys were included (Department of Health, Social Services and Public Safety 2013; Dunstan 2012; Field et al. 1995; Johnson et al. 2005; Johnson and Centre for Sexual and Reproductive Health Research 2015; National Centre for Social Research 2009; 2010; 2011; 2012; 2013; 2013a; 2014; 2015; 2015a; 2015b; Office for National Statistics 2011; Office of National Statistics 2011; Scottish Centre for Social Research, University

College London. Department of Epidemiology and Public Health 2015; 2015a; 2016, 2016a; 2016b) (Table 1).

Each population survey was repeated annually, except the National Survey of Sexual Attitudes and Lifestyles (Natsal), which is conducted every 10 years. All involved adults aged 16 years and over; however, the Natsal survey stopped at age 44 years for the first two iterations and at age 74 years for the third. The Scottish Health Survey (SHeS) had the smallest sample, and the General Lifestyle Survey (GLS) has the largest. The Welsh Health Survey (WHS) was the newest, having been first conducted in 2003. [\[Place Table 1 here.\]](#)

Eight cohort studies were included (Buck et al. 2011; Cade et al. 2015; Davey et al. 2003; Fone et al. 2013; Fraser et al. 2013; Hayat et al. 2014; Inskip et al. 2006; Robinson et al. 2004; Swerdlow et al. 2011; Wilkinson et al. 2006; University of Essex 2015) (Table 2). Data were available for the period between 1993 and 2009. Three cohorts were on a national scale, and five were regional. The Avon Longitudinal Study of Parents and Children (ALSPAC) and the Southampton Women's Study (SWS) focused on family health; the *Breakthrough Generations*, two European Prospective Investigations into Cancer (EPIC-Oxford and EPIC-Norfolk) and the UK Women's Cohort Study (UKWCS) projects focused upon cancer prevention. The remaining studies explored links between sociodemographics and health. Only limited data covering the entire cohort were available to the authors for the ALSPAC, and therefore these are only reported in the text. [\[Place Table 2 here.\]](#)

Two cohort studies were excluded. These were the UK Biobank, because this cohort profile was only available as an abstract (Hutchings et al. 2014), and the Southall and Brent Revisited study had a very small (n=1,198) target sample of women (Tillin et al. 2012).

### **Recruitment and fieldwork techniques**

The national population surveys used stratified samples based upon random selection of addresses from postcode sectors (Corbett et al. 2014; Erens et al. 2014; Mindell et al. 2012; Office of National Statistics 2011; Sadler et al. 2013). All surveys used face-to-face interviews, recorded using computerized assisted personal interviewing. In the Natsal and Northern Ireland Health Survey (NIHS), participants answered sensitive questions in writing to reduce embarrassment. In Natsal III a boost sample was used to increase the sensitivity of analyses in 16-34 year-olds (Erens et al. 2014).

More variation in study recruitment was observed for the cohort studies. The *Breakthrough Generations* study (Swerdlow et al. 2011) needed a committed group due to the study's requirements. They initially advertised through national cancer charities and then encouraged chain or snowball referrals from existing members to strengthen social ties between participants. The E-CATALyST study (Fone et al. 2013) employed postal contacts. Visits were organized to follow up missing questionnaires and verify property occupancy.

Both *EPIC* studies used postal recruitment using lists from family doctors (Davey et al. 2003; Hayat et al. 2014) but the *EPIC* Oxford also asked vegetarian and vegan societies to advertise the study to reach non-meat eaters. In both *EPIC* studies, family doctors and research nurses took biological samples.

The SWS sent leaflets and letters to eligible women (identified via family doctor registers) (Inskip et al. 2006). This was followed by a telephone call, a letter and visit to update contact details. To reach women who were not registered with a local family doctor, women were also approached in supermarkets and at local events.

The UKWCS, like *EPIC* Oxford, also wanted to recruit non-meat eaters (Cade, Burley, and Greenwood 2004). This study, however, contacted women via the World Cancer Research Fund using mailing lists targeted at women, including a number of charity databases. All

vegetarians were selected, and a systematic sampling procedure was devised to create a meat-eating comparison group.

*Understanding Society* used, like the population surveys, stratified sampling based upon postal code sectors to divide the UK into sampling units. A boost sample and inclusion of the British Household Panel Survey respondents allowed this study to develop an extremely large sample of minority groups. The sample was clustered in most of the UK with the exception of Northern Ireland (Buck et al. 2011). *Understanding Society* employed a variety of data collection techniques, including the collection of audio and video qualitative data plus record linkage.

### ***Patterns of Response by Sociodemographic Attributes***

#### **Age**

A consistent pattern was observed across surveys and over time of under-representation of lower age groups and over-representation of older age groups (Figure 1). This ranged from 10% under-representation in 16-24 year olds to 3% over-representation in > 44 year olds.

[\[Place Figure 1 here.\]](#)

Differences between year of data collection in the population surveys (Figure 1), comparing the same age group in the same survey, were all within 2%, apart from the SHeS in which a 3% reduction in under-representation occurred between 2012 and 2013 for the 25 to 34-year age group. The Natsal differed from the other population surveys in that the first two iterations demonstrated the same pattern of under-representation at younger ages noted above (Figure 2). However, the results in the Natsal III were reversed. Individuals younger than 35 years were over-represented, and older groups were under-represented. [\[Place Figure 2 here.\]](#)

In the cohort studies (Figure 3), with the exception of *Understanding Society*, the youngest age group was also under-represented. In two of the four studies, the oldest age group was under-represented. The *EPIC Norfolk* study recruited participants aged at least 40 years with the youngest and oldest age groups under-represented. [\[Place Figure 3 here.\]](#)

### **Socioeconomic Status (Occupation and Income)**

At the highest deprivation categories (according to the different measures used in the original datasets), three of four population surveys and one of three cohort studies with data showed under-representation of 2% to 15% (Figure 4). At the lowest level of deprivation, one of four population surveys and one of three cohort studies with data showed under-representation of 1% to 6%. The pattern for deprivation was less clear than for age, but the percentage differences between samples and population estimates were much larger in some cases. Notably *Understanding Society* over-represented women with high deprivation by 43% and *Breakthrough Generations* over-represented those with lower deprivation by 21%.

In addition, the ALSPAC study (not reported in Figure 4 due to the smaller number of categories) had a sample which over-represented women owning cars (7% higher levels than Avon estimates) and homeownership (10% higher than Avon estimates) (Fraser et al. 2013). Data were only available for genders combined in the *E-CATALyST* and population estimates from 2011 Census (England, Wales and Scotland). [\[Place Figure 4 here.\]](#)

### **Socioeconomic Status (Education)**

Two of four population surveys and two of five cohort studies demonstrated under-representation at the lowest level of education; two of four population studies and one of five cohort studies showed under-representation at the highest qualification level (Figure 5).

The SWS had the greatest under-representation at the lowest educational level (16%) (Wilkinson et al. 2006). While the WHS had the highest under-representation of women with

level 4 (university) education (10%), EPIC Oxford had the highest over-representation at this qualification level (28%). *EPIC* Norfolk was the most successful in obtaining representation of those with no qualifications (22%), and yet it was the study that did not achieve representative samples at level 4 and over. [\[Place Figure 5 here.\]](#)

### **Ethnicity**

Few included studies reported response rates by ethnicity. Individuals who classed themselves as white were over-represented by 1% in the HSE and 2% in the Natsal III (Figure 6). Asians were under-represented by 1% and 2% in the SHeS and Natsal III, respectively. *Understanding Society* employed a boost sample. ‘White’ women were under-represented by 9%. In the *ALSPAC* cohort, non-white mothers were under-represented by 2% compared to 1991 Census figures for Avon. No under-representation was observed in the ‘Black’ or ‘Other’ groups in the included studies. [\[Place Figure 6 here.\]](#)

### **Discussion**

Overall a clear and consistent under-representation was observed of women at younger ages. Patterns of under-representation, particularly within SES attributes, were more varied. The differences between percentages of women within samples and populations were large in some cohort studies, particularly if income/occupation was the attribute under consideration.

We found under-representation of women under the age of 35 years, notably in the population surveys. Lower response by young people has been attributed to availability. Stoop (2005) notes that younger individuals are difficult to contact and engage due to lack of landlines and time spent away from home for work/leisure activities. Older people may be more willing to participate due to greater ties to their social community (Groves and Couper 1998).

Some exceptions to under-representation at younger ages were observed, which could have been related to question topics. For example, *Understanding Society* included crime and

security. The different age profile in that study could suggest that reduced focus upon health may have attracted different ages.

Recruitment strategies could also have affected response. Surveys employing methods with reduced direct contact seemed to under-represent women in the youngest and oldest age groups. Reliance upon charity mailing lists could reduce the younger and older people within the sampling frame because they may have had less disposable income to donate to these organizations. Interestingly, even when only middle to older aged individuals were the recruitment target, the youngest and oldest age groups were still under-represented in *EPIC* Norfolk, suggesting that contact methods (e.g. survey literature) may not appeal to all ages.

Overall, the data showed greater under-representation at higher levels of deprivation, suggesting that women with lower socioeconomic status were less willing to participate. Groves and Couper (1998) have attributed this, in part, to pessimism in those with lower socioeconomic status, which is translated into reduced willingness to help organizations viewed as contributors to their difficulties. However, under-representation was also seen at the lowest level of deprivation. This finding could perhaps support the idea that availability may be a further barrier to participation. People with lower incomes may be less accessible to contact due to irregular shift patterns and those with higher income due to long work hours but also staying away from home during leisure hours due to their greater disposable income.

Differences between cohort studies SES also suggested that recruitment strategies and survey topic were influential in response. Thus, *Breakthrough Generations* had a relatively large over-representation at the lowest level of deprivation, and this study relied on charity mailing lists and chain or snowball referral methods to contact new participants. This may have resulted in wealthier individuals contacting people within their social networks who may have been of similar socioeconomic status levels. Finally, as stated above, the *Understanding*

*Society* study was not simply focused on health but also upon wider social and economic issues. It is possible that these topics may have had more appeal to those with higher levels of deprivation.

This review showed a pattern of greater under-representation at lower educational levels. This is not unexpected, in the light of trends found for deprivation, because higher-earners often have more education. Yet Groves and Couper (1998) isolated education to specify that greater levels of education may be associated with increased willingness to respond due to ability to see the societal benefits of research. It would also make intuitive sense that the amount of reading and interpretation required in study materials could be a barrier to participation for populations whose educational attainment is low.

The data suggest, however, that other factors were involved in the pattern of education and response. Both *EPIC* studies used postal recruitment through patient lists from family doctors. Yet *EPIC* Norfolk was more successful at representing individuals with lower levels of education than *EPIC* Oxford. This may have been due to the fact that the topics were slightly different; the older age range in *EPIC* Norfolk could have mitigated lower education. *EPIC* Oxford extended the sampling frame by using vegetarian/vegan organizations. Interest in this diet may have been more relevant to those with higher educational qualifications, because higher levels of education have been observed in vegetarian groups (e.g. Gale et al. 2007). More difficult to explain were the differences observed in the general population health surveys in which the questionnaires and recruitments strategies were similar. More research is needed to understand why the WHS seemed more successful at representing populations with lower levels of education. This cannot be explained by sampling, topic or by prevalence of level 4 education in the population because, while lower than other nations in the UK, this is only by 1% compared with Scotland and by 2% compared with England (Office for National Statistics 2011).



Regarding ethnicity, white women were more likely to be over-represented, and only women classed as ‘Asian’ were under-represented. The use of boost samples in *Understanding Society* was successful at recruiting ethnic minorities. However, the resultant sample could not be used to provide estimates of health risk without statistical adjustment.

Interestingly, the levels of underrepresentation for different age groups were similar over time within the population surveys. Achieving high overall response rates can mitigate the bias due to the ‘healthy volunteer effect’. In the current climate of declining response rates; however, it is unlikely that sufficient overall response will be achieved to overcome this source of bias. Addressing underrepresentation may be a useful strategy to reduce the healthy volunteer effect because it is likely to bring more people from sociodemographic groups at risk (Buck and Frosini 2012) into the sample. Arguably, some of the differences between population and sample distributions were quite small. The included studies aimed to provide summary statistics for a population. However, other research has demonstrated that the prevalence of certain risk behaviors is higher in the under-represented groups than in the rest of the population and that reported reductions in risk behaviors are driven by changes in individuals with higher socioeconomic status (Catto 2008). Therefore, even these small differences may translate to problematic underestimations of harmful behaviors in the groups who are at greatest risk. An additional limitation of this work is that we chose to focus upon response patterns by social determinants only in women. Therefore, we were unable to comment on representation by these factors in men or sex differences in response according to these factors.

The key strength of this work was the integration of multiple surveys which employed different recruitment strategies and approached different health topics. However, to achieve this broad overview, it must also be acknowledged that pragmatic decisions had to be taken to integrate such a large amount of data. Classification differences, related to differences over

time and study objectives, made it difficult to compare studies exactly. For the same reason, certain attributes had to be excluded or reduced to broad categories, as in the case of ethnicity. A reliance on secondary data also meant that some studies did not contribute data for each sociodemographic attribute. However, we believe that because the aim of this study was to look at overall patterns, these difficulties did not prevent the aims of the review from being met.

## **Conclusion**

The findings of these analyses suggest that health survey results have not been representative of women from all sociodemographic groups in the UK. If single studies were considered in isolation, clear associations appeared between sociodemographic characteristics and participation. However, this work is useful because it has integrated the findings from multiple surveys to demonstrate that topic, recruitment strategies and fieldwork strategies may affect response patterns. Future research should now consider how to identify, reach and engage women in a way that improves sample representativeness. Over- and under-representation among men included in surveys and studies, and gender differences in over- and under-representation should also be explored.

## **References**

- Brick, M. J. 2013. Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics* 29:329–53.
- Buck, D. and Frosini, F. 2012. *Clustering of unhealthy behaviours over time: Implications for policy and practice*. London: The King's Fund.
- Buck, N., H. Laurie, and V. Nolan. 2011. Appendix: Understanding Society design overview. in *Understanding Society: Early findings from the first wave of the UK's household longitudinal study* *Understanding Society: Early findings from the first wave of the UK's*

*household longitudinal study*, edited by S. L. McFall and C. Garrington. Colchester: Institute for Social and Economic Research.

Cade, J. E., V. J. Burley, N.A. Alwan, J. Hutchinson, N. Hancock, M.A. Morris, D.E.

Threapleton and D.C. Greenwood. 2015. Cohort profile: The UK Women's Cohort Study (UKWCS) *International Journal of Epidemiology* October: 1–11.

Cade, J. E., V. J. Burley, and D. C. Greenwood. 2004. The UK Women's Cohort Study: Comparison of vegetarians, fish-eaters and meat-eaters. *Public Health Nutrition* 7:871–78.

Catto, S. 2008. How much are people in Scotland Really Drinking? Glasgow: Health Scotland.

Centres for Disease Control and Prevention. 2012. Introduction to epidemiology. In *Principles of Epidemiology in Public Health Practice*.

<http://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section1.html> (accessed March 1, 2016).

Coggon, D., G. Rose, and D. J. P. Barker. 2016. Case control and cross sectional studies. In *Epidemiology for the Uninitiated*. <http://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/8-case-control-and-cross-sectional> (accessed March 1, 2016).

Corbett, J., J. Day, M. Doig, S. Dowling, S. Martin and A. Stannard. 2014. The Scottish Health Survey 2013 edition: Volume 2, technical report. Edinburgh: The Scottish Government.

Davey, G. K. E.A. Spencer, P. N. Appleby, N. E. Allen, K. H. Knox and T. J. Key. 2003.

EPIC-Oxford: Lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. *Public Health Nutrition* 6:259–69.

Department of Health, Social Services and Public Safety (Northern Ireland), 2013, *Northern Ireland Health Survey, 2010-2011*, [data collection], UK Data Service. SN: 7258.

<http://dx.doi.org/10.5255/UKDA-SN-7258-1> (accessed March 24 2016).

Dunstan, S. 2012. General Lifestyle Survey 2010: Technical appendices. London.

<http://www.ons.gov.uk/ons/rel/ghs/general-lifestyle-survey/2010/index.html> (accessed 12 September, 2015).

Erens, B. et al. 2014. Methodology of the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *Sexually Transmitted Infections* 90:84–89.

Field, J., A. M. Johnson, J. Wadsworth, K. Wellings. 1995. *National Survey of Sexual Attitudes and Lifestyles, 1990*, [data collection], UK Data Service. SN: 3434.

<http://dx.doi.org/10.5255/UKDA-SN-3434-1> (accessed March 24 2016).

Fone, D. L., F. Dunstan, J. White, M. Kelly, D. Farewell, G. John, R. A. Lyons and K. Lloyd. 2013. Cohort profile: The Caerphilly Health and Social Needs Electronic Cohort Study (E-CATALyST). *International Journal of Epidemiology* 42:1620–28.

Fraser, A. et al. 2013. Cohort profile: The Avon Longitudinal Study of Parents and Children: ALSPAC Mothers Cohort. *International Journal of Epidemiology* 42:97–110.

Gale, Catharine R, Ian J Deary, Ingrid Schoon and G.David Batty. 2007. IQ in childhood and vegetarianism in adulthood: 1970 British Cohort Study. *British Medical Journal* 334 (7587): 245. doi:10.1136/bmj.39030.675069.55.

Gorman, E. et al. 2014. Assessing the representativeness of population-sampled health surveys through linkage to administrative data on alcohol-related outcomes. *American Journal of Epidemiology* 180:941–48.

Griffith, J. R., J. L. Clasey, J.T. King, S. Gantz, and R. J. Kryscio. 2007. Role of parents in determining children's physical activity." *World Journal of Pediatrics* 3:265–70.

Groves, R. M. and M. P. Couper. 1998. *Nonresponse in household interview surveys*. New York: Wiley Interscience.

Hayat, S. A. et al. 2014. Cohort profile: A prospective cohort study of objective physical and cognitive capability and visual health in an ageing population of men and women in Norfolk (EPIC-Norfolk 3). *International Journal of Epidemiology* 43:1063–72.

Health Education East Midlands. 2013. Your future: Table to show equivalencies between GCSEs, O Levels and CSEs. <http://www.yourfuture-eastmidlands.co.uk/content.aspx?CategoryID=252>). (accessed October 25, 2015).

Hutchings, Sally et al. 2014. Using the UK Biobank Study to estimate occupational causes of chronic disease: Comparability with the UK national Population and adjustment for bias.” *Occupational and Environmental Medicine* 71(Suppl\_1):A79–A79.

Inskip, H. M., K. M. Godfrey, S. M. Robinson, C. M. Law, D. J. P. Barker, C. Cooper and SWS Study Group. 2006. Cohort profile: The Southampton Women’s Survey. *International Journal of Epidemiology* 35:42–48.

Johnson, A. M. and Centre for Sexual and Reproductive Health Research, London School of Hygiene and Tropical Medicine, 2015, *National Survey of Sexual Attitudes and Lifestyles III, 2010-2012* [data collection], UK Data Service. SN:7799.

<https://discover.ukdataservice.ac.uk/catalogue?sn=7799> (accessed March 24, 2016).

Mathur, R., E. Grundy, and L. Smeeth. 2013. Availability and use of UK based ethnicity data for health research. *National Centre for Research Methods Working Paper Series* 1–30.

Medical Research Council. 2015. Cohort directory <https://www.mrc.ac.uk/research/facilities-and-resources-for-researchers/cohort-directory/> (accessed 24 February, 2015).

Mindell, J., J. P. Biddulph, et al. 2012. Cohort profile: The Health Survey for England. *International Journal of Epidemiology* 41:1585–93.

Mindell, J., M. Aresu, L. Becares, and H. Tolonen. 2012. Representativeness of participants in a cross-sectional Health Survey by time of day and day of week of data collection. *The European Journal of Public Health* 22:364–69.

National Career Service. 2015. Courses and learning: Qualifications table.

<https://nationalcareersservice.direct.gov.uk/advice/courses/Pages/QualificationsTable.aspx>

(accessed October 25, 2015).

National Centre for Social Research. 2011. *Welsh Health Survey, 2010*, [data collection], UK Data Service. SN: 6895. <http://dx.doi.org/10.5255/UKDA-SN-6895-1> (accessed March 24, 2016).

National Centre for Social Research. 2013. *Welsh Health Survey, 2011*, [data collection], UK Data Service. SN: 7188. <http://dx.doi.org/10.5255/UKDA-SN-7188-1> (accessed March 24, 2016).

National Centre for Social Research. 2014. *Welsh Health Survey, 2009*, [data collection], UK Data Service, *3rd Edition*. SN: 6589. <http://dx.doi.org/10.5255/UKDA-SN-6589-2> (accessed March 24, 2016).

National Centre for Social Research. 2014a. *Welsh Health Survey, 2012*, [data collection], UK Data Service. SN: 7459. <http://dx.doi.org/10.5255/UKDA-SN-7459-1> (accessed March 24, 2016).

National Centre for Social Research. 2015. *Welsh Health Survey, 2013*, [data collection], UK Data Service. SN: 7632. <http://dx.doi.org/10.5255/UKDA-SN-7632-1> (accessed March 24, 2016).

National Centre for Social Research et al. 2005. *National Survey of Sexual Attitudes and Lifestyles II, 2000-2001*, [data collection], UK Data Service. SN: 5223.

<http://dx.doi.org/10.5255/UKDA-SN-5223-1> (accessed March 24, 2016).

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. 2013. *Health Survey for England, 2011*, [data collection], UK Data Service. SN: 7260. <http://dx.doi.org/10.5255/UKDA-SN-7260-1> (accessed March 24, 2016).

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. 2014. *Health Survey for England, 2012*, [data collection], UK Data Service. SN: 7480. <http://dx.doi.org/10.5255/UKDA-SN-7480-1> (accessed March 24, 2016).

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. 2015. *Health Survey for England, 2009*, [data collection], UK Data Service, *3rd Edition*. SN: 6732. <http://dx.doi.org/10.5255/UKDA-SN-6732-2> (accessed March 24, 2016).

National Centre for Social Research, University College London. Department of Epidemiology and Public Health. 2015a. *Health Survey for England, 2013*, [data collection], UK Data Service. SN: 7649. <http://dx.doi.org/10.5255/UKDA-SN-7649-1> (accessed March 24, 2016).

National Centre for Social Research, Royal Free and University College Medical School. Department of Epidemiology and Public Health. 2015b. *Health Survey for England, 2010*, [data collection], UK Data Service, *3rd Edition*. SN: 6986. <http://dx.doi.org/10.5255/UKDA-SN-6986-3> (accessed March 24, 2016).

National Records of Scotland. 2015. Scotland's Census: Census data explorer. <http://www.scotlandscensus.gov.uk/obs-web/home.html> (accessed October 2 2015).

Office for National Statistics. 1991. Census: Aggregate data (England and Wales) [computer file]. <http://casweb.mimas.ac.uk> (accessed September 7, 2015).

Office for National Statistics. 2010. Standard Occupational Classification 2010 volume 3, The National Statistics Socio-Economic Classification: (Rebased on the SOC2010) user manual.

<http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html> (accessed March 24, 2016).

Office for National Statistics. 2011. General Lifestyle Survey 2009: Technical appendices. London.

<http://www.sherpa.ac.uk/romeo/definitions.php?la=en&flDnum=|&mode=simple&version=#colours> (accessed 23 February, 2015).

Office for National Statistics. 2011a. General Lifestyle Survey 2011, Technical appendix B: Sample design and response. London. <http://www.ons.gov.uk/ons/rel/ghs/general-lifestyle-survey/2011/index.html> (accessed 23 February, 2015).

Office for National Statistics. 2011c. Census: Aggregate data (England and Wales) [computer file]. <http://infuse.mimas.ac.uk> (accessed September 7, 2015).

Office for National Statistics. 2015. Focus on gender: Release calendar.

<http://www.ons.gov.uk/ons/rel/social-trends-rd/focus-on-gender/index.html> (accessed 24 January, 2016).

Office for National Statistics. 2015a. Persistent poverty in the UK and EU, 2008-2013.

<http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/household-income/persistent-poverty-in-the-uk-and-eu/2008-2013/persistent-poverty-in-the-uk-and-eu--2008-2013.html> (accessed 24 January, 2016).

Office for National Statistics. 2015b. Population estimates.

<http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Population+Estimates> (accessed October 10, 2015).



Olson, K. 2013. Do non-response follow-ups improve or reduce data quality?: A review of the existing literature. *Journal of the Royal Statistical Society. Series A: Statistics in Society* 176:129–45.

Platt, L. 2013. Consultation on an immigrant boost and a refreshment of the ethnic minority boost for Understanding Society: The UK Household Longitudinal Study.

<https://www.understandingsociety.ac.uk/2013/08/12/redesign-of-immigrant-boost-sample-consultation-extended> (accessed 23 January, 2016).

Robinson, S. M., S. R. Crozier, S. E. Borland, J. Hammond, D. J. P. Barker, H. M. Inskip. 2004. Impact of educational attainment on the quality of young women's diets. *European Journal of Clinical Nutrition* 58:1174–80.

Rothman, K. J., J. E. J. Gallacher, and E. E. Hatch. 2013. Why representativeness should be avoided. *International Journal of Epidemiology* 42:1012–14.

Sadler, K., M. Doyle, D. Hussey, and R. Stafford. 2013. *Welsh Health Survey 2012: Technical report*. London: NatCen Social Research.

Scottish Centre for Social Research, University College London. Department of Epidemiology and Public Health, University of Glasgow. MRC/CSO Social and Public Health Sciences Unit. 2015. *Scottish Health Survey, 2012*, [data collection], UK Data Service, 3rd Edition. SN: 7417. <http://dx.doi.org/10.5255/UKDA-SN-7417-3> (accessed March 24, 2016).

Scottish Centre for Social Research. 2015a. *Scottish Health Survey, 2013*, [data collection], UK Data Service, 2nd Edition. SN: 7594 <http://dx.doi.org/10.5255/UKDA-SN-7594-2> (accessed March 24, 2016).

Scottish Centre for Social Research, University College London. Department of Epidemiology and Public Health. 2016. *Scottish Health Survey, 2009*, [data collection], UK

Data Service, *5th Edition*. SN: 6713. <http://dx.doi.org/10.5255/UKDA-SN-6713-3> (accessed March 24, 2016).

Scottish Centre for Social Research, University College London. Department of Epidemiology and Public Health. 2016a. *Scottish Health Survey, 2010*, [data collection], UK Data Service, *2nd Edition*. SN: 6987. <http://dx.doi.org/10.5255/UKDA-SN-6987-2> (accessed March 24, 2016).

Scottish Centre for Social Research, University College London. Department of Epidemiology and Public Health, University of Glasgow. MRC/CSO Social and Public Health Sciences Unit. 2016b. *Scottish Health Survey, 2011*, [data collection], UK Data Service, *4th Edition*. SN: 7247. <http://dx.doi.org/10.5255/UKDA-SN-7247-4> (accessed March 24, 2016).

Scottish Credit and Qualifications Framework Partnership. 2015. Scottish Credit and Qualifications Framework. <http://scqf.org.uk/> (accessed 7 September, 2015).

Scottish Public Health Observatory. 2015. Gender: Gender inequalities in wider determinants of health. <http://www.scotpho.org.uk/population-groups/gender/data/gender-inequalities-in-wider-determinants-of-health> (accessed 5 September, 2015).

Stoop, I. A. L. 2005. *The hunt for the last respondent: Non-response in sample surveys*. The Hague: Social and Cultural Planning Office of the Netherlands.

Swerdlow, A. J., M. E. Jones, M. J. Schoemaker, J. Hemming, D. Thomas, J. Williamson, A. Ashworth. 2011. The Breakthrough Generations Study: Design of a long-term UK cohort study to investigate breast cancer aetiology. *British Journal of Cancer* 105:911–17.

Tillin, T., N. G. Forouhi, P. M. McKeigue, N. Chaturvedi. 2012. Southall and Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins. *International Journal of Epidemiology* 41:33–42.

Uhrig, S. C. N. 2008. The nature and causes of attrition in the British Household Panel Survey. ISER Working Paper *Series* 05:1–85.

University of Essex. Institute for Social and Economic Research, NatCen Social Research.

2015. *Understanding Society: Waves 1-5, 2009-2014*, [data collection], UK Data Service, 7th Edition. SN: 6614. <http://dx.doi.org/10.5255/UKDA-SN-6614-7> (accessed March 24, 2016).

Wilkinson, R., H. Inskip, and Southampton Women's Survey Study Group. 2006. *The health and wellbeing of women in Southampton's Communities: An analysis of the Southampton Women's Survey*. Southampton.

[http://www.publichealth.southampton.gov.uk/Images/Southampton Women's Survey Briefing May'06.pdf](http://www.publichealth.southampton.gov.uk/Images/Southampton%20Women's%20Survey%20Briefing%20May'06.pdf) (accessed 23 February, 2015).

#### Table Titles

**Table 1 – Included large population surveys: sample characteristics, sample size, geographic regions, dates covered and study focus**

**Table 2 – Included cohort studies: cohort characteristics, sample size, geographical regions, recruitment period and study focus**

#### Figure Legends

**Figure 1: Difference between percentage of female respondents within general population survey samples and population estimates by age group: NIHS, HSE, SHeS, WHS and GLS for years 2009 to 2013**

**Figure 2: Difference between percentage of female respondents in Natsal I (1990-1991), II (1999-2001) and III (2010-2011) and population estimates for England and Wales by age group**

**Figure 3: Difference between percentage of female respondents within cohort study samples and population estimates by age group**

**Figure 4: Difference between percentage of female respondents within general population survey and cohort study samples and population estimates by levels of deprivation**

**Figure 5: Difference between percentage of female respondents within general population survey and cohort study samples and population estimates by highest level of education attained**

**Figure 6: Difference between percentage of female respondents within general population survey (Natsal -3, SHeS and HSE) and cohort study samples (*Understanding Society*) and population estimates by ethnic group**



**Patterns of response by sociodemographic characteristics and recruitment methods for women in UK population surveys and cohort studies.**

**Table 1 – Included large population surveys: sample characteristics, sample size, geographic regions, dates covered and study focus**

<b>Survey</b>	<b>Sample Characteristics</b>	<b>Latest Available Sample Size (Year)</b>	<b>Geographic Region</b>	<b>Dates Covered by Survey (Frequency)</b>	<b>Study Focus</b>
General Lifestyle Survey (GLS) <sup>a</sup>	Adults aged 16 and over in private households.	18,367 (2011)	Great Britain	1971 -2012 (annual)	Multipurpose – gathers information about households and families for use by education, health, business and charity sector. Sexual behaviour
National Survey of Sexual Attitudes and Lifestyles (Natsal)	Natsal 1&2 – adults aged 16 to 44 Natsal 3 – adults aged 16-74	15,162 (2010)	Great Britain	1990 – 2011 (every 10 years)	
Health Survey for England (HSE) <sup>b</sup>	Approximately 8000 Men and women in private households aged 16 and over	10,980 (2013)	England	1991 to 2013 <sup>b</sup> (annual)	Health – risk factors, past and current health status
Northern Ireland Health Survey (NIHS) <sup>b</sup>	Adults in private households aged 16 and over.	3059 (2010-2011)	Northern Ireland	2010/2011	
Scottish Health Survey (SHeS) <sup>b</sup>	Adults in private households aged 16 and over	6,733 (2013)	Scotland	1995, 1998, 2003, 2008 – 2013 <sup>b</sup> (annually)	
Welsh Health Survey (WHS) <sup>b</sup>	Adults aged 16 and over in private households	15,007 (2013)	Wales	2003/4 – 2013 <sup>b</sup>	

<sup>a</sup> This survey was named the ‘General Household Survey’ until 2008.

<sup>b</sup> These surveys continue beyond this date but data were unavailable for all surveys at the time of writing.

**Patterns of response by sociodemographic characteristics and recruitment methods for women in UK population surveys and cohort studies.**

**Table 2 – Included cohort studies: cohort characteristics, sample size, geographical regions, recruitment period and study focus**

<b>Cohort Study</b>	<b>Cohort Characteristics</b>	<b>Sample Size at Recruitment</b>	<b>Geographical Region</b>	<b>Recruitment Period</b>	<b>Study Focus</b>
<i>ALSPAC</i>	Mothers/ pregnant women aged 16 to 45 years	Mothers 14,541	Hampshire England	1991-1992	Health and educational outcomes for mothers and their children
<i>Breakthrough Generations</i>	Women aged >16 years.	112,798	British Isles	2003-2011	Links between lifestyle characteristics and cancer.
<i>The Caerphilly Health and Social Needs Electronic Cohort Study (E-CATALyST)</i>	Adults aged $\geq 18$ years	17,979	Caerphilly	1998-2008	Sources of health inequalities.
<i>European Prospective Investigation of Cancer, Norfolk (EPIC-Norfolk)</i>	Adults aged 40-79 years. Living in Norfolk and surrounding areas.	30,000	Norfolk	1993-1997	Diet, exercise and lifestyle and cancer.
<i>European Prospective Investigation of Cancer, Oxford (EPIC-Oxford)</i>	Men and women aged 17 to 98 years. Mainly vegetarians	65,000	Oxfordshire, Buckinghamshire and Greater Manchester	1993-2000	Influence of diet on cancer risk.
<i>Southampton Women's Study (SWS)</i>	Women aged 20-34 years	12,583 women	Southampton	1998-2002	Dietary and lifestyle factors.
<i>UK Women's Cohort Study (UKWCS)</i>	Women aged 35-62 years	35,372	England, Wales and Scotland	1994-2002	Links between diet and cancer in the UK
<i>Understanding Society</i>	Men and women of all ages	40,000 households (100 000 individuals) Health questions for 20 000 individuals	UK (including North Caledonian Canal)	2009-2014	Social and economic status and attitudes.

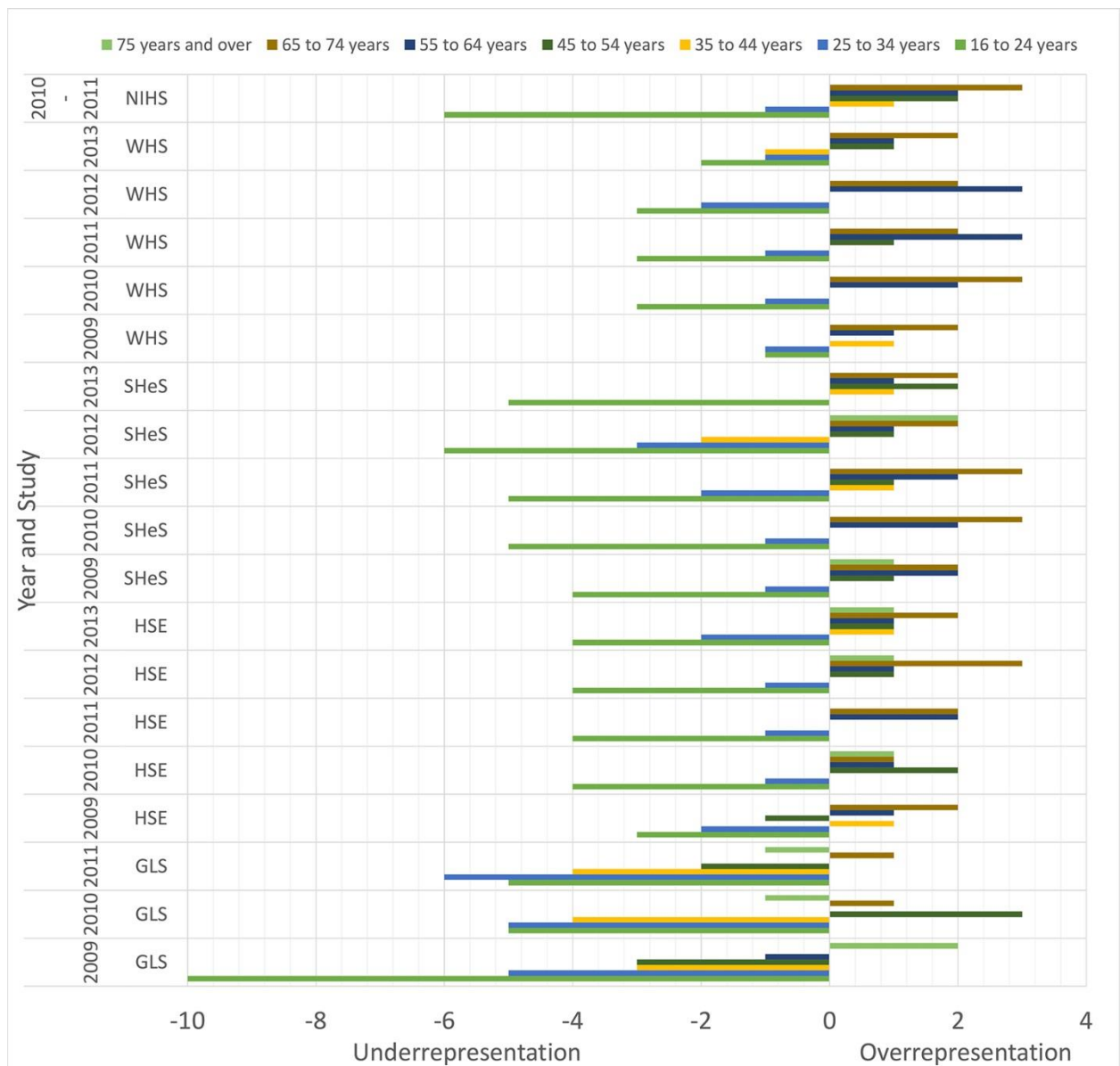


Fig 1



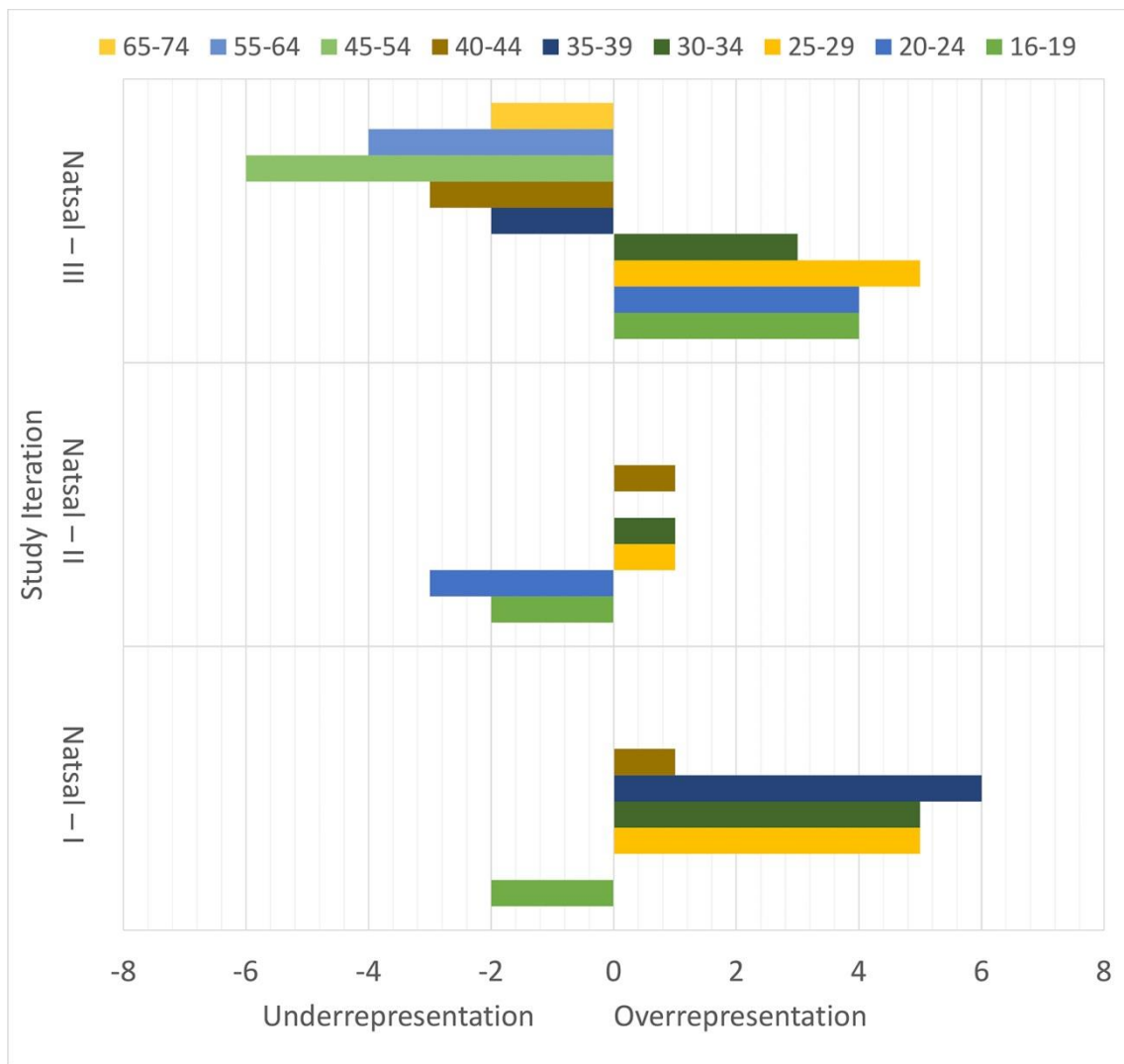


Fig 2

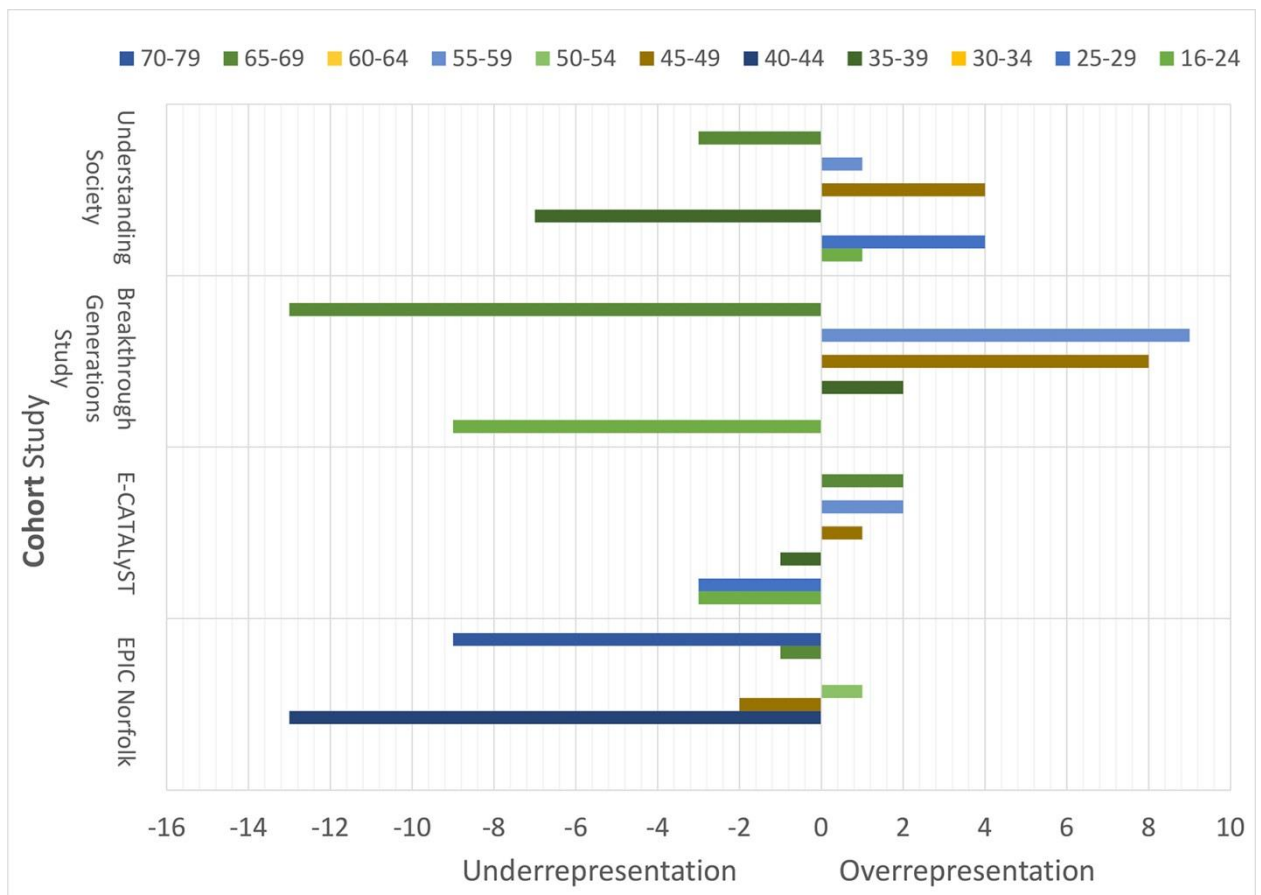


Fig 3

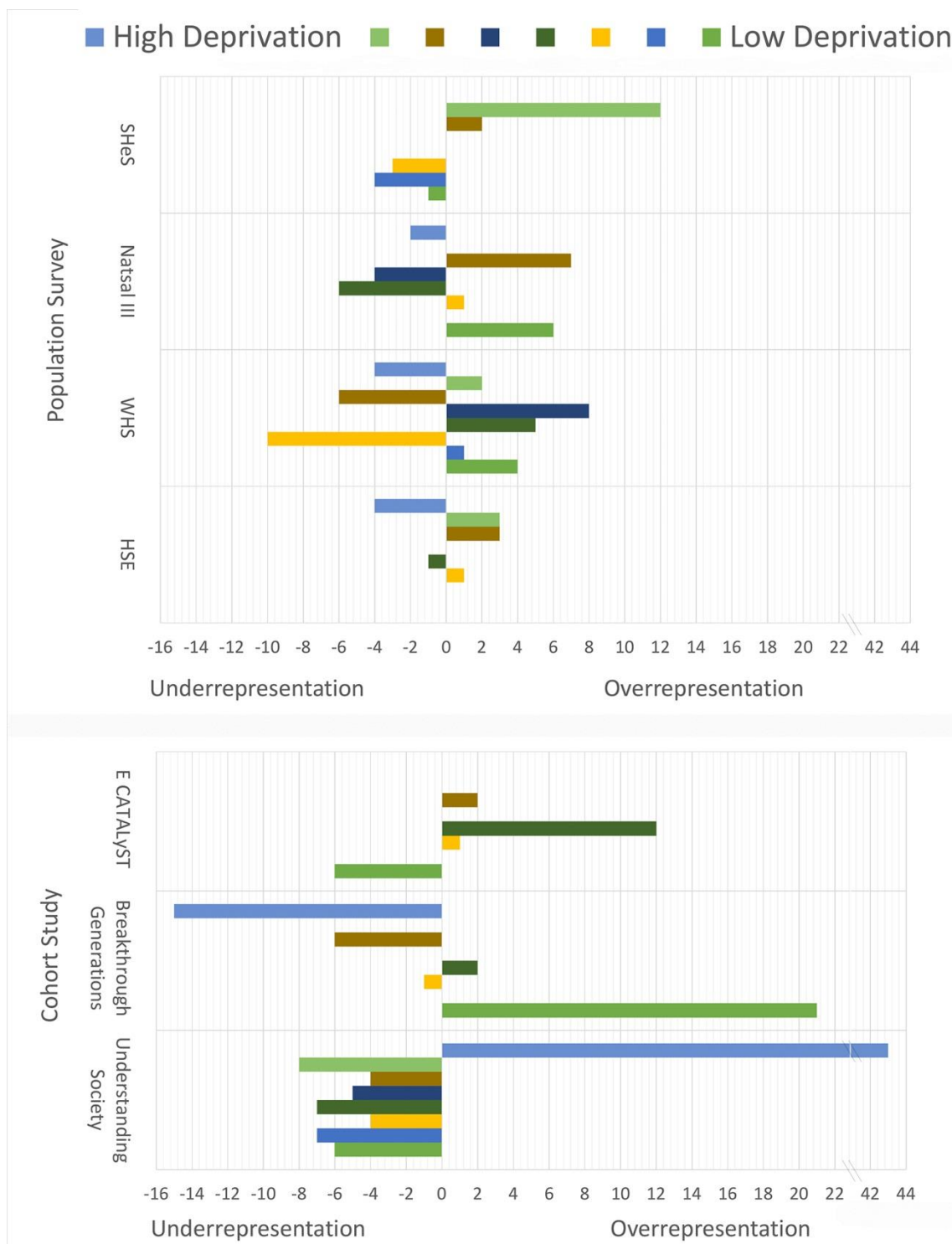


Fig 4

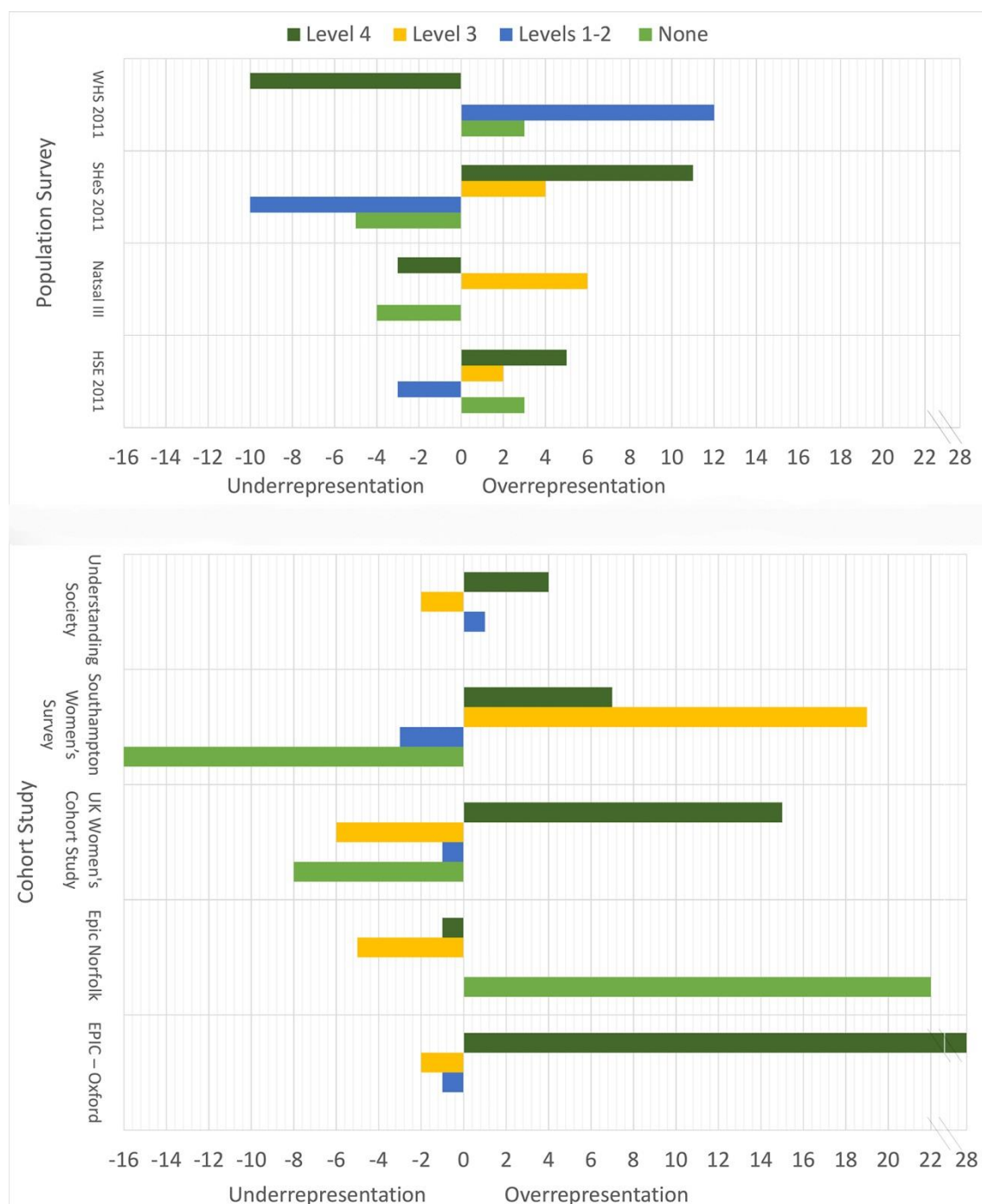


Fig 5

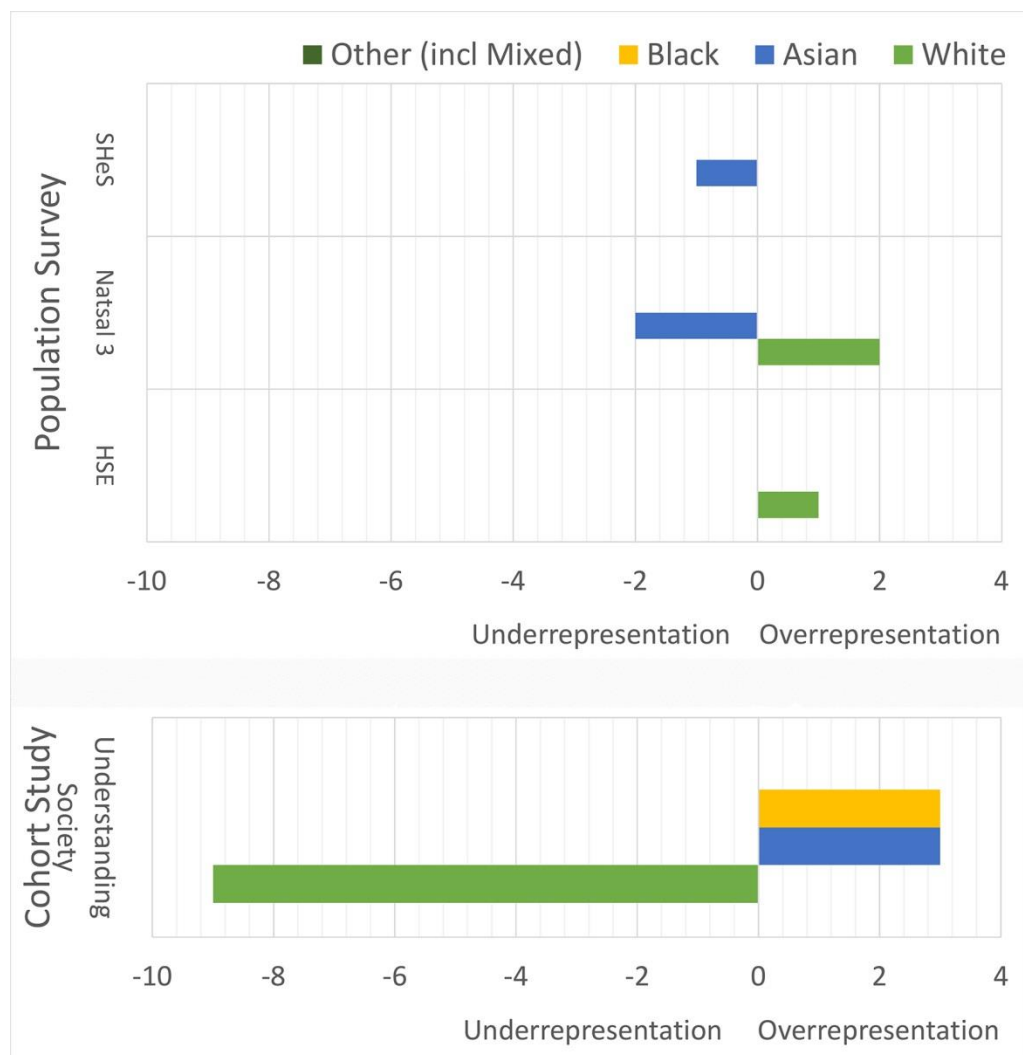


Fig 6