

Clustering of unhealthy behaviours over time

Implications for policy and practice

Authors: David Buck, Francesca Frosini

The sample

We used data from the Health Survey for England (HSE) in 2003 and 2008 covering the adult population aged 16 years and over. Both surveys were designed to be representative of the English population living in private households and adopted a multi-stage probability sampling design. The details of the sampling strategy are available from the University of Essex's Economic and Social Data Service website (see www.esds.ac.uk/government/hse/starting/), which hosts the survey, and The Information Centre (2009b).

In summary, postcode sectors were used as the primary sampling units (PSUs) and randomly selected after having been stratified by local authority and, within each local authority, by the percentage of households in the 2001 Census with a head of the household in a non-manual occupation. Within PSUs, postal addresses were randomly sampled and then used to identify households. All adults within a household were then selected for interviews.

The total dataset in 2003 consisted of 14,836 adult individuals at 8,867 household addresses that were selected from 720 postcode sectors. We excluded participants with missing values on lifestyle factors or socio-demographic characteristics (n=229), resulting in a final dataset comprising 14,607 individuals. The total dataset in 2008 comprised 15,102 adult individuals at 9,961 household addresses selected from 1,176 postcode sectors. We excluded participants with missing values on lifestyle factors or socio-demographic characteristics (n=190), resulting in a final dataset comprising 14,912 individuals.

Data definitions

We extracted data concerning four lifestyle factors: smoking, alcohol consumption, poor diet and low level of physical activity. We created four binary variables classified as adherent to national guidelines or non-adherent.

Smoking

Adherence to recommendation was based on the variable CIGNOW, (whether smoke cigarette nowadays). People with 'yes' were classified as non-adherent; people with 'no' were classified as adherent. Therefore, former smokers were considered adherent, regardless of the length of time since ceasing smoking.

Alcohol consumption

Adherence was based on current recommendations for sensible daily alcohol intake defined by consumption of <3 units of alcohol for women and <4 units of alcohol for men (Department of Health 2007). There were a number of questions that contribute to the construction of the alcohol consumption variable. Those respondents who stated they had drunk in the past week

(D7DAY) were asked questions about the amount of different types of alcohol (including normal and strong beer, spirits, sherry, wine and alcopops) drunk on the day in the previous week that they had drunk the most. These amounts were then converted into units of alcohol for each type of drink. In 2006 an updated method for converting volume of alcohol into units was developed, to reflect changes in assumptions about strength of drinks and glass sizes (Goddard 2007). To reflect these changes, in 2007 the Health Survey for England introduced new questions on wine consumption that distinguished between glass size (small (125ml), medium (175ml) and large (250ml)) and quantity of wine drunk in bottles. This is a substantial difference to the 2003 HSE questionnaire, which asked only about the number of glasses drunk with no definition of glass size. To make the 2003 and 2008 data comparable we applied the new conversion factors for drinks other than wine to both 2003 and 2008 data. For wine we used a temporary unit conversion of two units for a glass of wine for both 2003 and 2008 data (Goddard 2007). To calculate adherence to guidelines, units of each kind of alcohol were then added up and benchmarked against the guidelines limits. Those stating they never drank or they had not drunk in the previous week were considered as having consumed zero units of alcohol.

Poor diet

Adherence was based on the current guidelines that adults and children should aim to eat five or more portions of fruit and vegetables each day (Department of Health 2010). The description of the methods used in HSE to calculate fruit and vegetable consumption are described elsewhere (The Information Centre 2009a). In summary, people were asked about any fruit and vegetables consumed on the day before the interview. Foods contributing to the daily vegetable and fruit intake include pulses; salad; fresh, tinned and frozen vegetables; fresh, tinned, frozen and dried fruit; and fruit juices. People were asked to identify the amount in everyday units. For example, people were asked how many cereal bowls of salad or tablespoons of pulses they had eaten. For fresh fruit, people were first asked what kind of fruit they had eaten (from very large to very small) and then, depending on the size, how many slices, pieces or handfuls. A portion size is defined for different food items using these everyday units. For example, a portion of salad equates to one cereal bowl while a portion of vegetables equates to three tablespoons (see The Information Centre 2009a for details). In counting up portions, some items like fruit juice and pulses have been capped to one portion. To calculate adherence to national guidelines the number of portions from all the food included is added up and benchmarked against the guidelines limit.

Physical activity

The most recent guidelines (Department of Health 2011) recommend that adults (including older adults (65 and over)) should engage in at least 150 minutes of moderate intensity physical activity per week. It further indicates that 30 minutes on five or more days a week (which was the headline guideline in 2004 (Department of Health 2004)) is a way of achieving this target. The question structure within the HSE, especially the 2003 questionnaire, allowed calculation of the number of days with activity spells of 30 minutes' duration but did not allow us to establish the aggregate time spent in different bouts of activity on each day (and therefore over a week). Thus, in line with the approach followed in the HSE official documents (The Information Centre 2009a), we assessed adherence to recommendations concerning exercise based on engaging in at least 30 minutes of moderate activity on at least five days a week.

The 2008 survey included an extended physical activity questionnaire with two key differences from prior years: lower duration limit for an activity to be counted was 10 minutes rather than 30 as in 2003; more detailed questions on occupational activity rather than a single question as in 2003. To make the data comparable, the 2008 data has been analysed using 30 minutes as the lower duration for activities to be counted and including only the original question on occupational activity. This approach is detailed in the documentation for the 2008 HSE (The Information Centre 2009a).

To derive the number of days in a week on which a person has 30 minutes of moderate activity, we used the variable ADTOT30 for the 2003 survey and the variable A30TO06 (which uses the same definitions as earlier surveys and is used for trends analysis) for the 2008 survey (UK Data Archive 2003, 2008). In summary, people were asked whether in the previous four weeks they had engaged in a number of activities, including home activities (eg, heavy housework and gardening), walking, sports and occupational activity. For each category they were then asked the number of days they engaged in the above activity over the previous four weeks. In the 2003 survey this question was limited to the number of days they had done the activity for at least 30 minutes. In the 2008, they were then asked how long on average they had spent on the activity. We counted the days only when the respondent indicated 30 minutes or more on average. For some activities (like walking or sports), there were also questions about the intensity, which was accounted for in assessing whether the 30 minutes involved moderate or intense activity. For example, walk at a 'fairly brisk' or 'fast pace' was classified as moderate activity, 'slow' or 'average' pace was classified as light activity. Adherence to guidelines was calculated summing the number of days a person was moderately active for at least 30 minutes within the previous four weeks on the above activities. People who were active on 20 or more days over the previous four weeks were classified as adhering to the guidelines (ie, active five or more days a week).

Age

We used age at last birthday (variable AGE) categorised into five year bands: 16–24, 25–44, 45–54, 55–64, and 65 and over.

Educational qualification and social class

Income, education and occupation are often used interchangeably in studies of health behaviour as proxies for economic status or material wellbeing. In this study we use both education and an occupational measure of social class as complementary markers of socio-economic status, since they may capture, at least in part, different mechanisms influencing lifestyle behaviours.

Education is likely to be linked to health behaviour through two main mechanisms. Education can reflect greater material wellbeing as it is likely to influence opportunities for job and income. In turn, greater economic resources imply access to better food, safer environments, and better housing, all related to healthier lifestyle choices. But education can also reflect an important range of non-economic characteristics such as cognitive skills, literacy, knowledge, prestige and control. Education therefore increases a person's ability to access and process information and prompts greater influence over one's life, leading to healthier lifestyles. However, education is not a perfect proxy and has its problems. While it is fixed for most people early in life, it is a generally stable influence, and is useful in picking up long-run influences on health behaviours, it is less likely to reflect changes in personal – and particularly material – circumstances that will affect people's behaviour.

Social class, here measured by occupational classification of the head of the household, has been shown to reflect psychosocial links related to sense of empowerment, social integration, and stress – other mechanisms widely seen as important in the adoption of health behaviours. Social class, however, also has its weaknesses. While it is less fixed than education and therefore should be better at reflecting influences on behaviours that impact over the life-course itself, classifications based on occupation are problematic among adolescents, and older people, as those in employment will be a minority. In addition, standard occupational classifications are generally based on the 'head of the household', rather than the individual and therefore may not be as good at picking up affects, particularly women, who are not the head of household. For all these reasons we see education and social class as complementary explanatory variables rather than substitutes, and therefore include both in our analysis.

We measured education in terms of credentials earned using the derived variable TOPQUAL2 (highest educational qualification) to build a four categories qualification variable as follows:

- higher education – NVQ4/NVQ5/degree or equivalent and higher education below degree
- intermediate – NVQ3/GCE A Level equivalent/4 NVQ2/GCE O Level equivalent/5 NVQ1/CSE other grade equivalent
- no qualification
- full-time student.

We used the social class of the household representative according to the specific occupational groups from the Registrar General's occupation-based classification. Starting from the variable SCHRPF we derived four groups:

- I and II – professional and managerial/technical
- IIIIN – skilled non-manual
- IIIM – skilled manual
- IV and V – semi-skilled manual and unskilled manual.

Economic status

We used the derived variable ECONACT to build a three category indicator:

- economically active – in employment
- inactive – International Labour Office definition unemployed and other economically inactive
- retired.

Methodological notes

Accounting for the survey sample design

As described above, the Health Survey for England has a complex multi-stage sample design that incorporates clustering (individuals are nested within households, which are nested within the postcode sectors) and stratification (PSUs are stratified by local authority and households with a head of the household in a non-manual occupation). Therefore, the correct estimation of standard errors requires special techniques that account for the design effects of the sample design (Rafferty 2011). We used a design-based approach (ie, using the sampling design variables provided in the dataset) to incorporate these features into our analysis using the STATA 10 built-in survey estimation commands (ie, svyset commands). The key sample design variables provided were AREA and PSU to identify PSUs in 2003 and 2008 respectively, HSERIAL to identify clustering within household and CLUSTER to account for stratification.

In addition, HSE provides a set of weighting variables that are specific to the analysis conducted. For individual level analysis on the adult sample there are two sets of weights to be considered:

- a household weight that corrects for dwelling units and household selection and for the distribution of household members to match population estimates for sex/age groups and Government Office Region (GOR)
- an individual non-response weight that corrects for non response bias.

All our estimates are conducted on weighted data using the weighting variable Wt_int, which combines the household and individual weights.

Age standardisation

Data has been age standardised to allow comparisons between 2003 and 2008 and between groups (eg, social class or education). It should be noted that all analyses are reported separately for men and women. Standardisation has been carried out separately within each

sex. Therefore, when comparing data between sexes, no standardisation was undertaken to remove the effects of differential age distribution between sexes.

Age standardisation has been conducted using the direct method of standardisation based on the mid-year 2007 population estimates for England. The standardisation takes into account the sample design of the survey.

Confidence intervals for the observed to expected ratios

Confidence intervals for the observed to expected ratios were calculated using the exact method based on the Poisson distribution.

Multinomial logistic regression

To examine the relationship between socio-demographic variables and multiple risk factors an unordered multinomial logistic regression was developed to evaluate the factors associated with reporting one, two, three or four risk factors versus reporting no risk factors. Results are expressed as a series of logistic regressions in terms of odds ratios. In a nutshell, the odds ratio compares the odds of a higher number of lifestyle risk factors versus having none for one category of the independent variable compared to the reference category (eg, no qualifications vs higher education). To continue with the example of education, an odds ratio greater than one implies that those with no qualification are more likely to have a higher number of lifestyle risk factors than those with higher education. We ran separate models for education and social class adjusting for sex, age and economic status.

Given the ordered nature of the dependent variable, an ordered multinomial logistic model would have been appropriate. However, the required assumption of proportional odds was not met ($p < 0.000$).

The analysis was conducted adjusting for the survey sample design. Therefore estimates are weighted and standard errors are adjusted for clustering and stratification effects. Change between 2003 and 2008 was assessed using post-estimation tests for linear combinations of parameter estimations available for survey data.

Prevalence odds ratios

To examine the association between pairs of risk factors (eg, smoking and drinking) we estimated prevalence odds ratios through a series of logistic regressions where one risk factor (the dependent variable, eg, drinking) was regressed on another (the independent variable, eg, smoking), adjusting for age. Continuing with the example, the estimated odds ratio on smoking compares the odds of drinking for a smoker versus a non-smoker. If the odds ratio is greater than one and significant, it implies that smokers are more likely to drink than non-smokers.

As with the multinomial logistic regression we assessed change between 2003 and 2008 in prevalence odds ratios using post-estimation tests for linear combinations of parameter estimations available for survey data.

Summary measures of inequalities

To further analyse socio-economic inequalities and change in the socio-economic gap in multiple risk factors between 2003 and 2008, we calculated four summary measures of inequalities whose characteristics are described in the table below. The relative change between 2003 and 2008 was calculated as the ratio of the measures of 2008 to 2003. Similarly to previous studies (Drieskens *et al* 2009), we determined socio-economic inequalities and change over time using the interaction test (Altman *et al* 2003).

Measure	Characteristics	Definition	Interpretation
Prevalence difference	<ul style="list-style-type: none"> • Effect • Absolute 	The absolute difference between the age-adjusted prevalence of the lowest vs highest socio-economic status (SES) group.	Percentage points.
Odds ratio	<ul style="list-style-type: none"> • Effect • Relative 	Based on a logistic regression relating whether an individual has high-risk lifestyle (versus a low-risk lifestyle) to SES status, controlling for age.	The odds of having a risky lifestyle for somebody with lower SES versus somebody with higher SES.
Population attributable risk	<ul style="list-style-type: none"> • Impact • Relative 	The difference between the overall prevalence of multiple risk behaviours and the prevalence for the highest socio-economic group.	How much (percentage change) the prevalence of multiple behaviour would improve if the population had the same lifestyle of the highest socio-economic group (higher values indicate higher socio-economic inequalities).
Relative index of inequality	<ul style="list-style-type: none"> • Impact • Relative 	Based on a logistic regression relating to whether an individual has a high-risk lifestyle (versus low risk) to the position of each SES group on a scale from 0 to 1 (hence accounting for population size and SES position of groups).	Relative risk of those at the bottom of the SES ladder compared with those at the top assuming an association between multiple risk factors and SES position for all groups.

Study limitations

There are some limitations to this study. First, because we use survey data to measure lifestyle risk factors, non-response and misclassification are likely to influence different behaviours to differing degrees. For example, it is well known that people tend to under report alcohol consumption. Also, consumption of fruit and vegetables (poor diet) is based on a single 24-hour recall, and therefore likely to underestimate usual intake. In addition, the necessity to make data from 2003 and 2008 comparable might have introduced further issues; for example, we may have underestimated consumption of alcohol in 2008. However, our prime interest is in changes over time, which should reduce the impact of these problems as a whole.

Second, while the dichotomisation of risk factors and the choice of our cut-off points allowed the alignment of our definitions of risk factors to the current national health recommendations, it might have implications for the findings (Poortinga 2007) and might limit generalisation to context and populations that are different from the English context.

Finally, the cross-sectional nature of the data does not allow us to make strong causal claims or to test hypothesis of cohort effects or impact of SES dynamics over time, although we do comment on whether our results are consistent with hypotheses about changes over time.

References

Altman DG, Bland JM (2003) 'Interaction revisited: the difference between two estimates'. *British Medical Journal*, vol 326, p 219.

Drieskens S, Van Oyen H, Demarest S, Van der Heyden J, Gisle L, Tafforeau J (2010). 'Multiple risk behaviour: increasing socio-economic gap over time?' *European Journal of Public Health*, vol 20, no 6, pp 634–9 [ePub 2009].

Department of Health (2011). 'UK physical activity guidelines'. Department of Health website. Available at:

www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_127931 (accessed on 13 August 2012).

Department of Health (2010). '5 A DAY portion information'. Department of Health website. Available at:

http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publichealth/Healthimprovement/FiveADay/FiveADaygeneralinformation/DH_4001494 (accessed on 13 August 2012).

Department of Health (2007). *Safe. Sensible. Social. The next steps in the National Alcohol Strategy*. London: Department of Health. Available at:

www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_075218 (accessed on 20 August 2012).

Department of Health (2004). *At Least Five a Week: Evidence on the impact of physical activity and its relationship to health*. A report from the Chief Medical Officer. London: HM Government. Available at:

www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4080994 (accessed on 20 August 2012).

Goddard E (2007). 'Estimating Alcohol Consumption from Survey Data: Updated method of converting volumes to units'. National Statistics Methodological Series No 37. Office for National Statistics website. Available at:

www.ons.gov.uk/ons/search/index.html?newquery=Estimating+Alcohol+Consumption+from+Survey+Data%3A+Updated+method+of+converting+volumes+to+units (accessed on 20 August 2012).

The Information Centre (2009a). *Health Survey for England 2008. Volume 1 Physical activity and fitness* [online]. The Information Centre website. Available at:

www.ic.nhs.uk/webfiles/publications/HSE/HSE08/Volume_1_Physical_activity_and_fitness_revised.pdf (accessed on 20 August 2012).

The Information Centre (2009b). *Health Survey for England 2008. Volume 2. Methods and documentation* [online]. The Information Centre website. Available at:

www.ic.nhs.uk/webfiles/publications/HSE/HSE08/Volume_2_Methods_and_Documentation.pdf (accessed on 20 August 2012).

Poortinga W (2007). 'The prevalence and clustering of four major lifestyle risk factors in an English adult population'. *Preventive Medicine*, vol 44, no 2, pp 124–8.

Rafferty A (2011). *Introduction to Complex Sample Design in the UK Government Surveys. ESDS Government* [online]. Economic and Social Data Service website. Available at:

www.esds.ac.uk/government/docs/complexsampledesign.pdf (accessed on 21 August 2012).

UK Data Archive (2003) 'Health Survey for England, Cardiovascular Disease 03. number 5098. List of Variables'. Economic and Social Data Service website. Available at: www.esds.ac.uk/findingData/snDescription.asp?sn=5098 (accessed on 21 August 2012).

UK Data Archive (2008). 'Health Survey for England, Physical activity and fitness 2008. number 6397. List of Variables.' Economic and Social Data Service website. Available at: www.esds.ac.uk/findingData/snDescription.asp?sn=6397 (accessed on 21 August 2012).