Back pain prevalence models for small populations

Technical document produced for Arthritis Research UK

Samanta Adomaviciute, Michael Soljak, Julian Gardiner, Kim Foley, Hilary Watt, Roger Newson

Department Primary Care & Public Health
School of Public Health

v5.5 October 2018

Imperial College London

Contents

Back pain prevalence models

1 Executive Summary

Back pain is a very common condition and a major cause of disability. UK data from the 2016 Global Burden of Disease study shows that musculoskeletal (MSK) conditions cause the third largest loss of disability-adjusted life years (DALYs) after cardiovascular disease and cancer, and back pain causes the largest loss for any single disease. In a UK retrospective cohort study of health care costs associated with the treatment of chronic i.e. recurrent low back pain, total health care costs for patients were double those of the matched controls (£1074 vs. £516; $P < 0.05$).[1] Of the cost difference, 58.8% was accounted for by general practitioner's consultations, 22.3% by referrals to secondary care, and the rest by pain relief medications. We were commissioned by Arthritis Research UK (ARUK) to develop a small population prevalence model for lower back pain, using nationally-representative data on risk factors applied to matching local population data. Risk factors for back pain include age, biomechanical and occupational factors, community size, depression, educational level, gender, obesity, socioeconomic factors, low physical activity, psychosocial factors, and smoking.

We used Health Survey for England (HSfE) 2011 data[2] to fit a regression model to and to produce from that local estimates of lower back pain. There are several validated instruments which measure the severity of chronic pain. The Graded Chronic Pain Scale (GCPS), also known as the "von Korff"score after its developer, is a well-validated measure of pain intensity and pain-related disability related to a 6-month reference frame.[3] Out of 10,617 respondents 1,346 (12.68%) experienced back pain, while 7,415 (69.84%) did not experience back pain and 1,856 (17.48%) did not have information regarding back pain. We fitted univariate then multivariate logistic regression models for overall and severe back pain as described in previous publications.[4] A range of multivariate regression models were fitted in order to obtain the best performing. We included one additional variable at a time to observe the effects: the variance inflation factor (vif) command was used after model fitting to check for multicollinearity.

The local population estimates and their confidence intervals are produced in Stata software. In summary, the proportion of our population according to age and sex are known. Then, for example, the proportion by educational status can be applied to these numbers, taking account of the fact that the distribution by educational status differs by age group. This gives estimated proportion by age, sex and educational status and so on. A new set of variables is created, one for each combination of these risk factors pertinent to the logistic regression model for the chosen disease. We do not want to produce a synthetic estimate on just one population, but rather on many populations, for instance on each local authority separately. Therefore we produced a dataset containing information on the risk factors in all the different local authorities and other geographies, with one line of data per geography. There is uncertainty in these synthetic estimates of prevalence based on the imprecision in the estimated coefficients from the logistic regression equations. A bootstrap procedure can be used to construct confidence intervals on these synthetic estimates of prevalence, based on the imprecision in these logistic regression coefficients. We used the model developed from HSfE to produce the prevalence of overall back pain and severe back pain for Scotland and Wales.

We went through an extensive model fitting process. The two multivariate regression models M26b and M28 were chosen for overall back pain. We included one additional variable at a time to observe the effects. We next examined the receiver operating characteristics (ROC) curves for the various models. The best ROC curve which predicts data perfectly will touch the top-left corner of the plot (area/c-statistic 1.0), and the larger the area under the ROC curve the better the prediction. The best

performing model for overall back pain was pm7, with a c-statistic of 0.8345. The best performing model for severe back pain was the model excluding depression but including limiting activity and general health, with a c-statistic of 0.8672. These models were used to generate local estimates.

In the local estimates back pain prevalence varies widely depending on risk factor prevalence. The lowest prevalence of overall back pain is 4.18% for a male younger than 34 years with any BMI (excluding only above 30) who never smoked and has one of the following occupations: higher managerial/lower managerial/intermediate occupation/small employer and own account workers/routine occupations, or never worked. In contrast, the highest prevalence for overall back pain is 60.34% if a person is female aged over 75 being obese (BMI over 30), either current or ex-smoker, who has no education and is in routine occupation. The smallest prevalence for severe back pain is 1.21%. This is observed in males aged below 34, who never smoked with BMI under 18.4 or in the range of 18.4-25 with any education and are in either higher managerial/lower managerial/intermediate/small employer and own account worker occupation or never worked. The highest prevalence for severe back pain is similar to the maximal value for overall back pain – 63.82%. This prevalence is observed for obese females aged over 75 with no education, who are current smokers in routine occupations.

2 Background

Back pain is a very common condition and a major cause of disability. UK data from the 2016 Global Burden of Disease study shows that musculoskeletal (MSK) conditions cause the third largest loss of disability-adjusted life years (DALYs) after cardiovascular disease and cancer; and back pain causes the largest loss for any single disease.[5] Although there are no UK studies of overall consultation rates for back pain, in the US, it is one of the most common reasons for seeing a physician, with an estimated 440 million attendances per year, and it is also a frequent reason for lost working days and loss of worker productivity, estimated to be 149 million days per year.[6-8] Back or neck pain is the fifth most common reason for consulting a US physician, and results in $86 billion of health care spending annually. [9]

In the UK, because back pain has its highest incidence in midlife adults of working age, it has a similarly high impact on the economy.[10 11] In a UK retrospective cohort study of health care costs associated with the treatment of chronic, i.e. recurrent low back pain (lower back pain), total health care costs for patients were double those of the matched controls (£1074 vs. £516; $P < 0.05$).[1] Of the cost difference, 58.8% was accounted for by general practitioner's consultations, 22.3% by referrals to secondary care, and the rest by pain relief medications.

Nevertheless, most people who have acute lower back pain recover within months after the start of the disease, while others do not recover and progress to chronic lower back pain (lasting longer than 3 months).[7 8] A 2003 systematic review of lower back pain prognosis found rapid improvements in pain (mean reduction 58% of initial scores), disability (58%), and return to work (82% of those initially off work) occurred in one month.[12] Further improvement was apparent until about three months. Thereafter, levels for pain, disability, and return to work remained almost constant. Over 70% of patients had at least one recurrence within 12 months. A systematic review and meta-analysis of 41 studies found incidence rates for first-time lower back pain and transition to pain from a pain-free state were similar (~25%), regardless of community or occupational populations.[13] Risk factors for first-time lower back pain or transition to lower back pain from a baseline of a pain-free state were psychosocial and physically related.

Table 1: ARUK back pain risk factor list

Risk factor	References
Age	[8]
Biomechanical factors	[14]
Community size	[8]
Depression	[15-17]
Education	[8 18]
Gender	[6 8 19 20]
Obesity/BMI	[7 8 21-27]
Socioeconomic and Occupational class	[18]
Physical activity	[8 28-30]
Psychosocial factors	[14]
Sedentary lifestyle	[8 31]
Smoking	[8]

We were commissioned by Arthritis Research UK (ARUK) to develop a small population prevalence model for lower back pain, using nationally-representative data on risk factors applied to matching local population data. To initiate this ARUK provided the results of a review of risk factor research, which we supplemented with our own literature search. The following table of risk factors and

associated references was provided by ARUK (**Table 1**). We have summarised the evidence from both these sources in this section, by risk factor.

2.1 Risk factor – Obesity

It is still unclear whether obesity is a risk factor for lower back pain, as there are some studies that do not confirm the association between lower back pain and obesity. [7 8 26 27] The Cardiovascular Risk in Young Finns study investigated the association between weight related factors and lower back pain prevalence.[32] It suggested that abdominal obesity may increase risk of lower back pain in women. The MONICA population-based study found weak associations between body weight and lower back pain risk.[8]

2.2 Risk factor – Depression

A US study suggested that depressed people were three times more likely to develop chronic back pain within 2 years relative to non-depressed individuals.[7] Patients suffering from conditions with chronic pain are more likely to have some kind of psychiatric disorder.[16] Approximately 8-50% of patients with chronic pain have depression.[16 33] Moreover, depressed patients have a three times higher lower back pain prevalence compared to people in the general population.[16] A US cohort study examined the association between lifetime occurrence of depressive disorder and incident back pain over a 13-year period.[16] An association between lifetime back pain prevalence and occurrence of depressive disorder was observed at wave1 (OR=1.6; 95%CI 1.1-2.4). Moreover, throughout the whole follow-up period (13 years) higher risk for incident back pain was observed when depressive disorder was present at baseline (OR=1.9; 65%CI 1.2-3.1).

2.3 Risk factor – Sedentary lifestyle

A sedentary lifestyle is common, and is associated with obesity and therefore is linked to chronic health problems.[31] A systematic review of publications between 1998 and 2006 that examined the association between sedentary lifestyle and lower back pain did not find sufficient evidence to show that sedentary behaviour is a risk factor for development of lower back pain.[31] The review based its findings on analysis of 10 prospective cohorts and 5 case-control studies.[31] However, there was one high-quality study that showed a strong relationship between sitting and lower back pain among school children (OR=6.2, 95%CI 2.2-17.3). [31 34] The wide 95% confidence interval might indicate some uncertainty of the strength of association.[31] A Swedish population-based study found sedentary work was a risk factor.[8]

2.4 Risk factor – Physical activity

There are inconsistent findings regarding the effect of sports activities or other physical activities on lower back pain.[8 28-30] A Swedish population-based study showed a positive relationship between lower back pain and physical activity.[8] Individuals with physically demanding jobs but with light leisure physical activity were more likely to have lower back pain.

A wide variety of other risk factors have been identified in a variety of studies. Table 2 shows a fuller table of established risk factors and the scope of these.

Table 2 Risk factors and their ORs from various studies

Risk factor	Type of Odds Ratio	Odds Ratio	95% CI	Effect on Outcome
Age				
25-34	Unadjusted [8]	1.00		Reference
35-44	Unadjusted [8]	1.35	[1.07-1.70]	Risk factor
45-54	Unadjusted [8]	1.35	[1.08-1.70]	Risk factor
55-64	Unadjusted [8]	1.63	[1.31-2.04]	Risk factor
65-74	Unadjusted [8]	1.44	[1.15-1.81]	Risk factor
75-79	Unadjusted [8]	1.47	[1.04-2.09]	Risk factor
Gender				
Female	Unadjusted [8]	1.29	[1.17-1.44]	Risk factor
Education				
University	Unadjusted [8]	1.00		Reference
Secondary	Unadjusted [8]	1.42	[1.23-1.64]	Risk factor
Primary	Unadjusted [8]	1.50	[1.29-1.75]	Risk factor
Education (for females)				
University degree	Adjusted [18]	1.00		Reference
Professional qualification	Adjusted [18]	0.79	[0.50-1.26]	NS
A levels	Adjusted [18]	1.04	[0.61-1.52]	NS
O levels/CSE	Adjusted [18]	1.19	[0.82-1.71]	NS
None	Adjusted [18]	1.27	[0.89-1.81]	NS
Education (for males)				
University degree	Adjusted [18]	1.00		Reference
Professional qualification	Adjusted [18]	0.75	[0.31-1.84]	NS
A levels	Adjusted [18]	1.19	[0.85-1.68]	NS
O levels/CSE	Adjusted [18]	0.93	[0.67-1.29]	NS
None	Adjusted [18]	1.17	[0.87-1.57]	NS
Smoking				
Regular smoking	Unadjusted [8]	1.23	[1.07-1.41]	Risk factor
Smoking (for females)				
Non-smoker	Adjusted [18]	1.00		Reference
Current smoker	Adjusted [18]	1.38	[1.18-1.62]	Risk factor
Ex-smoker	Adjusted [18]	1.27	[1.06-1.51]	Risk factor
Smoking (for males)				
Non-smoker	Adjusted [18]	1.00		Reference
Current smoker	Adjusted [18]	1.29	[1.12-1.74]	Risk factor
Ex-smoker	Adjusted [18]	1.45	[1.15-1.83]	Risk factor
Social class (for females)				
Professional occupations	Adjusted [18]	1.00		Reference
Intermediate occupations	Adjusted [18]	0.90	[0.64-1.28]	NS
Skilled occupations: non-manual	Adjusted [18]	0.92	[0.64-1.34]	NS
Skilled occupations: manual	Adjusted [18]	1.13	[0.81-1.59]	NS
Partly skilled occupations	Adjusted [18]	1.20	[0.85-1.71]	NS
Unskilled occupations	Adjusted [18]	1.10	[0.70-1.66]	NS
Social class (for males)				
Professional occupations	Adjusted [18]	1.00		Reference
Intermediate occupations	Adjusted [18]	0.94	[0.61-1.47]	NS
Skilled occupations: non-manual	Adjusted [18]	0.99	[0.62-1.60]	NS

Risk factor	Type of Odds Ratio	Odds Ratio	95% CI	Effect on Outcome
Skilled occupations: manual	Adjusted [18]	1.13	[0.74-1.72]	NS
Partly skilled occupations	Adjusted [18]	1.03	[0.72-1.53]	NS
Unskilled occupations	Adjusted [18]	1.39	[0.83-2.33]	NS
Income (Sterling pounds per month for household) (for females)				
≤230	Adjusted [18]	1.44	[1.08-1.92]	Risk factor
231-415	Adjusted [18]	1.27	[0.97-1.67]	NS
416-750	Adjusted [18]	1.24	[0.94-1.63]	NS
751-995	Adjusted [18]	1.09	[0.71-1.39]	NS
≥996	Adjusted [18]	1.00		Reference
Income (Sterling pounds per month for household) (for males)				
≤230	Adjusted [18]	0.92	[0.64-1.33]	NS
231-415	Adjusted [18]	1.04	[0.78-1.40]	NS
416-750	Adjusted [18]	0.95	[0.72-1.26]	NS
751-995	Adjusted [18]	0.82	[0.58-1.15]	NS
≥996	Adjusted [18]	1.00		Reference
Obesity (for females)				
<25.0	Adjusted [6]	1.0		Reference
25.0-29.9	Adjusted [6]	1.2	[0.8-1.6]	NS
30.0-34.9	Adjusted [6]	1.4	[0.9-2.3]	NS
≥35	Adjusted [6]	3.1	[1.5-6.5]	Risk factor
≤21.0	Adjusted [18]	1.00		Reference
21.1-22.9	Adjusted [18]	1.16	[0.94-1.44]	NS
23.0-24.6	Adjusted [18]	1.18	[0.95-1.47]	NS
24.7-27.3	Adjusted [18]	1.36	[1.09-1.70]	Risk factor
>27.3	Adjusted [18]	1.45	[1.16-1.83]	Risk factor
Obesity (for males)				
<25.0	Adjusted [6]	1.0		Reference
25.0-29.9	Adjusted [6]	0.8	[0.6-1.1]	NS
30.0-34.9	Adjusted [6]	0.9	[0.9-1.4]	NS
≥35	Adjusted [6]	0.5	[0.2-1.4]	NS
≤21.0	Adjusted [18]	1.00		
21.1-22.9	Adjusted [18]	1.27	[0.93-1.72]	Reference
23.0-24.6	Adjusted [18]	1.16	[0.86-1.56]	NS
24.7-27.3	Adjusted [18]	1.18	[0.83-1.50]	NS
>27.3	Adjusted [18]	1.25	[0.92-1.71]	NS
Obesity				
≥25	Unadjusted [8]	1.28	[1.15-1.43]	Risk factor
Sedentary lifestyle	[34]	6.2	[2.2-17.3]	Risk Factor
Physical activity at work in the last year				
Sitting work	Unadjusted [8]	1.0		Reference
Light physical work	Unadjusted [8]	1.11	[0.93-1.31]	NS
Moderate heavy work	Unadjusted [8]	1.45	[1.21-1.73]	Risk Factor
Heavy work	Unadjusted [8]	1.44	[1.09-1.90]	Risk Factor
Physical activity at work in the last year				

Risk factor	Type of Odds Ratio	Odds Ratio	95% CI	Effect on Outcome
Sitting work	Adjusted [8]	1.00		Reference
Light physical work	Adjusted [8]	1.13	[0.95-1.35]	NS
Moderate heavy work	Adjusted [8]	1.37	[1.14-1.65]	Risk factor
Heavy work	Adjusted [8]	1.46	[1.09-1.94]	Risk factor
Occurrence of physically demanding work				
No never	Unadjusted [8]	1.0		Reference
No rarely	Unadjusted [8]	1.04	[0.86-1.26]	NS
Yes sometimes	Unadjusted [8]	1.31	[1.10-1.57]	Risk Factor
Yes often	Unadjusted [8]	1.97	[1.59-2.45]	Risk Factor
Low physical activity during leisure time in the last year	Unadjusted [8]	1.35	[1.19-1.53]	Risk factor
Occurrence of physically demanding work				
No never	Unadjusted [8]	1.00		Reference
No rarely	Unadjusted [8]	1.03	[0.84-1.25]	NS
Yes sometimes	Unadjusted [8]	1.22	[1.01-1.48]	Risk factor
Yes often	Unadjusted [8]	1.77	[1.41-2.22]	Risk factor
Low physical activity during leisure time in the last year	Unadjusted [8]	1.16	[1.02-1.33]	Risk factor
Community size				
>15,000 inhabitants	Unadjusted [8]	1.00		Reference
1,000-15,000	Unadjusted [8]	1.09	[0.96-1.24]	NS
<1,000	Unadjusted [8]	1.29	[1.13-1.46]	Risk factor

2.5 Back pain prevalence from the literature

A cross-sectional telephone survey was conducted in 1992 and repeated in 2006 using 4,437 and 5,357 households respectively.[7] This survey identified adults 21 years and older with chronic (lasting more than 3 months), impairing lower back pain or neck pain that resulted in restricting daily activities. The prevalence estimate increased from 3.9% in 1992 to 10.2% in 2006. As a consequence of the increase in the prevalence, a higher proportion of individuals sought care (increased from 73.1% to 84.0%), resulting in a substantial rise in lower back pain costs.[7] The study by Harkness et al. compared lower back pain, shoulder and widespread pain prevalence in the northwest region of England using two surveys (1956-58 and 1994-95, respectively). [11] The prevalence increased for all three symptoms from 2-fold to 4-fold between the two surveys.

Similarly, Palmer et al. compared two surveys over a 10 year period showed an increase in back pain prevalence from 36.4% to 49.1%.[10] This increase might be explained by increasing rates of psychological distress, increased reporting and/or increased awareness.[11] Various prevalence rates are shown from different surveys: 12-33% point prevalence, 22-65% 1-year prevalence and 11-84% life-time prevalence. [8] Lower back pain prevalence is higher in females than males in some studies, while others show quite equal distribution between two genders.[6 8 19 20]

Table 3 Prevalence estimates of chronic low back pain from literature

Publication[1]	Country	Type of study	Total*	Male*	Female*	Population age (years)
Freburger, et al. [7]	North Carolina, USA	Cross-sectional telephone survey	3.9% [3.4-4.4] in 1992 10.2% [9.3-11.0] in 2006	2.9% [2.2-3.6] in 1992 8.0% [6.8-9.2] in 2006	4.8% [4.0-5.6] in 1992 12.2% [10.9-13.5] in 2006	≥21
Bjorck-van Dijken, et al. [8]	Sweden	MONICA health survey	41%	37.8%	44.1%	25-79
Croft and Rigby [18]	United Kingdom	Cross-sectional health and lifestyle survey 1984-85	18.9%	16.2%	20.9%	≥18

Table 4 Prevalence of chronic low back pain by age and sex (shown as percentages)

Gender	Female					Male					Both sexes				
Age group	**21-34**	**35-44**	**45-54**	**55-64**	**≥65**	**21-34**	**35-44**	**45-54**	**55-64**	**≥65**	**21-34**	**35-44**	**45-54**	**55-64**	**≥65**
Freburger, et al. [7] (1992)	1.2	6.1	5.8	6.7	7.3	1.6	3.4	2.6	5.7	3.7	1.4	4.8	4.2	6.3	5.9
Freburger, et al. [7] (2006)	5.1	11.9	16.5	16.9	14.3	3.5	6.5	10.3	13.7	9.7	4.3	9.2	13.5	15.4	12.3
Harkness, et al. [11] 1956-58[2]	5.7	9.8	13.2	7.5	-	1.7	8.6	15.2	7.1	-	-	-	-	-	-
_ENREF_4Harkness, et al. [11] 1994-95[3]	14.8	12.9	24.6	24.7	-	15.7	14.1	22.1	22.9	-	-	-	-	-	-
_ENREF_3Palmer, et al. [104] 1987-8	25.9	33.1	44.8	34.6	-	34.9	38.1	37.0	40.2	-	-	-	-	-	-
Palmer, et al. [105] 1997-8	38.5	41.0	48.0	51.2	-	46.5	52.4	56.4	56.6	-	-	-	-	-	-

[1] Figures in brackets are 95% confidence intervals
[2] 25-34 age category instead of 21-34
[3] 25-34 age category instead of 21-34
[4] Age categories: 20-29, 30-39, 40-49, 50-59
[5] Age categories: 20-29, 30-39, 40-49, 50-59

3 Methods

3.1 Data source: Health Survey for England (2011)

This section describes the methods we used to fit a regression model to Health Survey for England (HSfE) 2011 data,[2] and to produce from that local estimates of lower back pain. The main focus of the HSfE in 2011 was cardiovascular disease. The survey also provided updates on core topics including smoking, drinking and fruit and vegetable consumption. Additional modules of questions were also included covering social care, chronic pain and well-being. In 2011, there was also a drinking diary designed to measure weekly consumption of alcohol.

The HSfE 2011 included a general population sample of adults and children, representative of the whole population at both national and regional level. For the sample, 8,992 addresses were randomly selected in 562 postcode sectors, issued over twelve months from January to December 2011. Where an address was found to have multiple dwelling units, one dwelling unit was selected at random and where there were multiple households at a dwelling unit, one household was selected at random. In each selected household, all individuals were eligible for inclusion in the survey. Where there were three or more children aged 0-15 in a household, two of the children were selected at random. A nurse visit was arranged for all participants who consented. A total of 8,610 adults aged 16 and over and 2,007 children aged 0-15 were interviewed. A household response rate of 66% was achieved for the core sample. Among the general population sample, 5,715 adults and 1,257 children had a nurse visit.

3.2 Outcome measures

The HSfE 2011 questionnaire contains a number of questions on chronic pain. The interviewer asked if the interviewee is currently troubled by pain (AnyPain), and if so, has it been for more than 3 months (More3m), and if that is the case, then the location of the pain (SitePain). The question includes "back pain" and "neck or shoulder pain"; in other words, "back pain" is likely to be interpreted as lower back **or** thoracic pain. Hence data from HSfE is likely to over-estimate the prevalence of lower back pain. One of the possible responses is back pain. The level of pain is rated by the interviewee in further questions (PainNow, WorstP, and UsualP). Illsm* variables also allow the recording of 'Back problems/slipped disc/spine/neck' with a value of '35'. (Respondents over 16 were asked these questions). The details of HSfE variable descriptions are shown in Section 6 Appendix 1: coding details of Health Survey for England variables.

Based on the answers to the three questions above a new variable **backpain** was generated:

- 0 "No Pain" (if More3m had label 2 OR -1 OR AnyPain had label 2)
- 1 "Pain" (if SitePai1 had label 1)
- Missing (if it was not coded as 'No pain' or 'Pain')

Out of 10,617 respondents, 1,346 (12.68%) experienced back pain, while 7,415 (69.84%) did not experience back pain and 1,856 (17.48%) did not have information regarding back pain. Note that there is a derived variable CPain that captures absence/presence of chronic pain based on the questions asked. This variable shows that there are 3,202 (30.16%) respondents with chronic pain and 5,397 (50.83%) do not experience chronic pain; with 2,007 (18.90%) missing. Therefore, this checked how many respondents have chronic pain and back pain – 1,346, which is the same as identified by using newly created backpain variable. The details of HSfE variable descriptions are shown in Section 6 Appendix 1: coding details of Health Survey for England variables.

There are several validated instruments which measure the severity of chronic pain. The Graded Chronic Pain Scale (GCPS), also known as the "von Korff"score after its developer, is a well-validated

measure of pain intensity and pain-related disability related to a 6-month reference frame.[3] The GCPS and the 12-Item Short-Form Health Survey (SF-12) are both widely used survey instruments that collect data on both mental and physical aspects of pain disorders – in both the telephone interview mode and the self-administration mode. Table 5 shows the pain grading based on GCPS Version 2.0.

Table 5 Back pain grading (based on GCPS Version 2.0)

Chronic pain grade based on GCPS version 2.0	Respondents with No Back Pain	Respondents with Back Pain	All respondents
Not applicable	2,018 (27.22%)	0	2,018 (23.03%)
Grade 0	5,397 (72.78%)	0	5,397 (61.60%)
Grade I – Low intensity	0	450 (33.43%)	450 (5.14%)
Grade II – High intensity	0	343 (25.48%)	343 (3.92%)
Grade III – Moderately limiting	0	198 (14.71%)	198 (2.26%)
Grade IV – Severely limiting	0	342 (24.41)	342 (3.90%)
Total	7,415 (100%)	1,346 (100%)	8,761 (100%)

The Graded Chronic Pain-Primary Care Scale (GCP-PCS) was developed to assess three components of catastrophizing: rumination, magnification, and helplessness.[35] Pain catastrophizing is significantly related to pain outcomes in patients with rheumatoid arthritis, osteoarthritis and neuropathic pain, with higher scores predicting chronicity. The GCP-PCS can be used to obtain measures of chronic pain intensity, interference with activities and pain duration, as well as an ordinal classification of chronic pain severity.

Table 6 Back pain grading (based on 3-item GCP-PCS)

Chronic pain grade based on 3-item GCP-PCS	Respondents with No Back Pain	Respondents with Back Pain	All respondents
Not applicable	2,018 (27.22%)	0	2,018 (23.03%)
Grade 0	5,397 (72.78%)	0	5,397 (61.60%)
Grade I – Low intensity	0	37 (2.75%)	37 (0.42%)
Grade II – High intensity	0	759 (56.39%)	759 (8.66%)
Grade III – Moderately limiting	0	150 (11.14%)	150 (1.71%)
Grade IV – Severely limiting	0	391 (29.05%)	391 (4.46%)
Total	7,415 (100%)	1,346 (100%)	8,761 (100%)

Table 7 explains the differences between the two different chronic pain grading methods. **Table 8** shows the prevalence stratified by age categories and sex (number of respondents and their proportion HSfE 2011 is presented). HSfE 2011 back pain prevalence rates are compared with the estimates published in Croft and Rigby paper. We chose **PainGrade1** (containing the four-item disability code) as it is the most commonly used. PainGrade1 was renamed painseverity, where 'Not applicable' and '-5' were given missing values (n=2,031). After further discussions and statistical considerations of various ways to categorise back pain severity, it was agreed to create a binary variable that would code for no severe back pain versus severe back pain.

Table 7: Explanations of two different chronic pain grading methods

CHRONIC PAIN GRADE based on GCPS Version 2.0	
PAIN FREE	
GRADE 0	No pain problem (prior 3 months)
GRADE I—Low intensity, Low interference	Characteristic Pain Intensity less than 15 *and* four-item Disability Score less than 17
GRADE II—High intensity	Characteristic Pain Intensity of 15 or greater *and* four-item Disability Score less than 17
GRADE III—Moderate interference	Four-item Disability Score of 17 to 24
GRADE IV—Severe interference	Four-item Disability Score of 25 to 40

CHRONIC PAIN GRADE based on three-item GCP-PCS	
PAIN FREE	
GRADE 0	No pain problem (prior 3 months)
GRADE I—Low intensity, Low interference	Usual Pain Intensity less than 5 *and* two-item Disability Score less than 9
GRADE II—High intensity	**Characteristic Pain Intensity of 5 or greater** *and* two-item Disability Score less than 9
GRADE III—Moderate interference	Two-item Disability Score of 9 to 12
GRADE IV—Severe interference	Two-item Disability Score of 13 to 20

There are two possibilities for creating the back pain severity variable:
1. generate **severity1**, where:
   - 0 "No severe back pain" (if PainGrade1 had a label either 0 **OR** 1 **OR** 3)
   - 1 "Severe back pain" (If PainGrade1 had a label either 2 **OR** 4)

   Out of 8,761 respondents 685 (7.82%) experienced severe back pain, while 6,045 (69.00%) did not experience severe back pain and 2,031 (23.18%) did not have information regarding the severity of back pain (using severity1).

2. generate **severity2**, where:
   - 0 "No severe back pain" (if PainGrade1 had a label either 0 **OR** 1)
   - 1 "Severe back pain" (if PainGrade1 had a label either 2 **OR** 3 **OR** 4)

   Out of 8,761 respondents 883 (10.08%) experienced severe back pain, while 5,847 (66.74%) did not experience severe back pain and 2,031 (23.18%) did not have information regarding the severity of back pain (using severity2).

Table 8 and Table: 9 show the prevalence of back pain as defined, stratified by age and sex, and pain severity stratified by age and sex from GCPS Version 2.0.

Table 8 Prevalence stratified by age and sex (N and %)

Gender	Female							Male							Both sexes						
Age group	**<25**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**≥75**	**<25**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**≥75**	**<25**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**≥75**
HSfE 2011	18 1.3%	67 10.4%	123 17.6%	164 25.7%	163 30.5%	154 35.8%	157 42.0%	9 0.7%	32 6.7%	81 14.5%	80 15.2%	114 24.7%	89 25.3%	95 32.5%	27 1.0%	99 8.8%	204 16.2%	244 21.0%	277 27.8%	243 31.1%	252 37.8%
Croft and Rigby [18]	15.7%	15.5%	18.5%	23.5%	22.9%	27.9%	31.1%	11.8%	13.8%	16.5%	18.2%	18.3%	17.4%	19.5%	13.8%	14.8%	17.6%	21.1%	20.9%	23.3%	26.4%

Table: 9 Back pain severity stratified by age and sex (N and %) (GCPS Version 2.0)

(D) Chronic Pain Grade based on GCPS (version 2.0)	Female								Male							
Age groups	<=25	25-34	35-44	45-54	55-64	65-74	=>75	Total	<=25	25-34	35-44	45-54	55-64	65-74	=>75	Total
Grade 0	400	576	576	474	370	275	216	2,887	336	447	475	446	347	263	196	2,510
	95.69%	89.58%	82.40%	74.29%	69.68%	64.10%	58.54%	77.46%	97.67%	93.51%	85.59%	84.79%	75.27%	74.72%	68.29%	83.58%
Grade I - Low intensity	5	29	45	56	48	38	31	252	3	16	39	35	42	37	26	198
	1.20%	4.51%	6.44%	8.78%	9.04%	8.86%	8.40%	6.76%	0.87%	3.35%	7.03%	6.65%	9.11%	10.51%	9.06%	6.59%
Grade II - High intensity	9	18	36	41	37	54	41	236	5	5	16	22	21	14	24	107
	2.15%	2.80%	5.15%	6.43%	6.97%	12.59%	11.11%	6.33%	1.45%	1.05%	2.88%	4.18%	4.56%	3.98%	8.36%	3.56%
Grade III - Moderately limiting	2	8	17	28	28	24	29	136	0	6	14	9	15	9	9	62
	0.48%	1.24%	2.43%	4.39%	5.27%	5.59%	7.86%	3.65%	0%	1.26%	2.52%	1.71%	3.25%	2.56%	3.14%	2.06%
Grade IV - Severely limiting	2	12	25	39	48	38	52	216	0	4	11	14	36	29	32	126
	0.48%	1.87%	3.58%	6.11%	9.04%	8.86%	14.09%	5.80%	0%	0.84%	1.98%	2.66%	7.81%	8.24%	11.15%	4.20%
Total	418	643	699	638	531	429	369	3,727	344	478	555	526	461	352	287	3,003
	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

3.3 Risk factors in HSfE 2011

Table 10 shows the list of risk factors related to back pain and the corresponding variables in HSfE 2011 survey data. The detailed coding for these variables is shown in Section 6: Appendix 1: coding details of Health Survey for England variables.

Table 10: list of risk factors in HSfE 2011 data

Risk Factor	Appropriate variable(s) in HSfE 2011	Selected variable for analysis
Age	Age	agegrp (derived from Age)
BMI	bmiok, bmi, bmival, bmivg5, bmicat1, bmicat2, bmicat3	bmigrp (renamed bmivg5)
Economic status	econact econact5 econact2	econact
Education	topqual2, topqual3	educ (renamed topqual3)
Ethnicity	Origin	ethn (derived from Origin)
Depression	Anxiety (combines both anxiety and depression)	anxiety (this variable had to be excluded from the model as it had no local data)
Gender	Sex	gender (renamed Sex)
Physical activity	Not available (2008 and 2012) In 2008 active wrkact24 hswrkhm hwrklsth workact workactg (merged 2008 and 2011 – only 267 respondents with back pain recorded pa: active and wrkact24 variables checked)	Not available in 2011
Sedentary lifestyle	Not available In 2008 there are questions that ask how many hours you are watching TV, sitting etc...	Not available in 2011
Socioeconomic status	STHNSSEC HRPNSSEC STNSSEC NSSEC nssec8 nssec5 nssec3 hpnssec8 hpnssec5 hpnssec3	nssec8
Smoking	StartSmk cigst1 cigsta3 cigst2 smkevr cignow cigevr cigreg	smoke (renamed cigsta3)
Extra variables:		
Activities limited due to illness (binary answer)	LimitAct LongIll	limitact (derived from LimitAct and LongIll)
Self-assessed general health	GenHelf genhelf2 (grouped)	genhealth (renamed GenHelf)

3.3.1 Depression variable (derived)

If a respondent indicated having longstanding illness (LongIll), then they were asked which illness they had (IllsM1-6). A value 4 of IllsM1-6 variables is coded for 'Mental illness/anxiety/depression". Depression is by far the commonest mental illness, so it was assumed that the majority of cases were related to depression therefore a '**depression'** variable was generated:

- 0 ("No depression") if LongIll equal to 1 or 2

- 1 ("Depression") if IllsM1 or Illsm2 or IllsM3 or IllsM4 or Illsm5 or IllsM6 equal to 4. A total of 6,847 (78.15%) of respondents did not have depression based on the newly derived depression variable, while 260 (2.97%) indicated having depression and 1,654 (18.88%) did not answer this question. Of these, 87 (33.46%) had depression and back pain, whereas 173 (66.54%) had depression but did not indicate to have back pain as well.

Statistical analyses were run using this depression variable but the overall back pain model with it (M27) performed worse compared to the model without this variable (M26B), while the severe back pain model with this variable performed similarly (6) to the model without this variable (5B) (see statistical analysis section). Therefore, the depression (derived) variable was not used in the final back pain models.

3.4 Comorbidities

3.4.1 Binary variable

The comorbidities variable was generated based on the information found in the HSfE 2011 data dictionary. The comorbidities section starts with a question whether a respondent has a longstanding illness. Therefore the comorbidity variable (**comorbid**) was generated using compm1_17 variables.

comorbid:

- 0 "No comorbidity" (if compm17 had label 1)
- 1 "Comorbidity present" (if compm1_15 had label 1)
- Missing (if it was not coded as 'No comorbidity' or 'Comorbidity')

More detail of coding is shown in Secion 6, Appendix 1: coding details of Health Survey for England variables.

There are 2,576 (29.40%) respondents with at least one comorbidity, 4,512 (51.50%) with no comorbidity, while 1,673 (19.10%) did not have this information. However, this variable was given label '1' if any compm1_15 indicated having a condition (even compm12 'Musculoskeletal diseases'). Comorbidity that was as a result of back pain, needed to be excluded from the comorbidity variable. Firstly, it was checked how compm12 variable was created; it was given values 34, 35 and 36 from variables IllsM1_6 that coded for:

Value = 34 Label = Arthritis/rheumatism/fibrositis
Value = 35 Label = Back problems/slipped disc/spine/neck
Value = 36 Label = Other problems of bones/joints/muscles

There were 450 respondents who indicated having Arthritis/rheumatism/fibrositis, 408 – Back problems/slipped disc/spine/neck, 297 – Other problems of bones/joints/muscles. A total of 408 respondents who indicated having back problems needed to be excluded from the comorbidity variable. Therefore, a new variable comorbid2 was generated:

- 0 "No comorbidity" (if compm17 had label 1)
- 1 "Comorbidity present" (if compm1_15 had label 1 **EXCLUDED if IllsM1_6 was equal to 35**)
- Missing (if it was not coded as 'No comorbidity' or 'Comorbidity')

As a result, there were 2,168 (24.75%) of respondents who indicated having a comorbidity, not related to back pain (IllsM1_6 value label 35), 4,920 (56.16%) had no comorbidity and 1,673 (19.10%) did not have an answer for these questions.

3.5 Statistical analysis

We fitted univariate then multivariate logistic regression models for overall and severe back pain as described in previous publications.[4] A range of multivariate regression models were fitted in order

to obtain the best performing. We included one additional variable at a time to observe the effects: the variance inflation factor (vif) command was used after model fitting to check for multicollinearity.

We looked for interactions between HSfE predictor variables for risk factors. There is an interaction between the effects of two exposures if the effect of one exposure varies according to the level of the other exposure.[36] For example, the protective effect of breastfeeding against infectious diseases in early infancy is more pronounced among infants living in poor environmental conditions than among those living in areas with adequate water supply and sanitation facilities. An alternative term for interaction is 'effect modification'. In this example, we can think of this as the quality of environmental conditions modifying the effect of breastfeeding. The most flexible approach to examine interactions is to use regression models, but when using Mantel-Haenszel methods to control for confounding, an alternative is to use a $\chi 2$ test for effect modification, commonly called a test of heterogeneity. Interaction, effect modification and heterogeneity are three different ways of describing the same thing. Log likelihoods are compared in the two models excluding and including the interaction parameters to test the null hypothesis that there is no interaction between selected variables.

In order to obtain the most parsimonious models, we then applied stepwise backward and forward variable selection using the *stepwise* command in Stata. Finally, we internally validated the models by generating receiver operating characteristic (ROC) curves, by using the ***predict*** regression post-estimation command to generate for each ELSA respondent the probability of having back pain, and by using these probabilities to examine sensitivity and specificity. All statistical analysis was carried out in StataSE12.

However, the variables included in the final model are also determined by the availability of local data to match with the model variables. Hence, variable selection has to be a compromise between the best model which can be produced from the ELSA data and the local variable available.

3.6 Estimating local prevalence

The local population estimates and their CIs are also produced in Stata. In summary, the proportion of our population according to age and sex are known. The proportion by educational status can be applied to these numbers, taking account of the fact that the distribution by educational status differs by age group. This gives estimated proportion by age, sex and educational status. This information is reflected in the variables names.

Within Stata, a new set of variables is created, one for each combination of these risk factors pertinent to the logistic regression model for the chosen disease. For instance, if there are two binary variables for age group included in the regression model, then there are three relevant age groups (those with the first variable=1, those with the second variable=1, and those where both variables=0 – it is not possible to have both variables =1 since this would imply being in two separate age groups at the same time). With a binary variable for gender included, we would need groups for each gender, and so on.

In practice, we do not want to find a synthetic estimate on just one population, but rather on many populations, for instance on each local authority separately. We have a dataset containing information on the risk factors in all the different local authorities (LAs) and also other regions, with one line of data per region. The above variables give the proportions for each specified combination of age/ sex/ education categories. There are other variables giving the proportions by each additional risk factor separately (e.g. the proportion of non-smokers, current smokers and ex-smokers).

There is uncertainty in these synthetic estimates of prevalence based on the imprecision in the estimated coefficients from the logistic regression equations. A bootstrap procedure can be used to

construct confidence intervals on these synthetic estimates of prevalence, based on the imprecision in these logistic regression coefficients. A detailed description of the steps involved can be found in Section 8 Appendix 3: synthetic estimation using Stata

4 Results

4.1 Baseline characteristics

Table 11 below shows the baseline characteristics of the HSfE respondents.

Table 11: baseline characteristics

	Back pain cases	Non-back pain cases	Total
Total number of respondents	1,346 (15.36%)	7,415 (84.64%)	8,761 **(1,856 out of 10,617 excluded as missing)**
Age (agegrp)[6]			
<=25	27 (2.01%)	2,746 (37.03%)	2,773 (31.65%)
25-34	99 (7.36%)	1,023 (13.80%)	1,122 (12.81%)
35-44	204 (15.16%)	1,055 (14.23%)	1,259 (14.37%)
45-54	244 (18.13%)	920 (12.41%)	1,164 (13.29%)
55-64	277 (20.58%)	718 (9.68%)	995 (11.36%)
65-74	243 (18.05%)	539 (7.27%)	782 (8.93%)
Over 75	252 (18.72%)	414 (5.58%)	666 (7.60%)
Age (agegrp2)			
<=34	126 (9.36%)	3,769 (50.83%)	3,895 (44.46%)
35-44	204 (15.16%)	1,055 (14.23%)	1,259 (14.37%)
45-54	244 (18.13%)	920 (12.41%)	1,164 (13.29%)
55-64	277 (20.58%)	718 (9.68%)	995 (11.36%)
65-74	243 (18.05%)	539 (7.27%)	782 (8.93%)
Over 75	252 (18.72%)	414 (5.58%)	666 (7.60%)
Gender			
Female	846 (62.85%)	3,870 (52.19%)	4,716 (53.83%)
Male	500 (37.15%)	3,545 (47.81%)	4,045 (46.17%)
Ethnicity			
White	1,237 (91.90%)	6,318 (85.21%)	7,555 (86.23%)
Mixed	8 (0.59%)	186 (2.51%)	194 (2.21%)
Asian	63 (4.68%)	571 (7.70%)	634 (7.24%)
Black/Black Caribbean	21 (1.56%)	272 (3.67%)	293 (3.34%)
Other	14 (1.04%)	38 (0.51%)	52 (0.59%)
Not stated	3 (0.22%)	30 (0.40%)	33 (0.38%)
Education			
NVQ4/NVQ5/Degree or equivalent	243 (18.05%)	1,428 (19.26%)	1,671 (19.07%)
Higher education below degree	141 (10.48%)	605 (8.16%)	746 (8.52%)
NVQ3/GCE A level equivalent	139 (10.33%)	902 (12.16%)	1,041 (11.88%)
NVQ2/GCE O level equivalent	257 (19.09%)	1,149 (15.50%)	1,406 (16.05%)
NVQ1/CSE other grade equivalent	56 (4.16%)	245 (3.30%)	301 (3.44%)
Foreign/other	31 (2.30%)	61 (0.82%)	92 (1.05%)
No qualification	476 (35.36%)	988 (13.32%)	1,464 (16.71%)
Missing	3 (0.22%)	2,037 (27.47%)	2,040 (23.29%)

[6] This variable was not used in the analyses

	Back pain cases	Non-back pain cases	Total
Socioeconomic status			
Higher managerial and professional	115 (8.54%)	720 (9.71%)	835 (9.53%)
Lower managerial and professional	266 (19.76%)	1,210 (16.32%)	1,476 (16.85%)
Intermediate occupations	204 (15.16%)	797 (10.75%)	1,001 (11.43%)
Small employers and own account workers	107 (7.95%)	399 (5.38%)	506 (5.78%)
Lower supervisory and technical	103 (7.65%)	325 (4.38%)	428 (4.89%)
Semi-routine occ.	249 (18.50%)	857 (11.56%)	1,106 (12.62%)
Routine occ.	244 (18.13%)	654 (8.82%)	898 (10.25%)
Never worked and long term unemployed	19 (1.41%)	112 (1.51%)	131 (1.50%)
Other	7 (0.52%)	216 (2.91%)	223 (2.55%)
Missing	32 (2.38%)	2,125 (28.66%)	2,157 (24.62%)
BMI (bmigrp)[7]			
<18.4 underweight	6 (0.45%)	80 (1.08%)	86 (0.98%)
18.5 – 24 normal	296 (21.99%)	1,728 (23.30%)	2,024 (23.10%)
25 – 29 overweight	398 (29.57%)	1,692 (22.82%)	2,090 (23.86%)
30 – 39 obese	317 (23.55%)	889 (11.99%)	1,206 (13.77%)
>40 obese	57 (4.23%)	85 (1.15%)	142 (1.62%)
Missing	272 (20.21%)	2,941 (39.66%)	3,213 (36.67%)
BMI (bmigrp2)			
<=24 normal (+underweight)	302 (22.44%)	1,808 (24.38%)	2,110 (24.08%)
25 – 29 overweight	398 (29.57%)	1,692 (22.82%)	2,090 (23.86%)
30 – 39 obese	317 (23.55%)	889 (11.99%)	1,206 (13.77%)
>40 obese	57 (4.23%)	85 (1.15%)	142 (1.62%)
Missing	272 (20.21%)	2,941 (39.66%)	3,213 (36.67%)
Smoking			
Current smoker	284 (21.10%)	1,067 (14.39%)	1,351 (15.42%)
Ex-regular smoker	448 (33.28%)	1,260 (16.99%)	1,708 (19.50%)
Never regular	611 (45.39%)	3,018 (40.70%)	3,629 (41.42%)
Missing	3 (0.22%)	2,070 (27.92%)	2,073 (23.66%)
Anxiety/depression			
Not anxious or depressed	669 (49.70%)	3,816 (51.46%)	4,485 (51.19%)
Moderately anxious or depressed	460 (34.18%)	860 (11.60%)	1,320 (15.07%)
Extremely anxious or depressed	64 (4.75%)	76 (1.02%)	140 (1.60%)
Missing	153 (11.37%)	2,663 (35.91%)	2,816 (32.14%)
Economic status			
In employment	533 (39.60%)	3,247 (43.79%)	3,780 (43.15%)
ILO Unemployed	41 (3.05%)	297 (4.01%)	338 (3.86%)
Retired	532 (39.52%)	1,055 (14.23%)	1,587 (18.11%)
Other economically inactive	236 (17.53%)	778 (10.49%)	1,014 (11.57%)
Missing	4 (0.30%)	2,038 (27.48%)	2,042 (23.31%)
Activities limited due to illness			
Yes	807 (59.96%)	546 (7.36%)	1,353 (15.44%)
No	539 (40.04%)	5,215 (70.33%)	5,754 (65.68%)
Missing	0	1,654 (22.31%)	1,654 (18.88%)

[7] This variable was not used in the analyses

	Back pain cases	Non-back pain cases	Total
Self-assessed general health			
Very good	173 (12.85%)	3,500 (47.20%)	3,673 (41.92%)
Good	442 (32.84%)	3,076 (41.48%)	3,518 (40.16%)
Fair	415 (30.83%)	734 (9.90%)	1,149 (13.11%)
Bad	216 (16.05%)	70 (0.94%)	286 (3.26%)
Very bad	100 (7.43%)	29 (0.39%)	129 (1.47%)
Missing	0	6 (0.08%)	6 (0.07%)

4.2 Comorbidities

Table 12 Comorbidities by back pain presence (also excluding those comorbidities related to back problems)

	Comorbidity			Comorbidity (excluding related to back problems)		
	No back pain	Back pain	Total	No back pain	Back pain	Total
No comorbidity	4,170 (56.24%)	342 (25.41%)	4,512 (51.50%)	4,215 (56.84%)	705 (52.38%)	4,920 (56.16%)
Comorbidity	1,577 (21.27%)	999 (74.22%)	2,576 (29.40%)	1,532 (20.66%)	636 (47.25%)	2,168 (24.75%)
Missing	1,668 (22.49%)	5 (0.37%)	1,673 (19.10%)	1,668 (22.49%)	5 (0.37%)	1,673 (19.10%)

4.2.1 Categorical variable

The *comorbidity count* variable (**comorbidcount2**) was generated using the newly created comorbid variable:

- 0 "No comorbidity" (if comorbid2 was equal to 0 OR IlsM1_6 was equal to 35)
- Number of comorbidities "Number of comorbidities present" (sum of compm1_15 variables)
- Missing (if it was not coded as 'No comorbidity' or 'Comorbidity')

The results are as follows:

Comorbidity count	Derived		Comorbidity (present in HSfE2011)	
	Frequency	Percentage	Frequency	Percentage
Not applicable	1,654	18.8	1,654	18.88
No comorbidity	4,531	51.72	4,512	51.50
1	1,519	17.34	1,539	17.57
2	668	7.62	668	7.62
3	249	2.84	248	2.83
4	107	1.22	107	1.22
5	30	0.34	30	0.34
6	3	0.03	3	0.03

The derived variable comorbidcount2 and condcnt produced almost identical results (with minor differences between no comorbidity and presence of one comorbidity, which might be explained by excluding value 35 of IllsM1_6 variables in our created variable).

4.2.2 Comorbidities and activities limited due to illness (limitact)

There are 1,353 respondents (15.44%) who indicated having activities limited due to illness, 5,754 (65.68%) had no limitations, and 1,654 (18.88%) did not have this information. A new ***limitact2*** variable was generated based on the values of ***limitact***:

- 1 "Yes" (if limitact was equal to 1)
- 2 "No" (if comorbid2 was equal to 0)
- Missing

In this scenario, activities that were limited due to back pain (value 35 of IllsM1_6) were coded as 'No'. There are 1,031 (11.77%) that indicated having activities limited due to illness (excluded if related to back pain), 6,076 (69.35%) had no limitations, and 1,654 (18.88%) did not have this information (322 respondents were coded as not having limited activities due to illness in limitact2 compared to limitact).

Table 13 Relation between different comorbidity variables and different limited activity variables

	Activities limited due to illness				Activities limited due to illness (excluding related to back problems)			
	Yes	**No**	**Missing**	**Total**	**Yes**	**No**	**Missing**	**Total**
No comorbidity	0	4,512 (78.42%)	0	4,512 (51.50%)	0	4,512 (74.26%)	0	4,512 (51.50%)
Comorbidity	1,343 (99.26%)	1,233 (21.43%)	0	2,576 (29.40%)	1,021 (99.03%)	1,555 (25.59%)	0	2,576 (29.40%)
Missing	10 (0.74%)	9 (0.16%)	1,654 (100%)	1,673 (19.10%)	10 (0.97%)	9 (0.15%)	1,654 (100%)	1,673 (19.10%)
No comorbidity (2)	322 (23.80%)	4,598 (79.91%)	0	4,920 (56.16%)	0	4,920 (80.97%)	0	4,920 (56.16%)
Comorbidity (2)	1,021 (75.46%)	1,147 (19.93%)	0	2,168 (24.75%)	1,021 (99.03%)	1,147 (18.88%)	0	2,168 (24.75%)

Table 14 shows the relation between comorbidity, limited activities and back pain.

4.2.3 Comorbidities and BMI

As observed in Table 15, there are 86 (0.98%) individuals with BMI lower than 18.4, meaning that they are underweight (6 of these respondents had back pain, while 80 did not have it). After deriving the comorbid and comorbidcountcat variables, we checked whether lower BMI might have resulted due to comorbidities. The majority of people with back pain indicated having a comorbidity/ies, while a much smaller proportion of people without back pain indicated having comorbidity. Six cases (100% of the 'Under 18.4' group) had comorbidity, therefore, the lower BMI might have resulted from comorbidity. However, only 22.50% of respondents with no back pain had comorbidity in the underweight category (Table 15).

Table 14 Relation between comorbidity, limited activities and back pain[8]

	Limitact==1 and backpain==1	Limitact==1 and backpain==0	Limitact==2 and backpain==1	Limitact==2 and backpain==0	Limitact2==1 and backpain==1	Limitact2==1 and backpain==0	Limitact2==2 and backpain==1	Limitact2==2 and backpain==0
No comorbidity	0	0	342	4,170	0	0	342	4,170
Comorbidity	802	541	197	1,036	504	517	495	1,060
No comorbidity (2)	298	24	407	4,191	0	0	705	4,215
Comorbidity (2)	504	517	132	1,015	504	517	132	1,015

Table 15 Comorbidities distribution by BMI and back pain

BMI	Under 18.4	18.5 – 24	25 - 29	30 - 39	Over 40	Missing
Back pain						
Comorbidity	6 (100%)	201 (67.91%)	273 (68.59%)	253 (79.81%)	50 (87.72%)	216 (79.41%)
No comorbidity	0	94 (31.86%)	123 (31.06%)	63 (19.94%)	7 (12.28%)	55 (20.22%)
No back pain						
Comorbidity	18 (22.50%)	358 (20.72%)	488 (28.84%)	347 (39.03%)	38 (44.71%)	328 (11.15%)
No comorbidity	62 (77.50%)	1,367 (79.11%)	1,195 (70.63%)	539 (60.63%)	47 (55.29%)	960 (32.64%)
Total						
Comorbidity	24 (27.91%)	559 (27.62%)	761 (36.41%)	600 (49.75%)	88 (61.97%)	544 (16.93%)
No comorbidity	62 (72.09%)	1,461 (72.18%)	1,318 (63.06%)	602 (49.92%)	54 (38.03%)	1,015 (31.59%)

We also checked how many comorbidities respondents had in different BMI categories.

[8] Explanation of Table 14:.
Out of these 1,353 with limited activities (limitact), 1,343 had a comorbidity (comorbid).
Out of these 1,353 with limited activities (limitact), 802 had a comorbidity (comorbid) and back pain (backpain==1).
Out of these 1,353 with limited activities (limitact), 0 did not have a comorbidity (comorbid) BUT had back pain (backpain==1).
Out of these 1,353 with limited activities (limitact), 541 had a comorbidity (comorbid) BUT no back pain (backpain==0).
Out of these 1,353 with limited activities (limitact), 1,021 had a comorbidity (comorbid2 - not related to back pain – not equal to value 35 in IllsM1_6).
Out of these 1,353 with limited activities, 504 had a comorbidity (comorbid2) and back pain (backpain==1).
Out of these 1,353 with limited activities, 298 did not have a comorbidity (comorbid2) BUT had back pain (backpain==1).
Out of these 1,353 with limited activities (limitact), 517 had a comorbidity (comorbid2) BUT no back pain (backpain==0).

Table 16 Number of comorbidities in respondents with back pain according to BMI categories

Number of comorbidities	0	1	2	3	4	5	6
Under 18.4	0	3 (50.00%)	1 (16.67%)	0	1 (16.67%)	0	1 (16.67%)
18.5 – 24	95 (32.09%)	103 (34.80%)	64 (21.62%)	25 (8.45%)	7 (2.36%)	2 (0.68%)	0
25 -29	125 (31.41%)	115 (28.89%)	95 (23.87%)	40 (10.05%)	15 (3.77%)	8 (2.01%)	0
30 -39	64 (20.19%)	89 (28.08%)	83 (26.18%)	40 (12.62%)	34 (10.73%)	6 (1.89%)	1 (0.32%)
Over 40	7 (12.28%)	12 (21.05%)	19 (33.33%)	11 (19.30%)	4 (7.02%)	4 (7.02%)	0
Missing	56 (20.59%)	84 (30.88%)	65 (23.90%)	41 (15.07%)	21 (7.72%)	5 (1.84%)	0

Since most of the respondents with back pain reported having comorbidities, we checked (Table 17) how the results changed if comorbidities due to back pain problems were excluded (comorbid2 and comorbidcount2 variables were used). After excluding comorbidities related to back pain, the proportion of people with back pain and a comorbidity decreases from around 70% to around 40-50%.

Table 17 Comorbidities (excluding any related to back pain) distribution by BMI and back pain

BMI	Under 18.4	18.5 – 24	25 - 29	30 - 39	Over 40	Missing
Back pain						
Comorbidity	4 (66.67%)	121 (40.88%)	172 (43.22%)	161 (50.79%)	35 (61.40%)	143 (52.57%)
No comorbidity	2 (33.33%)	174 (58.78%)	224 (56.28%)	155 (48.90%)	22 (38.60%)	128 (47.06%)
No back pain						
Comorbidity	18 (22.50%)	348 (20.14%)	470 (27.78%)	336 (37.80%)	38 (44.71%)	322 (10.95%)
No comorbidity	62 (77.50%)	1,377 (79.69%)	1,213 (71.69%)	550 (61.87%)	47 (55.29%)	966 (32.85)
Total						
Comorbidity	22 (25.58%)	469 (23.17%)	642 (30.72%)	497 (41.21%)	73 (51.41%)	465 (14.47%)
No comorbidity	64 (74.42%)	1,551 (76.63%)	1,437 (68.76%)	705 (58.46%)	69 (48.59%)	1,094 (34.05%)

4.2.4 Comorbidity variable based on QOF diseases

In contrast to the comborbidity variable above, a new variable **qofdis** was also generated based on the absence/ presence of Qualities & Outcomes Framework (QOF) diseases. QOF disease equivalents were recorded in HSfE 2011 variables IllsM1-6: cancer (neoplasm), diabetes, mental illness /anxiety/depression/nerves, epilepsy/fits/convulsions, stroke/cerebral haemorrhage/cerebral thrombosis, heart attack/angina, hypertension/high blood pressure/blood pressure and asthma.

Qofdis variable is generated based on the values of IllsM1-6 values:

- 1 "Yes" (if IllsM1-6 was equal to 1, or 2, or 4, or 6, or 15, or 16, or 17, or 23)
- 0 "No" (if otherwise)

Table 18 Number of comorbidities (excluding related to back pain) in respondents with back pain according to BMI

Number of comorbidities	0	1	2	3	4	5	6
Under 18.4	2 (33.33%)	3 (50.00%)	0	0	0	0	1 (16.67%)
18.5 – 24	175 (59.12%)	62 (20.95%)	39 (13.18%)	15 (5.07%)	3 (1.01%)	2 (0.68%)	0
25 -29	226 (56.78%)	81 (20.35%)	55 (13.82%)	21 (5.28%)	9 (2.26%)	6 (1.51%)	0
30 -39	156 (49.21%)	54 (17.03%)	54 (17.03%)	29 (9.15%)	21 (6.62%)	2 (0.63%)	1 (0.32%)
Over 40	22 (38.60%)	6 (10.53%)	12 (21.05%)	10 (17.54%)	4 (7.02%)	3 (5.26%)	0
Missing	129 (47.43%)	50 (18.38%)	48 (17.65%)	28 (10.29%)	14 (5.15%)	3 (1.10%)	0

Out of 8,761 respondents, 1,394 (15.36%) had a QOF disease or a combination of them. The HSfE variable coding is shown in Section 6, Appendix 1: coding details of Health Survey for England variables, QOF diseases.

4.2.5 Other variable changes

The **agegrp** variable has 7 categories, however, the first '<=25' category was quite small, especially between respondents with back pain (n=27; 2.01% of all respondents with back pain). Therefore, a new variable **agegrp2** was generated where the '<=24' and' 25-34' groups were merged together, making <34 the reference group used in analyses.

Socioeconomic status was analysed using **nssec8** variable. However, there are nine categories instead of eight. The last category is coded as 'Other'. According to HSfE 2011 documents, respondents were assigned to this group if they were full-time students or occupations were not stated/inadequate. Since it is impossible to distinguish between these groups in the nssec8 variable, the latter group was coded as missing - 223 (2.55%) values were coded as missing (7 (0.52%) for respondents with back pain and 216 (2.91%) for respondents with no back pain.

4.4 Statistical analyses: overall back pain

We first examined overall back pain cases versus non-cases/controls. Table 19 below shows the univariate logistic regression modelling results in the ELSA dataset containing cases and non-cases.

4.4.1 Logistic univariate analyses

Table 19 Univariate analyses

	Odds ratio (OR)	95% CI	p-value
Total number of respondents			
Age (agegrp)[9]			
<=25	1.00		
25-34	9.84	[6.39-15.15]	<0.001
35-44	19.67	[13.08-29.56]	<0.001
45-54	26.97	[18-40.42]	<0.001
55-64	39.24	[26.21-58.74]	<0.001
65-74	45.85	[30.48-68.96]	<0.001
Over 75	61.91	[41.08-93.29]	<0.001
Age (agegrp2)			
<34	1.00		
35-44	5.78	[4.58-7.3]	<0.001
45-54	7.93	[6.32-9.95]	<0.001
55-64	11.54	[9.21-14.46]	<0.001
65-74	13.49	[10.68-17.03]	<0.001
Over 75	18.21	[14.37-23.07]	<0.001
Gender			
Male	1.00		
Female	1.55	[1.38-1.75]	<0.001
Ethnicity			
White	1.00		
Mixed	0.22	[0.11-0.45]	<0.001
Asian	0.56	[0.43-0.74]	<0.001
Black/Black Caribbean	0.39	[0.25-0.62]	<0.001
Other	1.88	[1.02-3.48]	0.044
Not stated	0.51	[0.16-1.68]	0.268
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	1.37	[1.09-1.72]	0.007
NVQ3/GCE A level equivalent	0.91	[0.72-1.13]	0.387
NVQ2/GCE O level equivalent	1.31	[1.09-1.59]	0.005
NVQ1/CSE other grade equivalent	1.34	[0.97-1.85]	0.071
Foreign/other	2.99	[1.9-4.7]	<0.001
No qualification	2.83	[2.38-3.37]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.38	[1.09-1.75]	0.008
Intermediate occupations	1.60	[1.25-2.06]	<0.001

[9] This variable was not used in multivariate analyses. Its univariate analysis results are presented here.

	Odds ratio (OR)	95% CI	p-value
Small employers and own account workers	1.68	[1.26-2.24]	<0.001
Lower supervisory and technical	1.98	[1.48-2.67]	<0.001
Semi-routine occ.	1.82	[1.43-2.32]	<0.001
Routine occ.	2.34	[1.83-2.99]	<0.001
Never worked and long term unemployed	1.06	[0.63-1.79]	0.822
BMI (bmigrp)[10]			
<18.4 underweight	1.00		
18.5 – 24 normal	2.28	[0.99-5.28]	0.054
25 – 29 overweight	3.14	[1.36-7.24]	0.007
30 – 39 obese	4.75	[2.05-11.01]	<0.001
>40 obese	8.94	[3.65-21.88]	<0.001
BMI 2 (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.41	[1.2-1.66]	<0.001
30 – 39 obese	2.13	[1.79-2.55]	<0.001
>40 obese	4.01	[2.81-5.74]	<0.001
Obesity			
No	1.00		
Yes	1.92	[1.66-2.22]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.34	[1.13-1.58]	0.001
Never regular	0.76	[0.65-0.89]	0.001
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	3.05	[2.65-3.51]	<0.001
Extremely anxious or depressed	4.80	[3.41-6.77]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	0.84	[0.6-1.18]	0.317
Retired	3.07	[2.67-3.53]	<0.001
Other economically inactive	1.85	[1.56-2.19]	<0.001
Activities limited due to illness (limitact)			
Yes	1.00		
No	0.07	[0.06 – 0.08]	<0.001
Activities limited due to illness excluding related to back(limitact2)			
Yes	1.00		
No	0.16	[0.14 – 0.19]	<0.001
Self-assessed general health			
Very good	1.00		
Good	2.91	[2.42-3.49]	<0.001
Fair	11.44	[9.42-13.89]	<0.001
Bad	62.43	[45.8-85.1]	<0.001

[10] This variable was not used in multivariate analyses. Its univariate analyses results are presented.

	Odds ratio (OR)	95% CI	p-value
Very bad	69.76	[44.9-108.39]	<0.001
Comorbidities (comorbid)			
No comorbidity	1.00		
Comorbidity present	7.72	[6.74-8.85]	<0.001
Comorbidities excluding related to back pain (comorbid2)			
No comorbidity	1.00		
Comorbidity present	2.48	[2.2-2.8]	<0.001
Comorbidity count categories			
No comorbidity	1.00		
1	4.40	[3.76-5.15]	<0.001
2	11.56	[9.59-13.94]	<0.001
3	20.58	[15.56-27.22]	<0.001
4 or more	42.40	[28.03-64.12]	<0.001
Comorbidity count categories excluding related to back pain (2)			
No comorbidity	1.00		
1	1.40	[1.2-1.64]	<0.001
2	3.74	[3.09-4.53]	<0.001
3	7.13	[5.3-9.6]	<0.001
4 or more	14.17	[9.12-22.03]	<0.001
QOF diseases			
Absent	1.00		
Present	3.30	[2.89-3.77]	<0.001

4.4.2 Logistic multivariate analyses

We went through an extensive model fitting process. Table 20 and Table 21 show the final two multivariate regression models (M26b and M28) chosen for overall back pain. We included one additional variable at a time to observe the effects. The tables for all the other models are included in Section 7 Appendix 2: details of multivariate model fitting.

Table 20: overall back pain, autostepwise excluding depression and ethnicity (M26b)

	OR	95% CI	p-value
Total number of respondents	5,267		
Age (agegrp2)			
<34	1.00		
35-44	2.54	[1.95-3.31]	<0.001
45-54	3.16	[2.44-4.11]	<0.001
55-64	4.39	[3.39-5.7]	<0.001
65-74	5.30	[4.03-6.97]	<0.001
Over 75	6.71	[5.03-8.94]	<0.001
Gender			
Male	1.00		
Female	1.56	[1.33-1.82]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.33	[1.02-1.73]	0.034
Intermediate occupations	1.35	[1.01-1.79]	0.04

	OR	95% CI	p-value
Small employers and own account workers	1.54	[1.11-2.13]	0.01
Lower supervisory and technical	1.90	[1.36-2.67]	<0.001
Semi-routine occ.	1.42	[1.08-1.87]	0.013
Routine occ.	1.93	[1.46-2.55]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.55	[1.33-1.8]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.75	[0.65-0.87]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	2.00	[1.61-2.47]	<0.001

Table 21: overall back pain, stepwise backward and forward. Excluding economic activity (M28)

	OR	95% CI	p-value
Total number of respondents	5,269		
Age (agegrp2)			
<34	1.00		
35-44	2.44	[1.88-3.18]	<0.001
45-54	3.00	[2.32-3.89]	<0.001
55-64	4.07	[3.14-5.27]	<0.001
65-74	4.34	[3.29-5.72]	<0.001
Over 75	5.36	[3.99-7.2]	<0.001
Gender			
Male	1.00		
Female	1.66	[1.44-1.93]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.00		
Semi-routine occ.	1.42	[1.07-1.87]	0.014
Routine occ.	1.00		
Never worked and long term unemployed	1.47	[1.2-1.8]	<0.001
BMI (bmigrp3)			
<18.4	1.00		

	OR	95% CI	p-value
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.57	[1.34-1.82]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.74	[0.64-0.86]	<0.001
Education			
Any qualification	1.00		
No qualification	1.27	[1.05-1.52]	0.012

4.4.3 Interactions in the overall back pain models

In the next step, various interactions were tested in models M26B and M28. Interactions were tested between age and gender; socioeconomic status and economic activity; socioeconomic status and smoking status; socioeconomic status and education level; BMI and smoking status; BMI and economic activity; economic activity and education; age and education; age and socioeconomic status; age and smoking status; age and economic status; age and BMI (results summarised in Table 22 and Table 23).

Table 22 Overall back pain, interaction terms for model M26B, ROC=0.7051

Interaction		**Interaction status/p-value**	**Likelihood ratio test (Prob>chi2)**
Age	Gender	No interaction	0.2985
Age	**Education**	**Interaction below***	**NA** (observations number differ) ROC=0.7014
35-44	No qualification	P=0.007	
Age	**Socioeconomic status**	**Interaction below***	0.1276
55-64	Lower managerial and professional	P=0.006	
55-64	Intermediate occupations	P=0.038	
55-64	Small employers and own account workers	P=0.020	
55-64	Lower supervisory and technical	P=0.016	
55-64	Semi-routine occupations	P=0.002	
55-64	Routine occupations	P<0.001	
Age	Smoking status	No interaction	0.7272
Age	**Economic status**	**Interaction below***	0.0008 (ROC=0.7119)
45-54	Other economically inactive	P=0.001	
55-64	Other economically inactive	P=0.001	
Age	BMI	No interaction	0.0707
Socioeconomic status	Economic activity	No interaction	**NA** (observations number differ) ROC=0.7109
Socioeconomic status	Smoking status	No interaction	0.7189
Socioeconomic status	**Education**	**Interaction below***	

Interaction		Interaction status/p-value	Likelihood ratio test (Prob>chi2)
Intermediate occupation	NVQ3/GCE A level	P=0.045	**NA** (observations number differ) ROC=0.7104
Lower supervisory and technical	NVQ2/GCE O level	P=0.027	
Lower supervisory and technical	No qualification	P=0.003	
BMI	Smoking status	No interaction	0.6541
BMI	Economic activity	No interaction	0.2139
Economic activity	**Education**	**Interaction below***	**NA** (observations number differ) ROC=0.7106
Other economically inactive	NVQ2/GCE O level	P=0.015	
Other economically inactive	No qualification	P=0.007	

Table 23 Overall back pain, interaction terms model M28

Interaction		Interaction status/p-value	Likelihood ratio test (Prob>chi2)
Age	Gender	No interaction	0.2335
Age	**Education**	**Interaction below***	**NA** (observation number differ) ROC=0.7043
45-54	No qualification	P=0.006	
55-64	No qualification	P=0.002	
Age	**Socioeconomic status**		**0.2426**
55-64	Lower managerial and professional	P=0.006	
55-64	Intermediate occupations	P=0.050	
55-64	Small employers and own account workers	P=0.017	
55-64	Lower supervisory and technical	P=0.014	
55-64	Semi-routine occupations	P=0.003	
55-64	Routine occupations	P<0.001	
Age	Smoking status	No interaction	0.7248
Age	BMI	No interaction	0.0682
Socioeconomic status	Smoking status	No interaction	0.8701
Socioeconomic status	**Education**	**Interaction below***	**NA** (observation number differ) ROC=0.7039
Lower supervisory and technical	No qualification	P=0.004	
Semi-routine occupations	Higher education below degree	P=0.050	
BMI	Smoking status	No interaction	0.6730

4.5 Statistical analyses: severe back pain

We then carried out model fitting for severe back pain. Table 24 and Table 25 show the final two multivariate regression models (models 5B and 7) chosen for severe back pain. We included one additional variable at a time to observe the effects. The tables for all the other models are included in Section 7, Appendix 2: details of multivariate model fitting.

Table 24 Severe back pain, autostepwise excluding depression and ethnicity model 5B

	OR	95% CI	p-value
Total number of respondents	5,261		
Age (agegrp2)			
<34	1.00		
35-44	2.56	[1.79-3.65]	<0.001
45-54	3.43	[2.43-4.85]	<0.001
55-64	3.83	[2.67-5.5]	<0.001
65-74	3.40	[2.17-5.34]	<0.001
Over 75	4.41	[2.74-7.1]	<0.001
Gender			
Male	1.00		
Female	1.80	[1.49-2.18]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.80	[1.28-2.51]	0.001
Semi-routine occ.	1.36	[1.07-1.72]	0.011
Routine occ.	1.70	[1.32-2.18]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.33	[1.06-1.66]	0.012
>30 obese	2.10	[1.67-2.63]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.77	[0.6-0.99]	0.041
Never regular	0.56	[0.44-0.71]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.97	[1.42-2.72]	<0.001
Other economically inactive	3.39	[2.64-4.34]	<0.001
Education			
NVQ4/NVQ5 degree	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level	1.00		

	OR	95% CI	p-value
NVQ2/GCE O level	1.00		
NVQ1/CSE other grade	1.00		
Foreign/other	1.00		
No qualification	1.48	[1.19-1.83]	<0.001

Table 25: severe back pain, autostepwise excluding ethnicity, depression, economic activity model 7

	OR	95% CI	p-value
Total number of respondents	5,263		
Age (agegrp2)			
<34	1.00		
35-44	2.37	[1.67-3.36]	<0.001
45-54	3.02	[2.15-4.25]	<0.001
55-64	4.17	[2.97-5.85]	<0.001
65-74	4.28	[2.99-6.13]	<0.001
Over 75	5.84	[4-8.52]	<0.001
Gender			
Male	1.00		
Female	1.97	[1.64-2.37]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.78	[1.27-2.48]	0.001
Semi-routine occ.	1.36	[1.08-1.72]	0.01
Routine occ.	1.83	[1.43-2.35]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.32	[1.06-1.65]	0.013
>30 obese	2.14	[1.71-2.68]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.57	[0.45-0.72]	<0.001
Education			
NVQ4/NVQ5 degree	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level	1.00		
NVQ2/GCE O level	1.29	[1.03-1.63]	0.029
NVQ1/CSE other grade	1.00		
Foreign/other	1.00		
No qualification	1.82	[1.44-2.3]	<0.001

4.5.1 Interactions: severe back pain

Various interactions were tested in models 5B and 7. Interactions were tested between age and gender; socioeconomic status and economic activity; socioeconomic status and smoking status; socioeconomic status and education level; BMI and smoking status; BMI and economic activity; economic activity and education; age and education; age and socioeconomic status; age and smoking status; age and economic status; age and BMI (results summarised in Table 26 and Table 27).

Table 26 Severe back pain, interaction terms for model 5B ROC=0.7663

Interaction		Interaction status/p-value	Likelihood ratio test (Prob>chi2)
Age	Gender	No interaction	0.3457
Age	**Education**	**Interaction below***	**NA** (observation number differ) ROC=0.6993
45-54	No qualification	P=0.026	
Age	**Socioeconomic status**	**Interaction below***	**NA** (observation number differ) ROC=0.7030
55-64	Semi-routine occupations	P=0.031	
55-64	Routine occupations	P=0.008	
Age	Smoking status	No interaction	0.7698
Age	**Economic status**		0.0001 (ROC=0.7756)
55-64	Other economically inactive	P<0.001	
Age	BMI	No interaction	0.0762
Socioeconomic status	**Economic activity**	Interaction below*	**NA** (observation number differ)
Lower supervisory and technical	Retired	P=0.035	
Socioeconomic status	Smoking status	No interaction	0.6186
Socioeconomic status	**Education**	**Interaction below***	**NA** (observation number differ) ROC=0.7758
Lower supervisory and technical	No qualification	P=0.001	
BMI	Smoking status	No interaction	0.2322
BMI	Economic activity	No interaction	0.0472 (ROC=0.7732)
Economic activity	**Education**	**Interaction below***	
Retired	NVQ1/CSE other grade equivalent	P=0.047	**NA** (observation number differ) ROC=0.7735
Other economically inactive	NVQ2/GCE O level equivalent	P=0.050	

Table 27 Severe back pain, interaction terms for model 7

Interaction		Interaction status/p-value	Likelihood ratio test (Prob>chi2)
Age	Gender	No interaction	0.2488
Age	**Education**	**Interaction below***	**NA** (observation number differ) ROC=0.7592
45-54	No qualification	P=0.023	
55-64	No qualification	P=0.005	
Age	**Socioeconomic status**	**Interaction below***	

Interaction		Interaction status/p-value	Likelihood ratio test (Prob>chi2)
55-64	Small employers and own account workers	P=0.032	**NA** (observation number differ) ROC=0.7597
55-64	Lower supervisory and technical	P=0.052	
55-64	Semi-routine occ.	P=0.044	
55-64	Routine occ.	P=0.014	
Age	Smoking status	No interaction	0.7232
Age	BMI	No interaction	0.0897
Socioeconomic status	Smoking status	No interaction	0.7592
Socioeconomic status	**Education**	**Interaction below***	**NA** (observation number differ) ROC=0.7566
Lower supervisory and technical	NVQ2/GCE O level equivalent	P=0.047	
Lower supervisory and technical	No qualification	P=0.003	
BMI	Smoking status	No interaction	0.3723

4.6 ROC curves

4.6.1 Overall back pain

We next examined the receiver operating characteristics (ROC) curves for the various models (Table 28). The best ROC curve, which predicts data perfectly, will touch the top-left corner of the plot (area 1.0), and the larger the area under the ROC curve, the better the prediction. An area of 0.5 signifies a prediction no better than chance. However, as noted in the Methods, the choice of variables and hence the final model also depends on the availability of local data, so the final local model will not predict as well as the optimal/"gold standard", e.g. Model 5 below.

Table 28 Overall back pain, comparison of different models

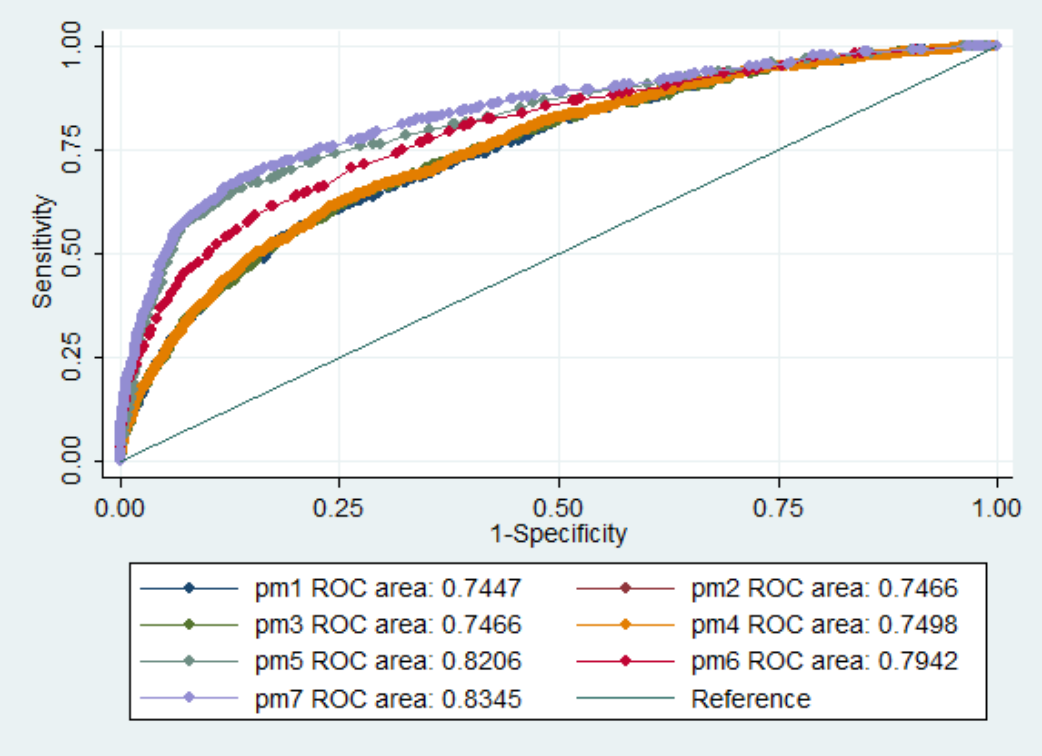
Model description	Model	ROC area	SE	95% CI
Auto stepwise forward model as in Table 45	M1	0.7447	0.0087	[0.7276 – 0.7619]
Auto stepwise backward model as in Table **46**	M2	0.7466	0.0887	[0.7295 – 0.7637]
Auto stepwise forward and backward models combined as in Table **47**	M3	0.7466	0.0087	[0.7295 – 0.7637]
Full model (with all covariates) as in Table 33	M4	0.7498	0.0087	[0.7328 – 0.7668]
Auto stepwise forward and backward model including limitact as in Table **48**	M5	0.8206	0.0080	[0.8050 – 0.8363]
Auto stepwise forward and backward model including genhealth as in Table 49	M6	0.7930	0.0083	[0.7768 – 0.8093]

Model description	Model	ROC area	SE	95% CI
Auto stepwise forward and backward model including limitact and genhealth as in Table **50**	M7	0.8345	0.0078	[0.8193 – 0.8498]
Auto stepwise forward and backward model including limitact2 (excluded if related to back pain) as in Table **51**	M8	0.7735	0.0085	[0.7568 – 0.7903]
Auto stepwise forward and backward model including comorbid as in Table **52**	M9	0.7999	0.0079	[0.7843 – 0.8154]
Auto stepwise forward and backward model including comorbid2 as in Table **53**	M10	0.7492	0.0087	[0.7322 – 0.7663]
Auto stepwise forward and backward model including comorbidity count (comorbidcountcat) as in Table **54**	M11	0.8063	0.0080	[0.7906 – 0.8220]
Auto stepwise forward and backward model including comorbidity count excluding related to back pain (comorbidcountcat2) as in Table 55	M12	0.7573	0.0086	[0.7404 – 0.7742]
Autostepwise forward and backward model including qofdis as in Table **56**	M13	0.7400	0.0090	[0.7224 – 0.7575]
Autostepwise forward and backward model including obesity as in Table **57**	M14	0.7369	0.0090	[0.7193 – 0.7544]
Autostepwise forward and backward model including obesity and qofdis as in Table **58**	M15	0.7384	0.0090	[0.7209 – 0.7560]
Autostepwise forward and backward model including limitact2 and comorbid2 as in Table **59**	M16	0.7726	0.0087	[0.7553 – 0.7896]
Autostepwise forward and backward model including qofdis and limitact2 as in Table 60	M17	0.7664	0.0088	[0.7492 – 0.7836]
Autostepwise forward and backward model including qofdis and limitact2 BUT excluding econact as in Table 61	M18	0.7649	0.0088	[0.7477 – 0.7821]
Autostepwise forward and backward model including qofdis, limitact2, bmigrp3 BUT excluding econact as in Table **62**	M19	0.7644	0.0088	[0.7472 – 0.7817]

Model description	Model	ROC area	SE	95% CI
Autostepwise forward and backward model including limitact2 and qofdis and bmigrp3 as in Table **63**	M20	0.7660	0.0088	[0.7488 – 0.7833]
Autostepwise forward and backward model including qofdis and bmigrp3 as in (excluded limiact2 as there is no local data for it) Table 64	M21	0.7384	0.0090	[0.7209 – 0.7560]
Autostepwise forward and backward model (bmigrp3) Table **65**	M22	0.7369	0.0090	[0.7193 – 0.7545]
Autostepwise forward and backward model (bmigrp3) BUT excluding anxiety/depression question Table **66**	M23	0.7066	0.0088	[0.6893 – 0.7239]
Autostepwise forward and backward model (bmigrp3) BUT excluding anxiety/depression question AND including limitact2 Table **67**	M24	0.7440	0.0087	[0.7269 – 0.7610]
Autostepwise forward and backward model (bmigrp3) BUT excluding anxiety/depression question AND including limitact2 (ethnicity excluded as only 52 individuals were in category 'Other') Table 68	M24b	0.7401	0.0088	[0.7229 – 0.7574]
Autostepwise forward and backward model (bmigrp3) BUT excluding anxiety/depression question AND including limitact2 AND genhealth Table **69**	M25	0.7904	0.0082	[0.7743 – 0.8065]
Autostepwise forward and backward model (bmigrp3) BUT excluding anxiety/depression question as seen in Table 70	M26	0.7066	0.0088	[0.6893 – 0.7239]
Autostepwise forward and backward model (bmigrp3) BUT excluding anxiety/depression question as seen in Table 20	M26b	0.7051	0.0089	[0.6877 – 0.7225]
Autostepwise forward and backward model (bmigrp3) including depression (derived) variable	M27	0.6697	0.0122	0.6458 – 0.6937
Autostepwise forward and backward model (bmigrp3) BUT excluding anxiety/depression question and economic activity as seen in Table 21	M28 (final model)	0.6945	0.0089	[0.6771-0.7119]

Figure 1 below illustrates some of the ROC curves for the tabulated models.

Figure 1 ROC curves comparison of different models



Explanations for different models:
Pm1 - Stepwise forward model
Pm2 - Stepwise backward model
Pm3 - Stepwise forward and backward
Pm4 - Full model
Pm5 - Stepwise forward model and backward + limitact variable added
Pm6 - Stepwise forward model and backward + genhealth variable added
Pm7 - Stepwise forward and backward + limitact + genhealth

The sections below concentrate on models pm3 and pm7.

4.6.2 Severe back pain

Models using outcome variable severity1 finish in 'A', while models using outcome as severity2 finish in 'B' (creation of these variables is described in section 6.1.1).

Table 29 Comparison of different models

Model description	Model	ROC area	SE	95% CI	Equivalent to 'No Back Pain v Back Pain' model
Model excluding depression but including limitact2 as seen in Table 72	1A (severity1)	0.8004	0.0105	[0.7798 – 0.8210]	M24
Model excluding depression but including limitact2 as seen in Table 73	1B (severity2)	0.8097	0.0092	[0.7918 – 0.8277]	M24

Model description	Model	ROC area	SE	95% CI	Equivalent to 'No Back Pain v Back Pain' model
Model excluding depression and ethnicity but including limitact2 as in Table 74	2A (severity1)	0.7984	0.107	[0.7774 – 0.8194]	M24b
Model excluding depression and ethnicity but including limitact2 as in Table 75	2B (severity2)	0.8065	0.0094	[0.7881 – 0.8248]	M24b
Model excluding depression but including limitact2 and genhealth	3A (severity1)	0.8672	0.0087	[0.8502 – 0.8842]	M25
Model excluding depression but including limitact2 and genhealth	3B (severity2)	0.8754	0.0074	[0.8609 – 0.8899]	M25
Model excluding depression as seen in Table 76	4B (severity2)	0.7705	0.0096	[0.7516 – 0.7894]	M24
Model excluding depression and ethnicity as seen in Table 24	5B (severity2)	0.7663	0.0099	[0.7470 – 0.7857]	M24b
Model including depression (derived) but ethnicity excluded	6 (severity2)	0.7689	0.0097	[0.7499 – 0.7879]	M27
Model excluding depression and ethnicity and economic activity as seen in	7 (severity2) (final model)	0.7490	0.0099	[0.7297-683]	M28

4.7 Internal validation of HSfE2011: How good is our model at predicting back pain caseness?

4.7.1 Overall back pain

We could use the HSfE2011 automatic stepwise forward and backward models combined to predict the probability of an individual having back pain in the HSfE2011 data set. In the figures below, the two box plots show the predicted probability of people having back pain among the non-back pain and back pain groups. . Since we have a binary response model, we can choose a cut-off point on the predicted probability to separate the predicted back pain cases (with higher predicted probability) from the predicted non-back pain cases (with lower predicted probability). We can tell from the box plots that people will not be mis-classified with a reasonable degree of certainty.

Figure 2 Predicted probabilities of having back pain (stepwise forward and backward model)

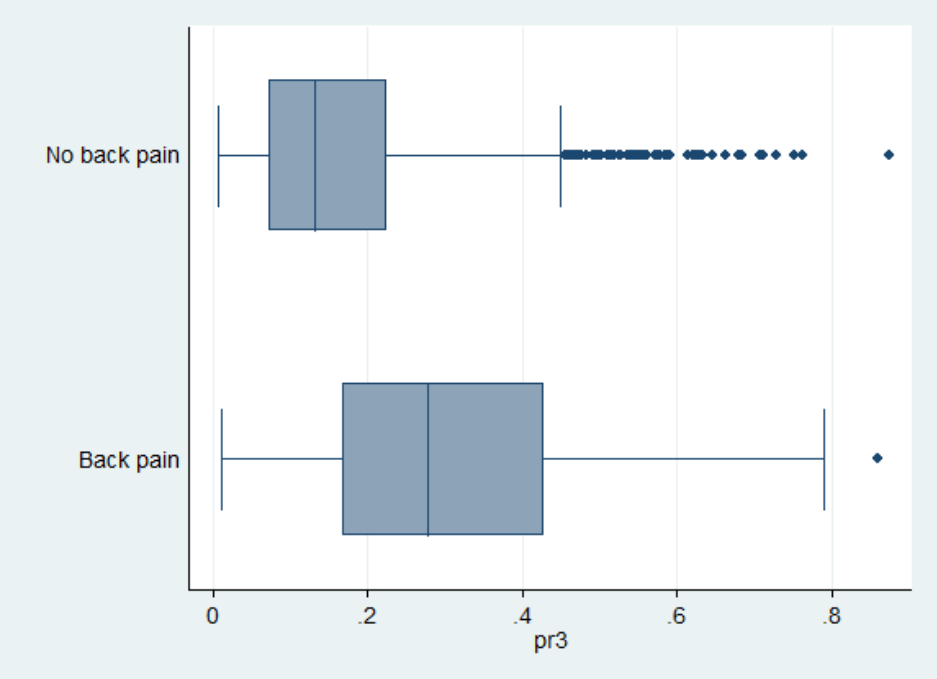


Figure 3 Predicted probabilities of having back pain (stepwise forward and backward model including limitact and genhealth)

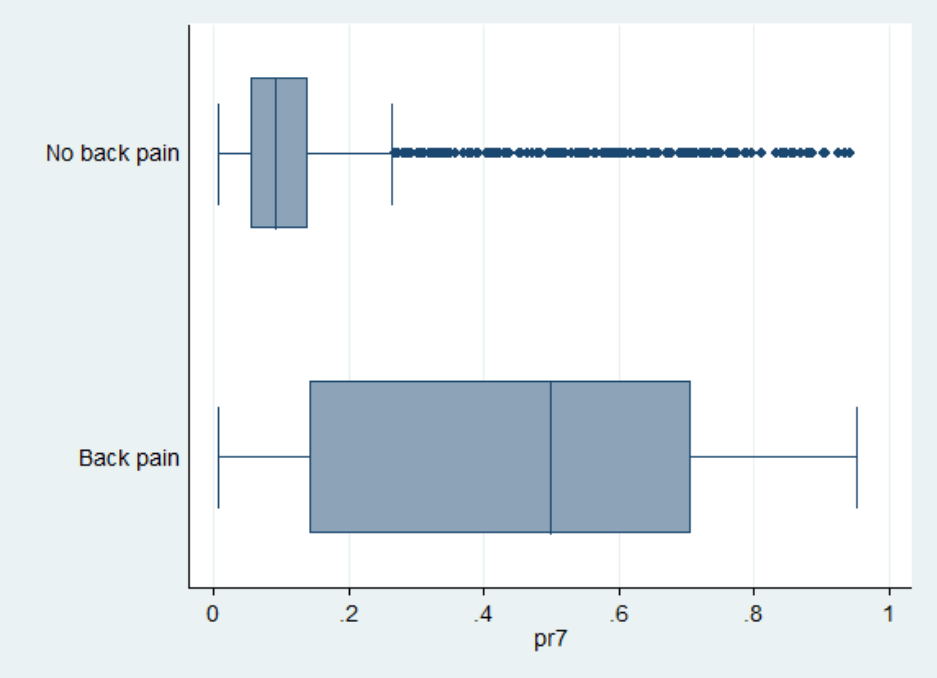


Figure 4 Sensitivity/specificity versus probability cut-off (stepwise forward and backward model)

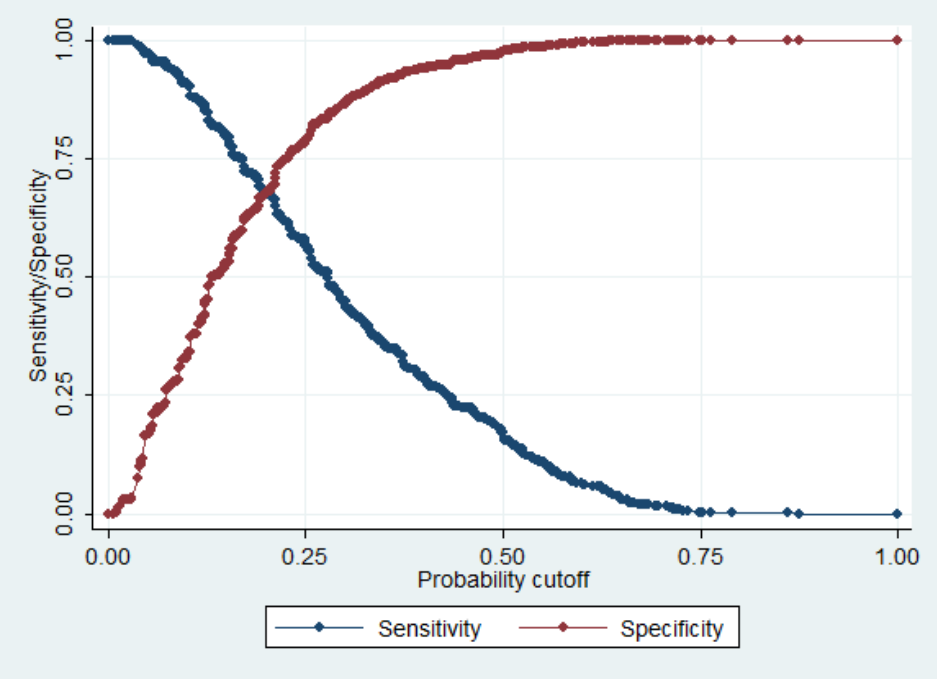
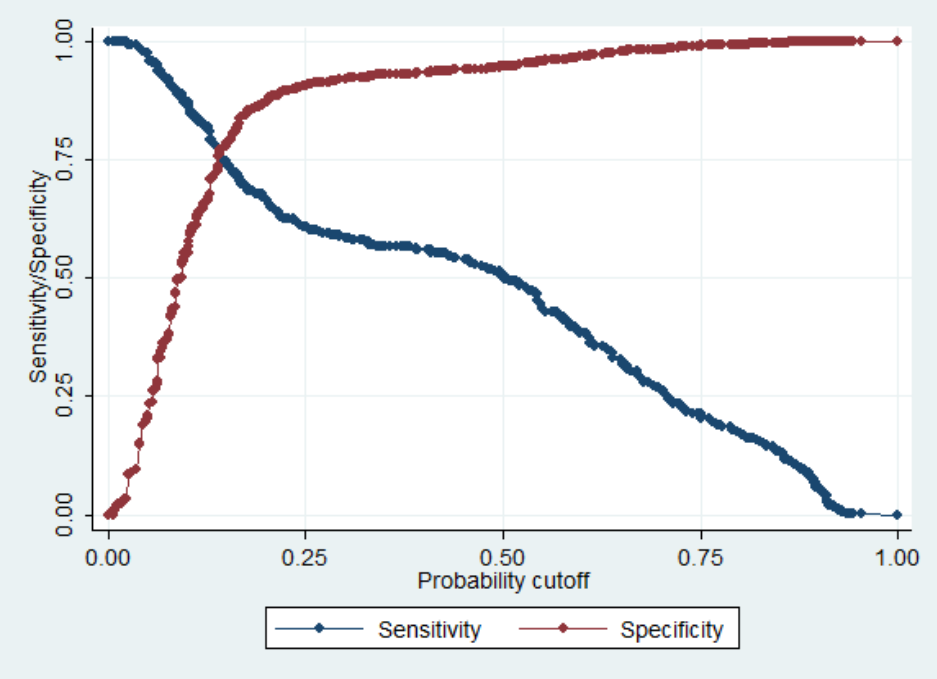


Figure 5 Sensitivity/specificity versus probability cut-off (stepwise forward and backward model including limitact and genhealth)



The sensitivity/specificity versus probability cut-off plots show us the corresponding sensitivity and specificity in each possible probability cut-off point (See Figure 4 and Figure 5). Higher sensitivity would usually yield low specificity and vice versa, the rule of thumb is to choose a cut-off probability to maximize both. We choose the cut-off probability where sensitivity and specificity lines cross. Applying different cut-off probabilities to our data, the following tables show the comparison between predicted and true cases of back pain in HSfE (Table 30 and Table 31).

Table 30 Predicted back pain caseness with different cut-off probabilities (stepwise forward and backward model)

Probability cut-off	0	0.05	0.10	0.15	0.20	0.21	0.22	0.25	0.30
Sensitivity (%)	100%	96.96%	90.87%	79.72%	67.85%	66.33%	62.27%	56.69%	45.03%
Specificity (%)	0%	16.81%	33.19%	52.77%	67.77%	69.29%	74.23%	78.88%	86.32%
True positive	986	956	896	786	669	654	614	559	444
False positive	4,086	3,399	2,730	1,930	1,317	1,255	1,053	863	559
True negative	0	687	1,356	2,156	2,769	2,831	3,033	3,223	3,527
False negative	0	30	90	200	317	332	372	427	542

Table 31 Predicted back pain caseness with different cut-off probabilities (stepwise forward and backward model including limitact and genhealth)

Probability cut-off	0	0.05	0.10	0.13	0.14	0.15	0.20	0.25	0.30
Sensitivity (%)	100%	97.57%	87.22%	79.01%	75.66%	74.54%	66.43%	**60.85%**	58.52%
Specificity (%)	0%	20.22%	55.28%	70.89%	75.67%	77.72%	87.20%	**90.67%**	92.07%
True positive	986	962	860	779	746	735	655	**600**	577
False positive	4,085	3,259	2,258	2,896	994	910	523	**381**	324
True negative	0	826	1,827	1,189	3,091	3,175	3,562	**3,704**	3,761
False negative	0	24	126	207	240	251	331	**386**	409

4.8 Synthetic estimation/ prevalence calculation

Back pain prevalence partly depends on respondents' risk factors. Therefore, the range of prevalence is quite wide. The lowest prevalence of overall back pain is 4.18% for a male younger than 34 years with any BMI (excluding only above 30), who never smoked and has one of the following occupations: higher managerial/lower managerial/intermediate occupation/small employer and own account workers/routine occupations or never worked. In contrast, the highest prevalence for overall back pain is 60.34%; if a person is female aged over 75 being obese (BMI over 30), either current or ex-smoker, who has no education and is in routine occupation.

The smallest prevalence for severe back pain is 1.21%. This is observed in males aged below 34' who never smoked with BMI under 18.4 or in the range of 18.4-25 with any education and are in either higher managerial/lower managerial/intermediate/small employer and own account worker occupation or never worked.

The highest prevalence for severe back pain is similar to the maximal value for overall back pain – 63.82%. This prevalence is observed for obese females aged over 75 with no education who are current smokers in routine occupations. As described in Methods section, the odds ratios and coefficients are on a logarithmic scale, so they were added instead of multiplied. For example, for a female over 75 years of age, who has never worked, with BMI over 30, but with high levels of physical activity, who is an ex-smoker with no educational qualifications the outcome would be calculated:

Outcome = -2.2649 + (-0.1663) + 0.597079 + 0 + 0.628003 + (-0.68873) + 0 + 0.131858 = -1.76299

ODDS = exp(-1.76299) = 0.1715

Prevalence = 0.1715/(1+0.1715) = 0.1464/14.6% (this shows the prevalence rate of total back pain for a person with those characteristics).

4.8.1 Estimating number of people (population) with different characteristics

Local population data for every risk factor that was used in a model are obtained. . For example, for a female over 75 years of age, who has never worked, with a BMI over 30, but with high levels of physical activity, who is an ex-smoker with no educational qualifications the population would be calculated at MSOA level (Hartlepool 001):

Population = Number of females over 75 * proportion of population that never worked * proportion of population whose BMI over 30 * proportion of population that has high physical activity levels * proportion of population that never smoked * proportion of population that do not have education.

This population (0.3221) is then multiplied by the prevalence, or proportion of cases in that population, from the regression model:

Proportion = 0.1464 * 0.3221 = 0.0472 (this shows number of cases of total back pain at Hartlepool 001 for a person with characteristics described above).

The sum of all values in this table is the number of expected cases of Hip or Knee OA in the selected MSOA/LA or practice/CCG.

Estimating number of severe/total back pain cases**:** the number of cases is calculated by multiplying the prevalence by the population in each demographic category for the selected MSOA/LA or practice/CCG. For example, for a female over 75 years of age, who has never worked, with BMI over

30, but with high levels of physical activity, who is an ex-smoker with no education, the proportion would be calculated:

*Proportion = Prevalence * Population*

Proportion = 0.1464 * 0.3221 = 0.0472 (this shows number of cases of total back pain at Hartlepool MLSOA 001 for a person with characteristics described above). The sum of all values in this table is the number of expected cases of back pain in the selected MSOA/LA or practice/CCG.

Example of small population prevalence estimates

Figure 6 and Figure 7 show the estimate of the prevalence from the model and number or proportion of cases, respectively in a practice/MLSOA. These are dependent on the number of cases in risk factor subgroups. There is obviously considerable random variation in these, and therefore in the population subgroup prevalence estimates. This variation is reduced by aggregating cells to produce the final practice/MLSOA totals, for which the CIs are calculated. For this reason the subgroup cells are not copiable or downloadable.

4.8.2 Example of small population prevalence estimates

Figure 6: example of back pain prevalence estimates for population subcategories, Hartlepool MLSOA 001

Predicted number of general Back Pain cases in Hartlepool 001 (E02002483) = 1623

Predicted prevalence of general Back Pain in Hartlepool 001 (E02002483) = 17.467 %

PREVALENCE RATE			Male & No Educ								Male & Any Educ							
Smoking Stauts	BMI	Age	Higher managerial and professional occupations	Lower managerial and professional occupations	Intermediate occupations	Small employers and own account workers	Lower supervisory and technical occupations	Semi-routine occupations	Routine occupations	Never worked and long term unemployed	Higher managerial and professional occupations	Lower managerial and professional occupations	Intermediate occupations	Small employers and own account workers	Lower supervisory and technical occupations	Semi-routine occupations	Routine occupations	Never worked and long term unemployed
SMOKER	Under 18.4	<34	0.0025	0.0073	0.0047	0.0020	0.0042	0.0054	0.0073	0.0014	0.0159	0.0457	0.0291	0.0123	0.0265	0.0335	0.0458	0.0089
		35 - 44	0.0023	0.0066	0.0042	0.0018	0.0037	0.0049	0.0064	0.0013	0.0116	0.0334	0.0212	0.0090	0.0188	0.0245	0.0324	0.0065
		45 - 54	0.0039	0.0111	0.0071	0.0030	0.0061	0.0082	0.0106	0.0022	0.0134	0.0384	0.0244	0.0103	0.0214	0.0281	0.0369	0.0074
		55 - 64	0.0046	0.0133	0.0085	0.0036	0.0072	0.0097	0.0123	0.0026	0.0114	0.0328	0.0208	0.0088	0.0180	0.0240	0.0309	0.0063
		65 - 74	0.0073	0.0209	0.0133	0.0056	0.0113	0.0153	0.0193	0.0041	0.0069	0.0200	0.0127	0.0054	0.0109	0.0146	0.0187	0.0039
		Over 75	0.0066	0.0190	0.0121	0.0051	0.0101	0.0139	0.0172	0.0037	0.0064	0.0183	0.0116	0.0049	0.0099	0.0134	0.0169	0.0035
	18.5-24	<34	0.0730	0.2101	0.1334	0.0565	0.1208	0.1538	0.2092	0.0407	0.4549	1.3094	0.8319	0.3525	0.7575	0.9587	1.3123	0.2537
		35 - 44	0.0659	0.1897	0.1205	0.0511	0.1056	0.1389	0.1820	0.0368	0.3321	0.9561	0.6074	0.2574	0.5378	0.7000	0.9285	0.1853
		45 - 54	0.1108	0.3190	0.2027	0.0859	0.1755	0.2336	0.3022	0.0618	0.3823	1.1004	0.6991	0.2962	0.6131	0.8057	1.0571	0.2132
		55 - 64	0.1323	0.3808	0.2419	0.1025	0.2056	0.2788	0.3531	0.0738	0.3259	0.9381	0.5960	0.2526	0.5141	0.6869	0.8846	0.1818
		65 - 74	0.2083	0.5995	0.3808	0.1614	0.3222	0.4389	0.5531	0.1162	0.1986	0.5717	0.3632	0.1539	0.3120	0.4186	0.5367	0.1108
		Over 75	0.1892	0.5445	0.3459	0.1466	0.2882	0.3987	0.4938	0.1055	0.1821	0.5240	0.3329	0.1411	0.2821	0.3837	0.4844	0.1015
	25-29	<34	0.0907	0.2610	0.1658	0.0703	0.1502	0.1911	0.2600	0.0506	0.5652	1.6270	1.0337	0.4380	0.9412	1.1913	1.6307	0.3153
		35 - 44	0.0819	0.2357	0.1498	0.0635	0.1312	0.1726	0.2261	0.0457	0.4127	1.1880	0.7547	0.3198	0.6683	0.8698	1.1537	0.2302

PREVALENCE RATE			Male & No Educ								Male & Any Educ							
Smoking Stauts	BMI	Age	Higher managerial and professional occupations	Lower managerial and professional occupations	Intermediate occupations	Small employers and own account workers	Lower supervisory and technical occupations	Semi-routine occupations	Routine occupations	Never worked and long term unemployed	Higher managerial and professional occupations	Lower managerial and professional occupations	Intermediate occupations	Small employers and own account workers	Lower supervisory and technical occupations	Semi-routine occupations	Routine occupations	Never worked and long term unemployed
		45 - 54	0.1377	0.3964	0.2518	0.1067	0.2181	0.2902	0.3755	0.0768	0.4750	1.3673	0.8686	0.3681	0.7618	1.0011	1.3135	0.2649
		55 - 64	0.1644	0.4732	0.3006	0.1274	0.2555	0.3465	0.4388	0.0917	0.4050	1.1657	0.7406	0.3138	0.6388	0.8535	1.0991	0.2259
		65 - 74	0.2588	0.7449	0.4732	0.2005	0.4003	0.5454	0.6873	0.1443	0.2468	0.7103	0.4513	0.1912	0.3877	0.5201	0.6668	0.1376
		Over 75	0.2351	0.6766	0.4299	0.1821	0.3581	0.4954	0.6136	0.1311	0.2262	0.6512	0.4137	0.1753	0.3505	0.4768	0.6019	0.1262
	Over 30	<34	0.1105	0.3180	0.2020	0.0856	0.1804	0.2328	0.3118	0.0616	0.6939	1.9973	1.2689	0.5377	1.1421	1.4624	1.9757	0.3870
		35 - 44	0.0954	0.2747	0.1745	0.0739	0.1489	0.2011	0.2558	0.0532	0.4880	1.4046	0.8923	0.3781	0.7725	1.0284	1.3298	0.2722
		45 - 54	0.1581	0.4550	0.2891	0.1225	0.2432	0.3332	0.4172	0.0882	0.5545	1.5960	1.0139	0.4297	0.8668	1.1686	1.4898	0.3093
		55 - 64	0.1840	0.5297	0.3365	0.1426	0.2766	0.3878	0.4733	0.1026	0.4624	1.3310	0.8456	0.3583	0.7078	0.9745	1.2135	0.2579
		65 - 74	0.2880	0.8290	0.5267	0.2232	0.4307	0.6070	0.7364	0.1606	0.2803	0.8069	0.5126	0.2172	0.4270	0.5908	0.7317	0.1563
		Over 75	0.2564	0.7380	0.4688	0.1987	0.3767	0.5403	0.6428	0.1430	0.2523	0.7262	0.4614	0.1955	0.3780	0.5317	0.6465	0.1407

Figure 7: back pain estimates for a local population (MLSOAs in Hartlepool LA)

MSOA Results			General Back Pain		Severe Back Pain	
MSOA Code	**MSOA Name**	**Total Population**	**Number**	**Prevalence**	**Number**	**Prevalence**
E02002483	Hartlepool 001	9,293	1,623	17.47%	1,012	10.89%
E02002484	Hartlepool 002	10,418	1,862	17.88%	1,206	11.58%
E02002485	Hartlepool 003	7,774	1,416	18.21%	925	11.90%
E02002487	Hartlepool 005	5,396	905	16.77%	586	10.86%
E02002488	Hartlepool 006	6,074	1,111	18.29%	654	10.77%
E02002489	Hartlepool 007	7,840	1,305	16.64%	815	10.39%
E02002490	Hartlepool 008	6,180	1,119	18.11%	739	11.95%
E02002491	Hartlepool 009	6,520	1,290	19.79%	836	12.82%
E02002492	Hartlepool 010	6,681	1,277	19.12%	806	12.06%
E02002493	Hartlepool 011	6,400	1,210	18.90%	738	11.54%
E02002494	Hartlepool 012	7,785	1,419	18.23%	955	12.26%
E02006909	Hartlepool 014	11,877	2,364	19.91%	1,478	12.44%
LA Results						
E06000001	Hartlepool	92,238	16,887	18.31% [17.26%-19.29%]	10,697	11.60% [10.76%-12.47%]

4.9 Production of Scottish local estimates

4.9.1 Methods

We used the model developed from HSfE to produce the prevalence of overall back pain and severe back pain for Scotland. However, risk factor data availability affects the risk factor variables included in the prediction model. Therefore, some variables may be dropped from the English model, if no local data available. The performance of the models was compared by c-statistics (ROC curves).

4.9.2 Results

Occupation data are not available in Scotland, so this variable was dropped from the Scotland model. Figure 8 and Figure 9 shown the performance of the overall back pain model for England (ROC curve=0.6840) and Scotland (ROC curve=0.6897). Figure 10 and Figure 11 shown the performance of the overall back pain model for England (ROC curve=0.7418) and Scotland (ROC curve=0.7412). There is minimal deterioration in discrimination as a result of this deletion.

Figure 8: Overall Back Pain England model

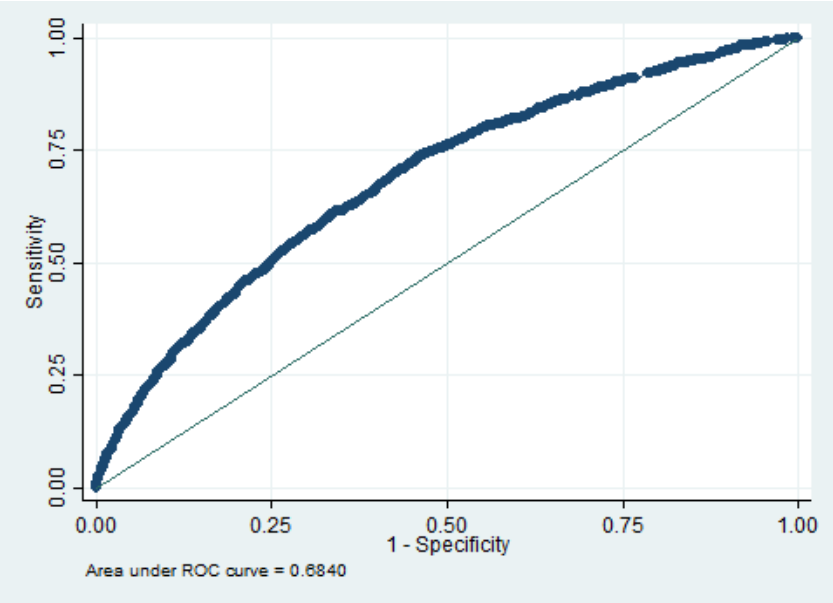


Figure 9: Overall Back Pain Scotland model

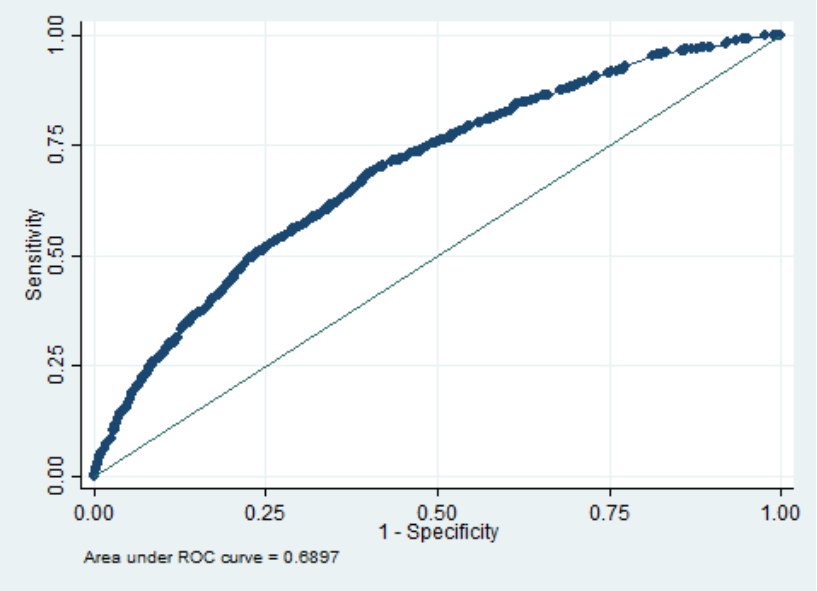


Figure 10: Severe Back Pain England Model

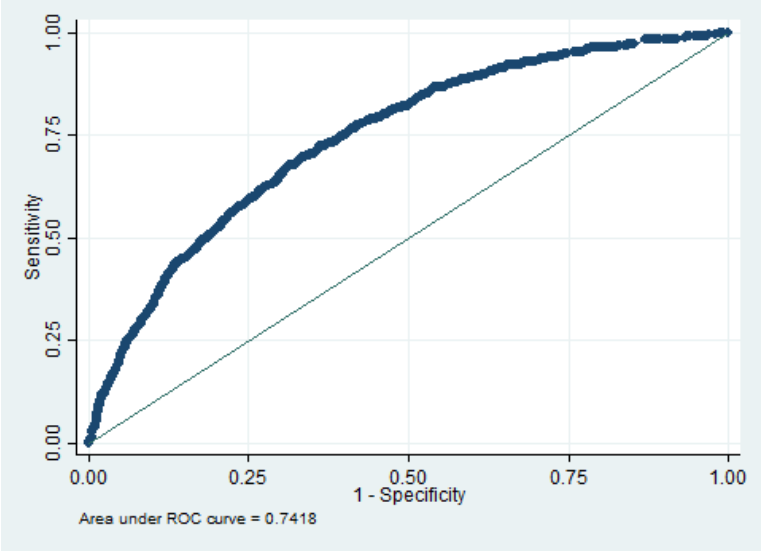


Figure 11: Severe Back Pain Scotland Model

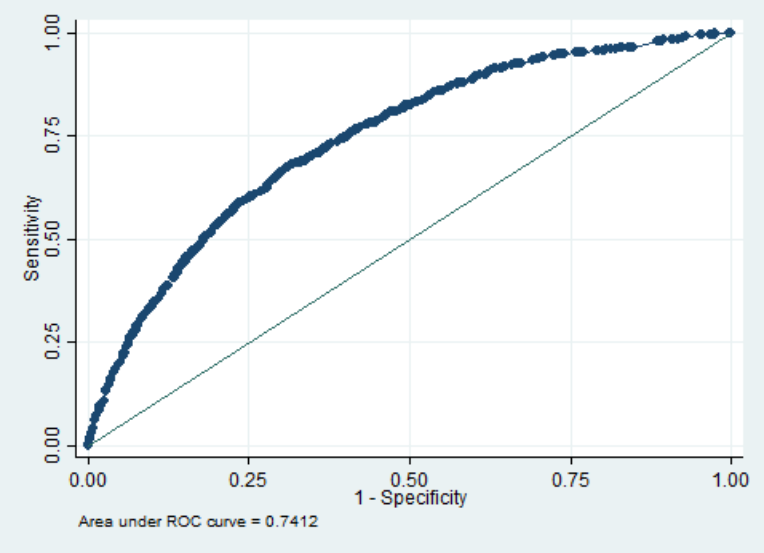


Figure 12 and Figure 13 shown the histogram of Scotland overall back pain prevalence at practice and LA levels. Figure 14 and Figure 15 shown the histogram of Scotland severe back pain prevalence at practice and LA levels. The average prevalence is around 19.35% for overall back pain and 11.85% for severe back pain.

Figure 12: Histogram of Scotland overall back prevalence at practice level

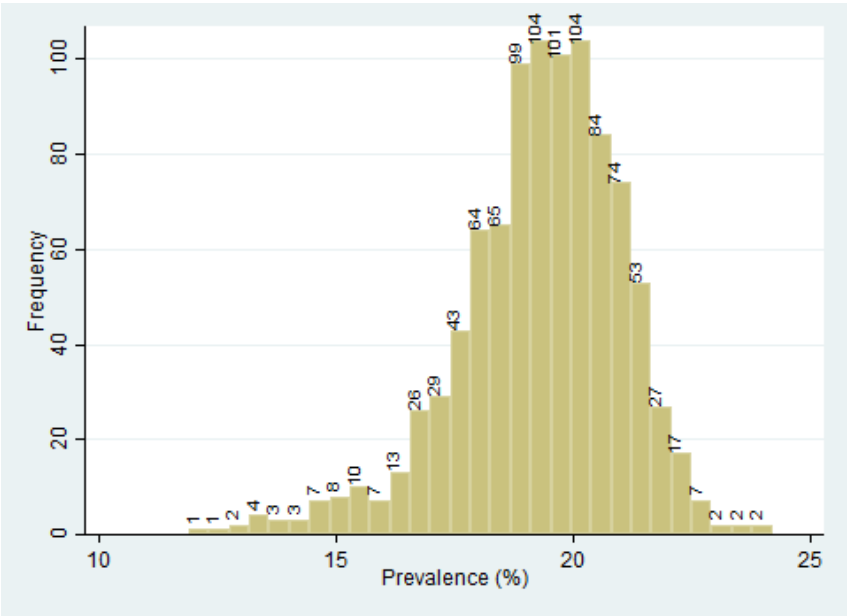


Figure 13: Histogram of Scotland overall back pain prevalence at LA level

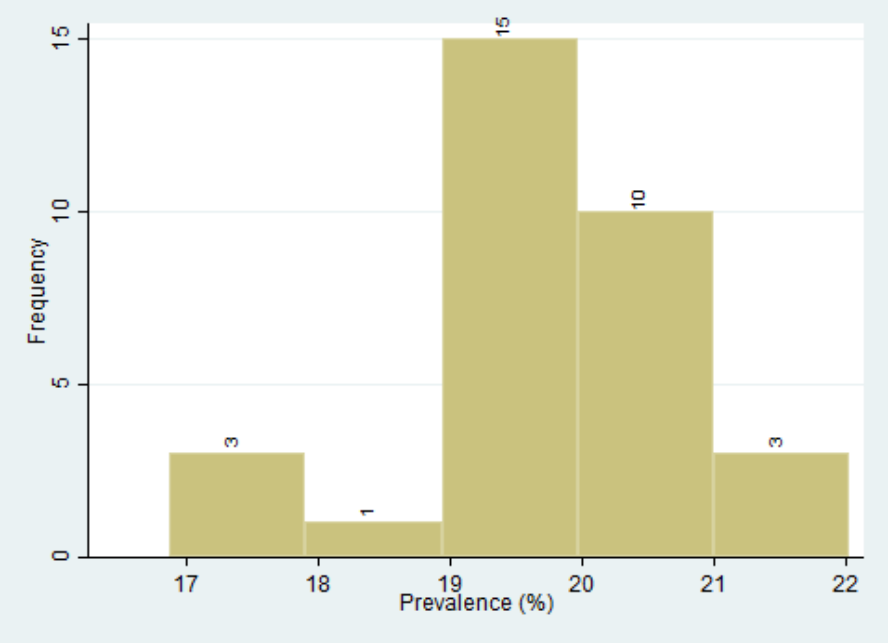


Figure 14: Histogram of Scotland severe back pain prevalence at practice level

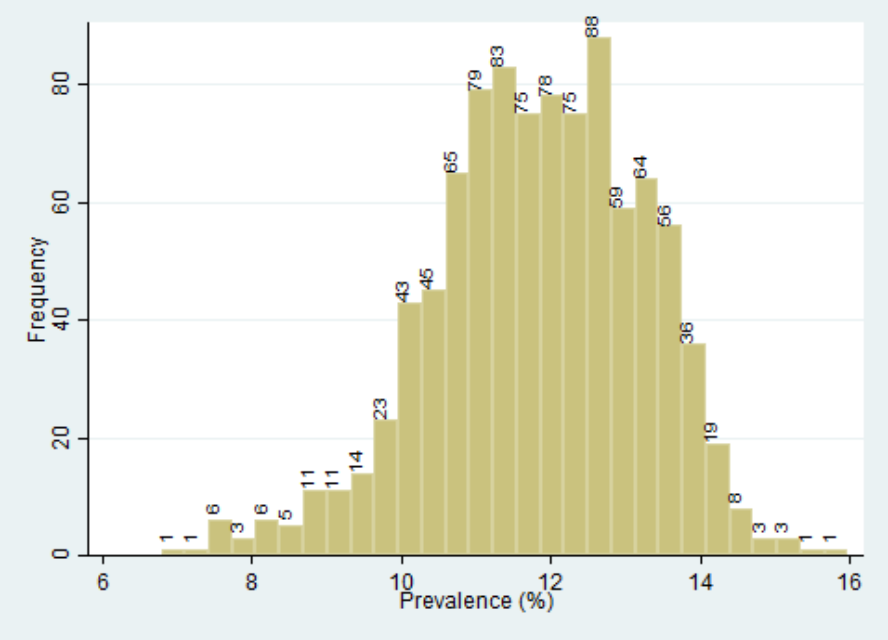
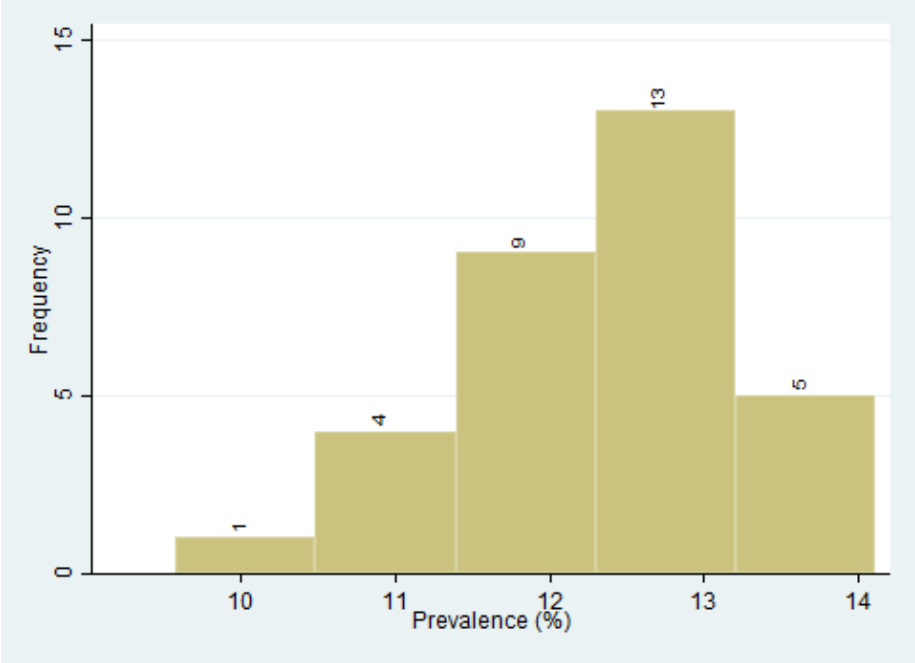


Figure 15: Histogram of Scotland severe back pain prevalence at LA level



ARUK commissioned an external validation of the prevalence model from Aberdeen University, using Scottish Biobank data. The report forms Section 9, Appendix 4: external validation of Health Survey for England 2011 models for low back pain: data from UK-BioBank Scotland.

4.10 Production of Wales local estimates

4.10.1 Methods

We were unable to produce MSK estimates for Wales during the main contract because lookup tables from GP practice populations to SOAs were unavailable. This meant we were unable to map key Census variables to GP practices and similarly unable to map practice variables to resident populations. For example, smoking data was available from practice QOF data, but because of the lack of a lookup table it could not be mapped to SOAs. However lookup tables became available in late 2017, so in 2018 we produced Wales estimate. We used the model developed from ELSA to produce the prevalence of overall and severe hip OA, and overall and severe knee OA for Wales. However, risk factor data availability affected the risk factor variables included in the prediction model. Therefore, some variables were dropped from the English model because no local data was available. The performance of the complete/English models were compared with the incomplete/Wales estimates using c-statistics (ROC curves).

As shown in Table 32, there were significant differences between the variables in the UK model and the local Wales data. It proved to be impossible to match alcohol consumption data between the HSfE and Welsh Health Survey lifestyle trends (2015) categories, as the latter included the categories: drinking above guidelines on a day in the past week, heavy (binge) drinking and very heavy drinking. We therefore dropped alcohol from the England model and local Wales risk factor data. Therefore, the final model included only age, gender, BMI (four categories), smoking (three categories), deprivation (fifths) and ethnicity (five categories).

Table 32: choice of Wales local risk factor data

Geography 1	Risk factor	Definition	Action in national regression model	Geography2	Name of source
Practice/LSOA	Age & gender	Same		LA, MLSOA	Population estimates by lower super output area and age group
Practice/LSOA	Age & gender	Same		Health Board/practice	Population estimates by age group & gender
LSOA	Socioeconomic status	Very similar	No change	Mapped to Health Board/practice from LSOA data	Welsh Index of Multiple Deprivation
LA	BMI	Overweight or obese and obese only (so overweight only can be calculated, giving three categories), we used 4 categories	Combine underweight and normal range to obtain 3 categories to match Wales	Practice/ Health Board	Welsh Health Survey lifestyle trends (2015)
LA	Smoking	Current smokers + ecigarette users only in Welsh Health Survey. we used three categories in England data. We mapped QOF practice level smoking data to MSOAs.	No action required, same categories	Health Board/ practice	QOF data (smoker/ex-smoker/non-smoker)
LA	Education status	Highest qualification		LA	2011 Census: Qualifications1 and students, local authorities in the United Kingdom
LA	Alcohol consumption	Heavy drinking & binge drinking,		National level data	Welsh Health Survey lifestyle trends (2015)
LA	Ethnicity	Same	Same	Council area	Ethnic Group Demographics
LSOA	Deprivation	Rank	Same	Mapped to practices via lookup table	Welsh Index of Multiple Deprivation

For English estimates derived from English national surveys e.g. HSfE and ELSA (or from CPRD data for that matter as it only includes 40 Welsh practices) we internally validated local estimates by aggregating the numbers of cases to Regional level and compared them to national data disaggregated to Regional level. However as no national Wales data was used for to develop the regression models we were unable to do this for back pain.

4.10.2 Results

Figure 16 shows the discrimination (ROC curve) of Wales's severe back pain regression model. The C-statistic is 0.7418, which is the same as the England model (Figure 10 in the Scotland model section). A decreased number of variables are included.

Figure 16: discrimination (ROC curve) of Wales's severe back pain regression model

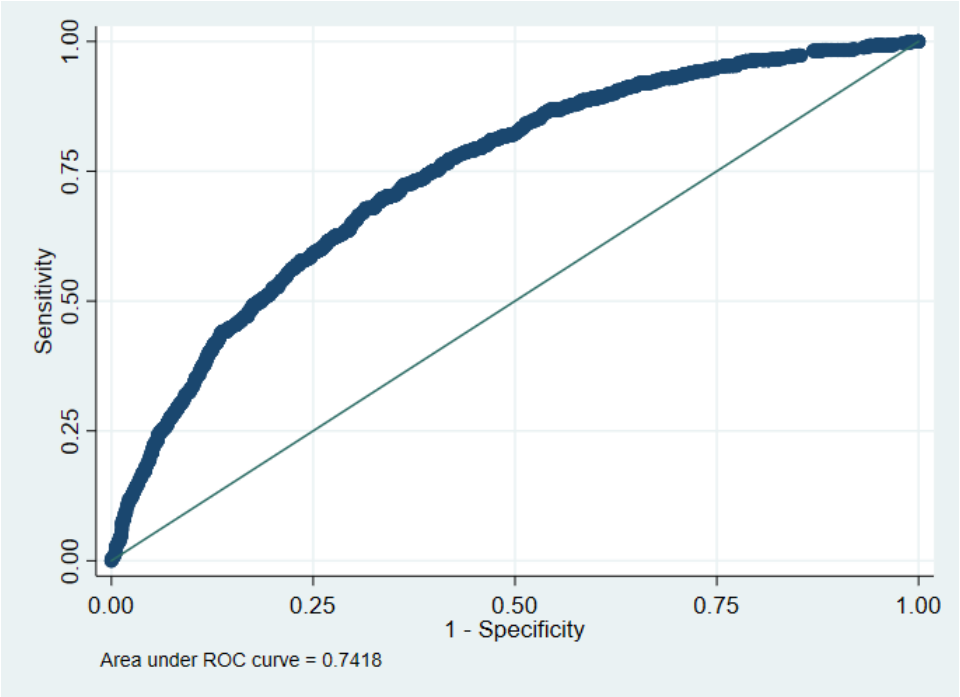
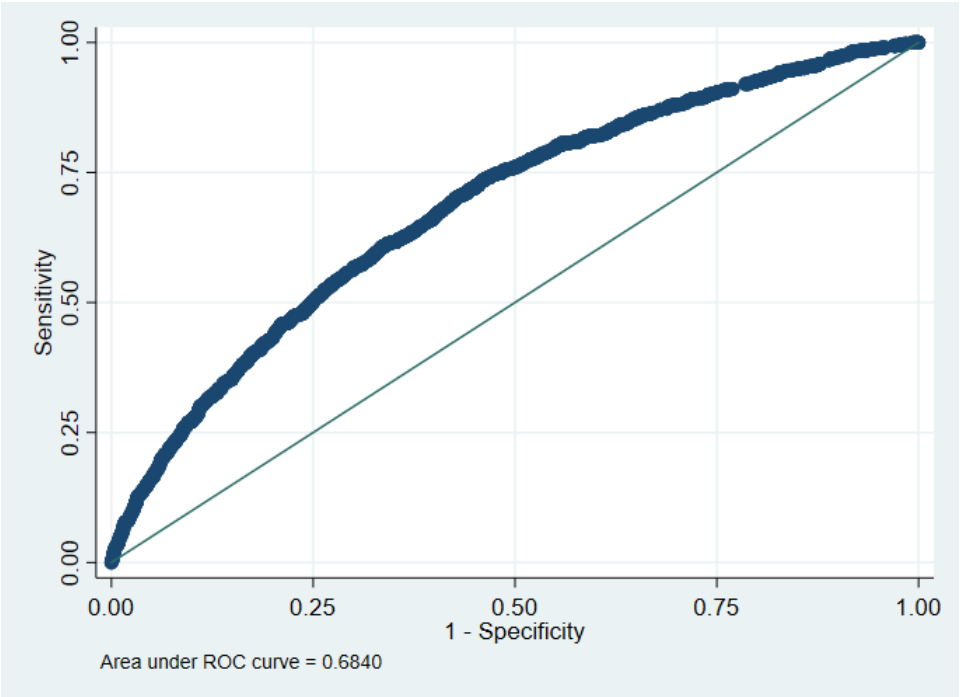


Figure 17 shows the discrimination (ROC curve) of Wales's general back pain regression model. The C-statistic is 0.684, which is the same as the England model (Figure 8 in the Scotland model section).

Figure 17: discrimination (ROC curve) of Wales's general back pain regression model



We were not able to carry out an external validation against QOF-registered data, because there are no QOF indicators for back pain.

5 References

1. Hong J, Reed C, Novick D, Happich M. Costs Associated With Treatment of Chronic Low Back Pain: An Analysis of the UK General Practice Research Database. Spine 2013;38(1):75-82. doi: 10.1097/BRS.0b013e318276450f Link: http://journals.lww.com/spinejournal/Fulltext/2013/01010/Costs_Associated_With_Treatment_of_Chronic_Low.13.aspx.

2. Health & Social Care Information Centre. Health Survey for England 2011 Volume2 - Methods and documentation, 2012 Link.

3. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50:133-49. doi: http://dx.doi.org/10.1016/0304-3959(92)90154-4 Link: http://www.sciencedirect.com/science/article/pii/0304395992901544.

4. Nacul L, Soljak M, Meade T. Model for estimating the population prevalence of chronic obstructive pulmonary disease: cross sectional data from the Health Survey for England. Population Health Metrics 2007;5(1):8. Link: http://www.pophealthmetrics.com/content/5/1/8.

5. Murray CJL, Richards MA, Newton JN, Fenton KA, Anderson HR, et al. UK health performance: findings of the Global Burden of Disease Study 2010. The Lancet 2015;381(9871):997-1020. doi: 10.1016/S0140-6736(13)60355-4 Link: http://dx.doi.org/10.1016/S0140-6736(13)60355-4.

6. Shiri R, Solovieva S, Husgafvel-Pursiainen K, Taimela S, Saarikoski LA, et al. The association between obesity and the prevalence of low back pain in young adults: the Cardiovascular Risk in Young Finns Study. Am J Epidemiol 2008;167(9):1110-9. doi: 10.1093/aje/kwn007 Link: http://www.ncbi.nlm.nih.gov/pubmed/18334501.

7. Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, et al. The Rising Prevalence of Chronic Low Back Pain. Archives of Internal Medicine 2009;169(3):251-58. Link: http://archinte.ama-assn.org/cgi/content/abstract/169/3/251.

8. Bjorck-van Dijken C, Fjellman-Wiklund A, Hildingsson C. Low back pain, lifestyle factors and physical activity: a population based-study. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine 2008;40(10):864-9. doi: 10.2340/16501977-0273 Link: http://www.medicaljournals.se/jrm/content/?doi=10.2340/16501977-0273.

9. Mafi JN, McCarthy EP, Davis RB, Landon BE. Worsening trends in the management and treatment of back pain. JAMA Internal Medicine 2013;173(17):1573-81. doi: 10.1001/jamainternmed.2013.8992 Link: http://dx.doi.org/10.1001/jamainternmed.2013.8992.

10. Palmer KT, Walsh K, Bendall H, Cooper C, Coggon D. Back pain in Britain: comparison of two prevalence surveys at an interval of 10 years. BMJ 2000;320(7249):1577-8. Link: http://www.bmj.com/content/320/7249/1577.long.

11. Harkness EF, Macfarlane GJ, Silman AJ, McBeth J. Is musculoskeletal pain more common now than 40 years ago?: Two population-based cross-sectional studies. Rheumatology (Oxford) 2005;44(7):890-5. doi: 10.1093/rheumatology/keh599 Link: http://www.ncbi.nlm.nih.gov/pubmed/15784630.

12. Pengel LH, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. BMJ 2003;327(7410):323. doi: 10.1136/bmj.327.7410.323 Link: http://www.bmj.com/content/327/7410/323.long.

13. Taylor JB, Goode AP, George SZ, Cook CE. Incidence and risk factors for first-time incident low back pain: a systematic review and meta-analysis. The Spine Journal 2014;14(10):2299-319. doi: http://dx.doi.org/10.1016/j.spinee.2014.01.026 Link: http://www.sciencedirect.com/science/article/pii/S152994301400103X.

14. Kerr MS, Frank JW, Shannon HS, Norman RW, Wells RP, et al. Biomechanical and psychosocial risk factors for low back pain at work. Am J Public Health 2001;91(7):1069-75. Link: http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.91.7.1069.

15. Croft PR, Papageorgiou AC, Ferry S, Thomas E, Jayson MI, et al. Psychologic distress and low back pain. Evidence from a prospective study in the general population. Spine (Phila Pa 1976) 1995;20(24):2731-7. Link: http://www.ncbi.nlm.nih.gov/pubmed/8747252.

16. Larson SL, Clark MR, Eaton WW. Depressive disorder as a long-term antecedent risk factor for incident back pain: a 13-year follow-up study from the Baltimore Epidemiological Catchment Area sample. Psychological medicine 2004;34(2):211-9. Link: http://www.ncbi.nlm.nih.gov/pubmed/14982127.

17. Currie SR, Wang J. More data on major depression as an antecedent risk factor for first onset of chronic back pain. Psychological medicine 2005;35(9):1275-82. doi: 10.1017/S0033291705004952 Link: http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=328842&fileId=S0033291705004952.

18. Croft PR, Rigby AS. Socioeconomic influences on back problems in the community in Britain. Journal of epidemiology and community health 1994;48:166-70. Link: http://jech.bmj.com/content/48/2/166.abstract?sid=83a5c817-a782-4072-9593-59f259b4d283.

19. Schneider S, Randoll D, Buchner M. Why do women have back pain more than men? A representative prevalence study in the federal republic of Germany. The Clinical journal of pain 2006;22(8):738-47. doi: 10.1097/01.ajp.0000210920.03289.93 Link: http://www.ncbi.nlm.nih.gov/pubmed/16988571.

20. Wijnhoven HA, de Vet HC, Picavet HS. Prevalence of musculoskeletal disorders is systematically higher in women than in men. The Clinical journal of pain 2006;22(8):717-24. doi: 10.1097/01.ajp.0000210912.95664.53 Link: http://www.ncbi.nlm.nih.gov/pubmed/16988568.

21. Leboeuf-Yde C. Body weight and low back pain. A systematic literature review of 56 journal articles reporting on 65 epidemiologic studies. Spine (Phila Pa 1976) 2000;25(2):226-37. Link: http://www.ncbi.nlm.nih.gov/pubmed/10685488.

22. Lake JK, Power C, Cole TJ. Back pain and obesity in the 1958 British birth cohort. cause or effect? J Clin Epidemiol 2000;53(3):245-50. Link: http://www.sciencedirect.com/science/article/pii/S0895435699001559.

23. Janke EA, Collins A, Kozak AT. Overview of the relationship between pain and obesity: What do we know? Where do we go next? Journal of rehabilitation research and development 2007;44(2):245-62. Link: http://www.rehab.research.va.gov/jour/07/44/2/Janke.html.

24. Strine TW, Hootman JM. US national prevalence and correlates of low back and neck pain among adults. Arthritis and rheumatism 2007;57(4):656-65. doi: 10.1002/art.22684 Link: http://onlinelibrary.wiley.com/doi/10.1002/art.22684/abstract;jsessionid=8BAEA2372B89553DE16E56B2FC12617C.f03t02.

25. Deyo RA, Bass JE. Lifestyle and low-back pain. The influence of smoking and obesity. Spine (Phila Pa 1976) 1989;14(5):501-6. Link: http://journals.lww.com/spinejournal/Abstract/1989/05000/Lifestyle_and_Low_Back_Pain__The_Influence_of.5.aspx.

26. Heliovaara M, Makela M, Knekt P, Impivaara O, Aromaa A. Determinants of sciatica and low-back pain. Spine (Phila Pa 1976) 1991;16(6):608-14. Link: http://journals.lww.com/spinejournal/Abstract/1991/06000/Determinants_of_Sciatica_and_Low_Back_Pain_.2.aspx.

27. Power C, Frank J, Hertzman C, Schierhout G, Li L. Predictors of low back pain onset in a prospective British study. Am J Public Health 2001;91(10):1671-8. Link: http://ajph.aphapublications.org/doi/full/10.2105/AJPH.91.10.1671.

28. Hoogendoorn WE, van Poppel MNM, Bongers PM, Koes BW, Bouter LM. Physical load during work and leisure time as risk factors for back pain. Scandinavian journal of work, environment & health 1999:387-403. Link: http://www.jstor.org/stable/10.2307/40966920files/840/40966920.html.

29. Croft PR, Papageorgiou AC, Thomas E, Macfarlane GJ, Silman AJ. Short-term physical risk factors for new episodes of low back pain - Prospective evidence from the South Manchester Back Pain Study. Spine 1999;24(15):1556-61. doi: Doi 10.1097/00007632-199908010-00009 Link: http://journals.lww.com/spinejournal/pages/articleviewer.aspx?year=1999&issue=08010&article=00009&type=abstract.

30. Hildebrandt VH, Bongers' PM, Dul J, van Dijk FJH, Kemper HCG. The relationship between leisure time, physical activities and musculoskeletal symptoms and disability in worker populations. International archives of occupational and environmental health 2000;73(8):507-18. doi: DOI 10.1007/s004200000167 Link: http://link.springer.com/article/10.1007%2Fs004200000167.

31. Chen SM, Liu MF, Cook J, Bass S, Lo SK. Sedentary lifestyle as a risk factor for low back pain: a systematic review. International archives of occupational and environmental health 2009;82(7):797-806. doi: 10.1007/s00420-009-0410-0 Link: http://link.springer.com/article/10.1007%2Fs00420-009-0410-0.

32. Shiri R, Solovieva S, Husgafvel-Pursiainen K, Telama R, Yang X, et al. The role of obesity and physical activity in non-specific and radiating low back pain: the Young Finns study. Semin Arthritis Rheum 2013;42(6):640-50. doi: 10.1016/j.semarthrit.2012.09.002 Link: http://www.ncbi.nlm.nih.gov/pubmed/23270761.

33. Smith GR. The Epidemiology and Treatment of Depression When It Coexists with Somatoform Disorders, Somatization, or Pain. Gen Hosp Psychiat 1992;14(4):265-72. doi: Doi 10.1016/0163-8343(92)90097-T Link: <Go to ISI>://WOS:A1992HY76300007.

34. Sjolie AN. Persistence and change in nonspecific low back pain among adolescents: a 3-year prospective study. Spine (Phila Pa 1976) 2004;29(21):2452-7. Link: http://www.ncbi.nlm.nih.gov/pubmed/15507810.

35. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment 1995;7(4):8. Link: http://psycnet.apa.org/record/1996-10094-001.

36. Kirkwood BR, Sterne JAC. Regression modelling. In: K M, ed. Medical Statistics. USA: Blackwell Publishing company 2003:339-42.

6 Appendix 1: coding details of Health Survey for England variables

6.1 Outcome variables

AnyPain
Pos. = 2181 **Variable =** AnyPain **Variable label =** Whether currently troubled by pain or discomfort
This variable is *numeric*, the SPSS measurement level is *nominal.*
SPSS user missing values = -9 thru -1

Value label information for AnyPain
Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = Yes
Value = 2 Label = No

IF AnyPain = Yes THEN
More3m
Pos. = 2182 **Variable =** More3m **Variable label =** Had pain or discomfort for more than 3 months
This variable is *numeric*, the SPSS measurement level is *nominal.*
SPSS user missing values = -9 thru -1

Value label information for More3m
Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = Yes
Value = 2 Label = No

IF More3m = Yes THEN
SitePain
Where is this pain or discomfort?
CODE ALL THAT APPLY
PROBE: Where else?
1 Back pain
2 Neck or shoulder pain
3 Headache, facial or dental pain
4 Stomach ache or abdominal pain
5 Pain in your arms, hands, hips, legs or feet
6 Chest pain
7 Other pain

SitePai1 was used as it is the one related to back pain.
Pos. = 2183 **Variable =** SitePai1 **Variable label =** Site of pain: Back pain
This variable is *numeric*, the SPSS measurement level is *nominal.*
SPSS user missing values = -9 thru -1

Value label information for SitePai1
Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 0 Label = Not mentioned
Value = 1 Label = Mentioned

Pos. = 2216 **Variable =** CPain **Variable label =** (D) Whether has chronic pain
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1

Value label information for CPain
Value = -9 Label = Refusal

Value = -8 Label = Don't know
Value = -1 Label = Item not applicable
Value = 1 Label = Yes
Value = 2 Label = No

6.1.1 Back Pain severity categories

Respondents experiencing chronic pain were asked to rate their pain and the responses were captured in PainGrade1 and PainGrade2 (See below). .

Pos. = 2214 **Variable =** PainGrade1 **Variable label =** (D) Chronic Pain Grade based on GCPS Version 2.0

This variable is *numeric*, the SPSS measurement level is *scale.*

SPSS user missing values = -9 thru -1

Value label information for PainGrade1
Value = -1 Label = Not applicable
Value = 0 Label = Grade 0
Value = 1 Label = Grade I - Low intensity
Value = 2 Label = Grade II - High intensity
Value = 3 Label = Grade III - Moderately limiting
Value = 4 Label = Grade IV - Severely limiting

PainGrade1 was derived as follows:

```
compute PainGrade1=-5.
if CPI>=0 and CPI<15 and FIDS>=0 and FIDS<17 PainGrade1=1.
if CPI>=15 and FIDS<17 PainGrade1=2.
if FIDS>=17 and FIDS<=24 PainGrade1=3.
if FIDS>=25 PainGrade1=4.
if Anypain=2 PainGrade1=0.
if more3m=2 PainGrade1=0.
if PainGrade1=-5 and more3m=-1 PainGrade1=-1.
exe.
variable labels PainGrade1 '(D) Chronic Pain Grade based on GCPS Version 2.0'.
value labels PainGrade1 -1 'Not applicable'
0 'Grade 0 '
1 'Grade I - Low intensity'
2 'Grade II - High intensity'
3 'Grade III - Moderately limiting'
4 'Grade IV - Severely limiting.
```

Pos. = 2215 **Variable =** PainGrade2 **Variable label =** (D) Chronic Pain Grade based on 3-item GCP-PCS

This variable is *numeric*, the SPSS measurement level is *scale.*

SPSS user missing values = -9 thru -1

Value label information for PainGrade2
Value = -1 Label = Not applicable
Value = 0 Label = Grade 0
Value = 1 Label = Grade I - Low intensity
Value = 2 Label = Grade II - High intensity
Value = 3 Label = Grade III - Moderately limiting
Value = 4 Label = Grade IV - Severely limiting

PainGrade2 was derived as follows:

```
compute PainGrade2=-5.
*if usualp=0 and TIDS=0 PainGrade2=0.
*if usualp=0 and TIDS>=1 and TIDS<9 PainGrade2=1.
*if usualp>=1 and usualp<5 and TIDS<9 PainGrade2=1.
if usualp>=0 and usualp<5 and TIDS>=0 and TIDS<9 PainGrade2=1.
if CPI>=5 and TIDS<9 PainGrade2=2.
if TIDS>=9 and TIDS<=12 PainGrade2=3.
if TIDS>=13 PainGrade2=4.
if Anypain=2 PainGrade2=0.
if more3m=2 PainGrade2=0.
if PainGrade2=-5 and more3m=-1 PainGrade2=-1.
exe.
variable labels PainGrade2 '(D) Chronic Pain Grade based on 3-item GCP-PCS'.
value labels PainGrade2 -1 'Not applicable'
```

0 'Grade 0 '
1 'Grade I - Low intensity'
2 'Grade II - High intensity'
3 'Grade III - Moderately limiting'
4 'Grade IV - Severely limiting.

6.2 HSfE risk factor variables

Age grouped into categories in a newly created **agegrp** variable:
1 if Age<25
2 if Age>=25 and Age<=34
3 if Age>=35 and Age<=44
4 if Age>=45 and Age<=54
5 if Age>=55 and Age<=64
6 if Age>=65 and Age<=74
7 if Age>=75

Pos. = 306 **Variable =** bmivg5 **Variable label =** (D) Valid BMI (grouped:<18.5,18.5-25,25-30,30-40 40+)
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1
Value label information for bmivg5
Value = -9 Label = Refused
Value = -8 Label = Don't know
Value = -7 Label = Refused/not obtained
Value = -6 Label = Schedule not obtained
Value = -2 Label = Schedule not applicable
Value = -1 Label = Not applicable
Value = 1 Label = Under 18.5
Value = 2 Label = 18.5 and below 25
Value = 3 Label = 25 and below 30
Value = 4 Label = 30 and below 40
Value = 5 Label = Over 40

Pos. = 178 **Variable =** econact5 **Variable label =** (D) Economic status (5 groups)
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1
Value label information for econact5
Value = -9 Label = Refusal
Value = -8 Label = Don't know
Value = -1 Label = Not applicable
Value = 1 Label = In employment - employee
Value = 2 Label = In emloyment - self-employed
Value = 3 Label = ILO unemployed
Value = 4 Label = retired
Value = 5 Label = other economically inactive

Pos. = 131 **Variable =** topqual3 **Variable label =** (D) Highest Educational Qualification
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1
Value label information for topqual3
Value = -9 Label = Refused
Value = -8 Label = Don't know
Value = -7 Label = Refused/not obtained
Value = -6 Label = Schedule not obtained
Value = -2 Label = Schedule not applicable
Value = -1 Label = Not applicable
Value = 1 Label = NVQ4/NVQ5/Degree or equiv
Value = 2 Label = Higher ed below degree
Value = 3 Label = NVQ3/GCE A Level equiv
Value = 4 Label = NVQ2/GCE O Level equiv
Value = 5 Label = NVQ1/CSE other grade equiv
Value = 6 Label = Foreign/other
Value = 7 Label = No qualification

Pos. = 186 **Variable =** Origin **Variable label =** Ethnic origin of individual
This variable is *numeric*, the SPSS measurement level is *nominal.*

SPSS user missing values = -9 thru -1

Value label information for Origin

Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = White - English/Welsh/Scottish/Northern Irish/British
Value = 2 Label = White - Irish
Value = 3 Label = White - Gypsy or Irish Traveller
Value = 4 Label = Any other white background
Value = 5 Label = White and Black Caribbean
Value = 6 Label = White and Black African
Value = 7 Label = White and Asian
Value = 8 Label = Any other mixed/multiple ethnic background
Value = 9 Label = Indian
Value = 10 Label = Pakistani
Value = 11 Label = Bangladeshi
Value = 12 Label = Chinese
Value = 13 Label = Any other Asian background
Value = 14 Label = African
Value = 15 Label = Caribbean
Value = 16 Label = Any other Black/African/Caribbean background
Value = 17 Label = Arab
Value = 18 Label = Any other ethnic group (please describe)

Ethn variable was created grouping ethnicity values into 6 groups:

- 'White' 1 if Origin label equal to 1, 2, 3 , 4
- 'Mixed' 2 if Origin label equal to 5, 6, 7, 8
- 'Asian' 3 if Origin label equal to 9, 10, 11, 12, 13
- 'Black/Black Caribbean' 4 if Origin label equal to 14, 15, 16
- 'Other' 5 if Origin label equal to 17, 18
- 'Not stated' 6 if Origin label equal to -9, -8, -2, -1

Pos. = 1901 **Variable =** Anxiety **Variable label =** Anxiety/Depression

This variable is *numeric*, the SPSS measurement level is *nominal.*

SPSS user missing values = -9 thru -1

Value label information for Anxiety

Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = Not anxious or depressed
Value = 2 Label = Moderately anxious or depressed
Value = 3 Label = Extremely anxious or depressed

Pos. = 31 **Variable =** Sex **Variable label =** Sex

This variable is *numeric*, the SPSS measurement level is *nominal.*

SPSS user missing values = -9 thru -1

Value label information for Sex

Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = Male
Value = 2 Label = Female

Pos. = 172 **Variable =** nssec8 **Variable label =** (D) NS-SEC 8 variable classification (individual)

This variable is *numeric*, the SPSS measurement level is *scale.*

SPSS user missing values = -9 thru -1

Value label information for nssec8

Value = -9 Label = Refused
Value = -8 Label = Don't know
Value = -7 Label = Refused/not obtained
Value = -6 Label = Schedule not obtained
Value = -2 Label = Schedule not applicable
Value = -1 Label = Not applicable
Value = 1 Label = Higher managerial and professional occupations
Value = 2 Label = Lower managerial and professional occupations
Value = 3 Label = Intermediate occupations
Value = 4 Label = Small employers and own account workers
Value = 5 Label = Lower supervisory and technical occupations
Value = 6 Label = Semi-routine occupations

Value = 7 Label = Routine occupations
Value = 8 Label = Never worked and long term unemployed
Value = 99 Label = Other

Pos. = 2238 **Variable =** cigsta3 **Variable label =** (D) Cigarette Smoking Status: Current/Ex-Reg/Never-Reg
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1
Value label information for cigsta3
Value = -9 Label = Refused
Value = -8 Label = Don't know
Value = -7 Label = Refused/not obtained
Value = -6 Label = Schedule not obtained
Value = -2 Label = Schedule not applicable
Value = -1 Label = Not applicable
Value = 1 Label = Current cigarette smoker
Value = 2 Label = Ex-regular cigarette smoker
Value = 3 Label = Never regular cigarette smoker

Pos. = 1977 **Variable =** LongIll **Variable label =** Whether has longstanding illness
This variable is *numeric*, the SPSS measurement level is *nominal.*
SPSS user missing values = -9 thru -1
Value label information for LongIll
Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = Yes
Value = 2 Label = No

Pos. = 1978-83 **Variable =** IllsM1-6 **Variable label =** Type of illness - 1st-6th
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1
Value label information for IllsM1
Value = -1 Label = Item not applicable
Value = 1 Label = Cancer (neoplasm)
Value = 2 Label = Diabetes
Value = 3 Label = Other endocrine/metabolic
Value = 4 Label = Mental illness/anxiety/depression/nerves (nes)
Value = 5 Label = Mental handicap
Value = 6 Label = Epilepsy/fits/convulsions
Value = 7 Label = Migraine/headaches
Value = 8 Label = Other problems of nervous system
Value = 9 Label = Cataract/poor eye sight/blindness
Value = 10 Label = Other eye complaints
Value = 11 Label = Poor hearing/deafness
Value = 12 Label = Tinnitus/noises in the ear
Value = 13 Label = Menieres disease/ear complaints causing balance problems
Value = 14 Label = Other ear complaints
Value = 15 Label = Stroke/cerebral haemorrhage/cerebral thrombosis
Value = 16 Label = Heart attack/angina
Value = 17 Label = Hypertension/high blood pressure/blood pressure (nes)
Value = 18 Label = Other heart problems
Value = 19 Label = Piles/haemorrhoids including Varicose Veins in anus
Value = 20 Label = Varicose veins/phlebitis in lower extremities
Value = 21 Label = Other blood vessels/embolic
Value = 22 Label = Bronchitis/emphysema
Value = 23 Label = Asthma
Value = 24 Label = Hayfever
Value = 25 Label = Other respiratory complaints
Value = 26 Label = Stomach ulcer/ulcer (nes)/abdominal hernia/rupture
Value = 27 Label = Other digestive complaints
Value = 28 Label = Complaints of bowel/colon
Value = 29 Label = Complaints of teeth/mouth/tongue
Value = 30 Label = Kidney complaints
Value = 31 Label = Urinary tract infection
Value = 32 Label = Other bladder problems/incontinence
Value = 33 Label = Reproductive system disorders
Value = 34 Label = Arthritis/rheumatism/fibrositis
Value = 35 Label = Back problems/slipped disc/spine/neck
Value = 36 Label = Other problems of bones/joints/muscles
Value = 37 Label = Infectious and parasitic disease
Value = 38 Label = Disorders of blood and blood forming organs and immunity disorders
Value = 39 Label = Skin complaints
Value = 40 Label = Other complaints
Value = 41 Label = Unclassifiable (no other codable complaint)

Value = 42 Label = Complaint no longer present
Value = 97 Label = Not Answered/Refusal
Value = 99 Label = Not Answered / Refusal

Pos. = 1984 **Variable =** LimitAct **Variable label =** Activities limited due to illness
This variable is *numeric*, the SPSS measurement level is *nominal.*
SPSS user missing values = -9 thru -1
Value label information for LimitAct
Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = Yes
Value = 2 Label = No

Pos. = 2147 **Variable =** GenHelf **Variable label =** Self-assessed general health
This variable is *numeric*, the SPSS measurement level is *nominal.*
SPSS user missing values = -9 thru -1
Value label information for GenHelf
Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = ...very good,
Value = 2 Label = good,
Value = 3 Label = fair,
Value = 4 Label = bad, or
Value = 5 Label = very bad?

Pos. = 2148 **Variable =** genhelf2 **Variable label =** (D) Self-assessed general health - grouped
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1
Value label information for genhelf2
Value = 1 Label = Very good/good
Value = 2 Label = Fair
Value = 3 Label = Bad/very bad

Pos. = 1977 **Variable =** LongIll **Variable label =** Whether has longstanding illness
This variable is *numeric*, the SPSS measurement level is *nominal.*
SPSS user missing values = -9 thru -1
Value label information for LongIll
Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = Yes
Value = 2 Label = No

If a person answers 'Yes' to this question, then she/he is asked to mention the illness and the answers are captured in six variables (IllsM1-6).

Pos. = 1978-83 **Variable =** IllsM1-6 **Variable label =** Type of illness - 1st-6th
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1
Value label information for IllsM1
Value = -1 Label = Item not applicable
Value = 1 Label = Cancer (neoplasm)
Value = 2 Label = Diabetes
Value = 3 Label = Other endocrine/metabolic
Value = 4 Label = Mental illness/anxiety/depression/nerves (nes)
Value = 5 Label = Mental handicap
Value = 6 Label = Epilepsy/fits/convulsions
Value = 7 Label = Migraine/headaches
Value = 8 Label = Other problems of nervous system
Value = 9 Label = Cataract/poor eye sight/blindness
Value = 10 Label = Other eye complaints
Value = 11 Label = Poor hearing/deafness
Value = 12 Label = Tinnitus/noises in the ear
Value = 13 Label = Menieres disease/ear complaints causing balance problems
Value = 14 Label = Other ear complaints
Value = 15 Label = Stroke/cerebral haemorrhage/cerebral thrombosis
Value = 16 Label = Heart attack/angina
Value = 17 Label = Hypertension/high blood pressure/blood pressure (nes)
Value = 18 Label = Other heart problems
Value = 19 Label = Piles/haemorrhoids including Varicose Veins in anus

Value = 20	Label = Varicose veins/phlebitis in lower extremities
Value = 21	Label = Other blood vessels/embolic
Value = 22	Label = Bronchitis/emphysema
Value = 23	Label = Asthma
Value = 24	Label = Hayfever
Value = 25	Label = Other respiratory complaints
Value = 26	Label = Stomach ulcer/ulcer (nes)/abdominal hernia/rupture
Value = 27	Label = Other digestive complaints
Value = 28	Label = Complaints of bowel/colon
Value = 29	Label = Complaints of teeth/mouth/tongue
Value = 30	Label = Kidney complaints
Value = 31	Label = Urinary tract infection
Value = 32	Label = Other bladder problems/incontinence
Value = 33	Label = Reproductive system disorders
Value = 34	Label = Arthritis/rheumatism/fibrositis
Value = 35	Label = Back problems/slipped disc/spine/neck
Value = 36	Label = Other problems of bones/joints/muscles
Value = 37	Label = Infectious and parasitic disease
Value = 38	Label = Disorders of blood and blood forming organs and immunity disorders
Value = 39	Label = Skin complaints
Value = 40	Label = Other complaints
Value = 41	Label = Unclassifiable (no other codable complaint)
Value = 42	Label = Complaint no longer present
Value = 97	Label = Not Answered/Refusal
Value = 99	Label = Not Answered / Refusal

Values 34, 35 and 36 were coding for:

Musculo-skeletal - complaints of bones/joints/muscles

34 Arthritis/rheumatism/fibrositis
Arthritis as result of broken limb
Arthritis/rheumatism in any part of the body
Gout *(previously code 03)*
Osteoarthritis, rheumatoid arthritis, polymyalgia rheumatica
Polyarteritis Nodosa *(previously code 21)*
Psoriasis arthritis (also code psoriasis)
Rheumatic symptoms
Still's disease

35 Back problems/slipped disc/spine/neck
Back trouble, lower back problems, back ache
Curvature of spine
Damage, fracture or injury to back/spine/neck
Disc trouble
Lumbago, inflammation of spinal joint
Prolapsed invertebral discs
Schuermann's disease
Spondylitis, spondylosis
Worn discs in spine - affects legs
Exclude if damage/injury to spine results in paralysis - code 06
Sciatica or trapped nerve in spine - code 06

36 Other problems of bones/joints/muscles
Absence or loss of limb eg. lost leg in war, finger amputated, born without arms
Aching arm, stiff arm, sore arm muscle
Bad shoulder, bad leg, collapsed knee cap, knee cap removed
Brittle bones, osteoporosis
Bursitis, housemaid's knee, tennis elbow
Cartilage problems
Chondrodystrophia
Chondromalacia
Cramp in hand
Deformity of limbs eg. club foot, claw-hand, malformed jaw
Delayed healing of bones or badly set fractures
Deviated septum
Dislocations eg. dislocation of hip, clicky hip, dislocated knee/finger
Disseminated lupus
Dupuytren's contraction
Fibromyalgia
Flat feet, bunions,
Fracture, damage or injury to extremities, ribs, collarbone, pelvis, skull, eg. knee injury, broken leg, gun shot wounds in leg/shoulder, can't hold arm out flat - broke it as a child, broken nose
Frozen shoulder
Hip infection, TB hip
Hip replacement (nes)
Legs won't go, difficulty in walking
Marfan Syndrome
Osteomyelitis
Paget's disease
Perthe's disease
Physically handicapped (nes)
Pierre Robin syndrome
Schlatter's disease
Sever's disease
Stiff joints, joint pains, contraction of sinews, muscle wastage
Strained leg muscles, pain in thigh muscles
Systemic sclerosis, myotonia (nes)
Tenosynovitis
Torn muscle in leg, torn ligaments, tendonitis
Walk with limp as a result of polio, polio (nes), after affects of polio (nes)
Weak legs, leg trouble, pain in legs
Muscular dystrophy - code 08

The information from these variables is used to generate **compm1_18** variables:
COMPM1 (D) II Neoplasms & benign growths Derived
COMPM2 (D) III Endocrine & metabolic Derived
COMPM3 (D) V Mental disorders Derived
COMPM4 (D) VI Nervous system Derived
COMPM5 (D) VI Eye complaints Derived
COMPM6 (D) VI Ear complaints Derived
COMPM7 (D) VII Heart & circulatory system Derived
COMPM8 (D) VIII Respiratory system Derived
COMPM9 (D) IX Digestive system Derived

COMPM10 (D) X Genito-urinary system Derived
COMPM11 (D) XII Skin complaints Derived
COMPM12 (D) XIII Musculoskeletal system Derived
COMPM13 (D) I Infectious disease Derived
COMPM14 (D) IV Blood & related organs Derived
COMPM15 (D) Other complaints Derived
COMPM17 (D) No long-standing illness Derived
COMPM18 (D) No longer present Derived
COMPM99 (D) Unclass/NLP/inadeq.describe

6.3 Coding of comorbidity variables

Pos. = 1977 **Variable =** LongIll **Variable label =** Whether has longstanding illness
This variable is *numeric*, the SPSS measurement level is *nominal.*
SPSS user missing values = -9 thru -1

Value label information for LongIll
Value = -9 Label = Refusal
Value = -8 Label = Don't Know
Value = -2 Label = Schedule not applicable
Value = -1 Label = Item not applicable
Value = 1 Label = Yes
Value = 2 Label = No

If a person answers 'Yes' to this question, then she/he is asked to mention the illness and the answers are captured in six variables (IllsM1-6).
Pos. = 1978-83 **Variable =** IllsM1-6 **Variable label =** Type of illness - 1st-6th
This variable is *numeric*, the SPSS measurement level is *scale.*
SPSS user missing values = -9 thru -1

Value label information for IllsM1
Value = -1 Label = Item not applicable
Value = 1 Label = Cancer (neoplasm)
Value = 2 Label = Diabetes
Value = 3 Label = Other endocrine/metabolic
Value = 4 Label = Mental illness/anxiety/depression/nerves (nes)
Value = 5 Label = Mental handicap
Value = 6 Label = Epilepsy/fits/convulsions
Value = 7 Label = Migraine/headaches
Value = 8 Label = Other problems of nervous system
Value = 9 Label = Cataract/poor eye sight/blindness
Value = 10 Label = Other eye complaints
Value = 11 Label = Poor hearing/deafness
Value = 12 Label = Tinnitus/noises in the ear
Value = 13 Label = Menieres disease/ear complaints causing balance problems
Value = 14 Label = Other ear complaints
Value = 15 Label = Stroke/cerebral haemorrhage/cerebral thrombosis
Value = 16 Label = Heart attack/angina
Value = 17 Label = Hypertension/high blood pressure/blood pressure (nes)
Value = 18 Label = Other heart problems
Value = 19 Label = Piles/haemorrhoids including Varicose Veins in anus
Value = 20 Label = Varicose veins/phlebitis in lower extremities
Value = 21 Label = Other blood vessels/embolic
Value = 22 Label = Bronchitis/emphysema
Value = 23 Label = Asthma
Value = 24 Label = Hayfever
Value = 25 Label = Other respiratory complaints
Value = 26 Label = Stomach ulcer/ulcer (nes)/abdominal hernia/rupture
Value = 27 Label = Other digestive complaints
Value = 28 Label = Complaints of bowel/colon
Value = 29 Label = Complaints of teeth/mouth/tongue
Value = 30 Label = Kidney complaints
Value = 31 Label = Urinary tract infection
Value = 32 Label = Other bladder problems/incontinence
Value = 33 Label = Reproductive system disorders
Value = 34 Label = Arthritis/rheumatism/fibrositis
Value = 35 Label = Back problems/slipped disc/spine/neck
Value = 36 Label = Other problems of bones/joints/muscles
Value = 37 Label = Infectious and parasitic disease
Value = 38 Label = Disorders of blood and blood forming organs and immunity disorders
Value = 39 Label = Skin complaints
Value = 40 Label = Other complaints
Value = 41 Label = Unclassifiable (no other codable complaint)
Value = 42 Label = Complaint no longer present
Value = 97 Label = Not Answered/Refusal

Value = 99 Label = Not Answered / Refusal

Values 34, 35 and 36 were coding for:

Musculo-skeletal - complaints of bones/joints/muscles

34 Arthritis/rheumatism/fibrositis
Arthritis as result of broken limb
Arthritis/rheumatism in any part of the body
Gout *(previously code 03)*
Osteoarthritis, rheumatoid arthritis, polymyalgia rheumatica
Polyarteritis Nodosa *(previously code 21)*
Psoriasis arthritis (also code psoriasis)
Rheumatic symptoms
Still's disease

35 Back problems/slipped disc/spine/neck
Back trouble, lower back problems, back ache
Curvature of spine
Damage, fracture or injury to back/spine/neck
Disc trouble
Lumbago, inflammation of spinal joint
Prolapsed invertebral discs
Schuermann's disease
Spondylitis, spondylosis
Worn discs in spine - affects legs
Exclude if damage/injury to spine results in paralysis - code 06
Sciatica or trapped nerve in spine - code 06

36 Other problems of bones/joints/muscles
Absence or loss of limb eg. lost leg in war, finger amputated, born without arms
Aching arm, stiff arm, sore arm muscle
Bad shoulder, bad leg, collapsed knee cap, knee cap removed
Brittle bones, osteoporosis
Bursitis, housemaid's knee, tennis elbow
Cartilage problems
Chondrodystrophia
Chondromalacia
Cramp in hand
Deformity of limbs eg. club foot, claw-hand, malformed jaw
Delayed healing of bones or badly set fractures
Deviated septum
Dislocations eg. dislocation of hip, clicky hip, dislocated knee/finger
Disseminated lupus
Dupuytren's contraction
Fibromyalgia
Flat feet, bunions,
Fracture, damage or injury to extremities, ribs, collarbone, pelvis, skull, eg. knee injury, broken leg, gun shot wounds in leg/shoulder, can't hold arm out flat - broke it as a child, broken nose
Frozen shoulder
Hip infection, TB hip
Hip replacement (nes)
Legs won't go, difficulty in walking
Marfan Syndrome
Osteomyelitis
Paget's disease
Perthe's disease
Physically handicapped (nes)
Pierre Robin syndrome
Schlatter's disease
Sever's disease
Stiff joints, joint pains, contraction of sinews, muscle wastage
Strained leg muscles, pain in thigh muscles
Systemic sclerosis, myotonia (nes)
Tenosynovitis
Torn muscle in leg, torn ligaments, tendonitis
Walk with limp as a result of polio, polio (nes), after affects of polio (nes)
Weak legs, leg trouble, pain in legs
Muscular dystrophy - code 08

The information from these variables is used to generate **compm1_18** variables:
COMPM1 (D) II Neoplasms & benign growths Derived
COMPM2 (D) III Endocrine & metabolic Derived
COMPM3 (D) V Mental disorders Derived
COMPM4 (D) VI Nervous system Derived
COMPM5 (D) VI Eye complaints Derived
COMPM6 (D) VI Ear complaints Derived

COMPM7 (D) VII Heart & circulatory system Derived
COMPM8 (D) VIII Respiratory system Derived
COMPM9 (D) IX Digestive system Derived
COMPM10 (D) X Genito-urinary system Derived
COMPM11 (D) XII Skin complaints Derived
COMPM12 (D) XIII Musculoskeletal system Derived
COMPM13 (D) I Infectious disease Derived
COMPM14 (D) IV Blood & related organs Derived
COMPM15 (D) Other complaints Derived
COMPM17 (D) No long-standing illness Derived
COMPM18 (D) No longer present Derived
COMPM99 (D) Unclass/NLP/inadeq.describe

6.3.1 QOF diseases

Pos. = 1978-83 **Variable =** IllsM1-6 **Variable label =** Type of illness - 1st-6th

This variable is *numeric*, the SPSS measurement level is *scale.*

SPSS user missing values = -9 thru -1

Value label information for IllsM1

Value = -1 Label = Item not applicable
Value = 1 Label = Cancer (neoplasm)
Value = 2 Label = Diabetes
Value = 3 Label = Other endocrine/metabolic
Value = 4 Label = Mental illness/anxiety/depression/nerves (nes)
Value = 5 Label = Mental handicap
Value = 6 Label = Epilepsy/fits/convulsions
Value = 7 Label = Migraine/headaches
Value = 8 Label = Other problems of nervous system
Value = 9 Label = Cataract/poor eye sight/blindness
Value = 10 Label = Other eye complaints
Value = 11 Label = Poor hearing/deafness
Value = 12 Label = Tinnitus/noises in the ear
Value = 13 Label = Menieres disease/ear complaints causing balance problems
Value = 14 Label = Other ear complaints
Value = 15 Label = Stroke/cerebral haemorrhage/cerebral thrombosis
Value = 16 Label = Heart attack/angina
Value = 17 Label = Hypertension/high blood pressure/blood pressure (nes)
Value = 18 Label = Other heart problems
Value = 19 Label = Piles/haemorrhoids including Varicose Veins in anus
Value = 20 Label = Varicose veins/phlebitis in lower extremities
Value = 21 Label = Other blood vessels/embolic
Value = 22 Label = Bronchitis/emphysema
Value = 23 Label = Asthma
Value = 24 Label = Hayfever
Value = 25 Label = Other respiratory complaints
Value = 26 Label = Stomach ulcer/ulcer (nes)/abdominal hernia/rupture
Value = 27 Label = Other digestive complaints
Value = 28 Label = Complaints of bowel/colon
Value = 29 Label = Complaints of teeth/mouth/tongue
Value = 30 Label = Kidney complaints
Value = 31 Label = Urinary tract infection
Value = 32 Label = Other bladder problems/incontinence
Value = 33 Label = Reproductive system disorders
Value = 34 Label = Arthritis/rheumatism/fibrositis
Value = 35 Label = Back problems/slipped disc/spine/neck
Value = 36 Label = Other problems of bones/joints/muscles
Value = 37 Label = Infectious and parasitic disease
Value = 38 Label = Disorders of blood and blood forming organs and immunity disorders
Value = 39 Label = Skin complaints
Value = 40 Label = Other complaints
Value = 41 Label = Unclassifiable (no other codable complaint)
Value = 42 Label = Complaint no longer present
Value = 97 Label = Not Answered/Refusal
Value = 99 Label = Not Answered / Refusal

7 Appendix 2: details of multivariate model fitting

The tables below show the full range of multivariate models we fitted in order to select the final models.

7.1 Overall back pain

Table 33 Multivariate analysis (M1, M2, M3 and M4)

	OR	95% CI	p-value
Total number of respondents	5,072		
Age (agegrp2)			
<34	1.00		
35-44	2.46	[1.85-3.26]	<0.001
45-54	3.06	[2.3-4.05]	<0.001
55-64	3.99	[2.96-5.37]	<0.001
65-74	4.53	[3.08-6.67]	<0.001
Over 75	5.64	[3.69-8.6]	<0.001
Gender			
Male	1.00		
Female	1.46	[1.23-1.72]	<0.001
Ethnicity			
White	1.00		
Mixed	0.47	[0.16-1.34]	0.158
Asian	1.11	[0.77-1.59]	0.574
Black/Black Caribbean	0.79	[0.44-1.43]	0.436
Other	1.53	[0.41-5.7]	0.526
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.97	[0.73-1.28]	0.809
NVQ3/GCE A level equivalent	0.90	[0.68-1.19]	0.469
NVQ2/GCE O level equivalent	0.86	[0.66-1.11]	0.247
NVQ1/CSE other grade equivalent	0.92	[0.6-1.4]	0.688
Foreign/other	0.97	[0.54-1.75]	0.928
No qualification	1.04	[0.78-1.39]	0.766
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.36	[1.01-1.82]	0.041
Intermediate occupations	1.28	[0.92-1.79]	0.149
Small employers and own account workers	1.46	[1-2.13]	0.048
Lower supervisory and technical	1.81	[1.23-2.68]	0.003
Semi-routine occ.	1.34	[0.95-1.88]	0.094
Routine occ.	1.68	[1.18-2.4]	0.004
Never worked and long term unemployed	0.94	[0.45-1.97]	0.861
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.17	[0.97-1.41]	0.098
30 – 39 obese	1.57	[1.28-1.92]	<0.001
>40 obese	3.03	[2.01-4.57]	<0.001
Smoking			

	OR	95% CI	p-value
Current smoker	1.00		
Ex-regular smoker	0.94	[0.75-1.18]	0.623
Never regular	0.76	[0.61-0.93]	0.009
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.85	[2.42-3.36]	<0.001
Extremely anxious or depressed	3.77	[2.45-5.82]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	0.99	[0.66-1.48]	0.952
Retired	1.23	[0.93-1.64]	0.152
Other economically inactive	1.75	[1.36-2.24]	<0.001

The variance inflation factor (vif) command was used after model fitting to check for multicollinearity. Mean variance inflation factor (vif) was 1.65 and none of the individual vifs exceeded 4. Therefore no collinearity was observed between the variables.

Table 34 Multivariate analysis including *limitact* variable (M5)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.18	[1.61-2.95]	<0.001
45-54	2.47	[1.82-3.34]	<0.001
55-64	3.02	[2.18-4.17]	<0.001
65-74	3.71	[2.41-5.69]	<0.001
Over 75	3.46	[2.16-5.55]	<0.001
Gender			
Male	1.00		
Female	1.53	[1.27-1.84]	<0.001
Ethnicity			
White	1.00		
Mixed	0.50	[0.17-1.52]	0.223
Asian	1.03	[0.69-1.53]	0.885
Black/Black Caribbean	0.95	[0.5-1.8]	0.872
Other	1.83	[0.49-6.81]	0.365
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	1.04	[0.76-1.41]	0.825
NVQ3/GCE A level equivalent	0.88	[0.65-1.2]	0.437
NVQ2/GCE O level equivalent	0.84	[0.63-1.12]	0.233
NVQ1/CSE other grade equivalent	1.02	[0.64-1.64]	0.923
Foreign/other	1.08	[0.56-2.09]	0.81
No qualification	1.04	[0.76-1.44]	0.793
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.23	[0.89-1.69]	0.207
Intermediate occupations	1.21	[0.84-1.74]	0.301

	OR	95% CI	p-value
Small employers and own account workers	1.30	[0.86-1.96]	0.217
Lower supervisory and technical	1.70	[1.1-2.61]	0.016
Semi-routine occ.	1.29	[0.89-1.88]	0.172
Routine occ.	1.50	[1.02-2.22]	0.041
Never worked and long term unemployed	1.00	[0.44-2.28]	0.996
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.15	[0.94-1.41]	0.185
30 – 39 obese	1.25	[1-1.57]	0.053
>40 obese	1.82	[1.14-2.9]	0.013
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.09	[0.85-1.41]	0.494
Never regular	0.86	[0.68-1.09]	0.212
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	1.98	[1.64-2.39]	<0.001
Extremely anxious or depressed	1.38	[0.85-2.24]	0.193
Economic status			
In employment	1.00		
ILO Unemployed	0.90	[0.57-1.4]	0.633
Retired	0.76	[0.54-1.05]	0.094
Other economically inactive	0.99	[0.74-1.32]	0.936
Activities limited due to illness			
Yes	1.00		
No	0.09	[0.08-0.11]	<0.001

Table 35 Multivariate analysis including *genhealth* variable (M6)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.29	[1.71-3.07]	<0.001
45-54	2.53	[1.89-3.39]	<0.001
55-64	3.22	[2.36-4.39]	<0.001
65-74	4.18	[2.78-6.29]	<0.001
Over 75	4.47	[2.85-7.03]	<0.001
Gender			
Male	1.00		
Female	1.72	[1.44-2.06]	<0.001
Ethnicity			
White	1.00		
Mixed	0.50	[0.17-1.47]	0.208
Asian	0.84	[0.58-1.24]	0.389
Black/Black Caribbean	0.78	[0.42-1.44]	0.43
Other	1.48	[0.36-6.11]	0.591
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.89	[0.66-1.2]	0.455
NVQ3/GCE A level equivalent	0.83	[0.62-1.11]	0.205
NVQ2/GCE O level equivalent	0.77	[0.58-1.01]	0.059
NVQ1/CSE other grade equivalent	0.77	[0.49-1.22]	0.271
Foreign/other	0.86	[0.46-1.61]	0.64
No qualification	0.86	[0.63-1.17]	0.335
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.31	[0.96-1.77]	0.088
Intermediate occupations	1.14	[0.8-1.62]	0.464
Small employers and own account workers	1.31	[0.88-1.95]	0.185
Lower supervisory and technical	1.64	[1.08-2.47]	0.019
Semi-routine occ.	1.19	[0.83-1.7]	0.337
Routine occ.	1.41	[0.97-2.06]	0.072
Never worked and long term unemployed	0.88	[0.38-2.01]	0.756
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.14	[0.94-1.39]	0.19
30 – 39 obese	1.23	[0.99-1.53]	0.064
>40 obese	1.59	[1.01-2.51]	0.045
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.10	[0.86-1.4]	0.461
Never regular	0.95	[0.76-1.19]	0.674
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	1.93	[1.61-2.31]	<0.001

	OR	95% CI	p-value
Extremely anxious or depressed	1.24	[0.75-2.05]	0.396
Economic status			
In employment	1.00		
ILO Unemployed	0.93	[0.61-1.43]	0.748
Retired	0.81	[0.59-1.11]	0.187
Other economically inactive	1.04	[0.78-1.38]	0.804
Self-assessed general health			
Very good	1.00		
Good	2.08	[1.68-2.57]	<0.001
Fair	5.32	[4.13-6.85]	<0.001
Bad	34.15	[21.31-54.71]	<0.001
Very bad	42.73	[19.12-95.52]	<0.001

Table 36 Multivariate analysis including *limitact* and *genhealth* variables (M7)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.14	[1.57-2.91]	<0.001
45-54	2.33	[1.72-3.17]	<0.001
55-64	2.81	[2.03-3.91]	<0.001
65-74	3.73	[2.41-5.76]	<0.001
Over 75	3.31	[2.04-5.36]	<0.001
Gender			
Male	1.00		
Female	1.65	[1.36-2]	<0.001
Ethnicity			
White	1.00		
Mixed	0.51	[0.17-1.55]	0.235
Asian	0.90	[0.6-1.35]	0.622
Black/Black Caribbean	0.92	[0.49-1.75]	0.804
Other	1.67	[0.43-6.44]	0.456
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.97	[0.71-1.33]	0.864
NVQ3/GCE A level equivalent	0.85	[0.62-1.16]	0.306
NVQ2/GCE O level equivalent	0.79	[0.59-1.05]	0.109
NVQ1/CSE other grade equivalent	0.90	[0.55-1.47]	0.684
Foreign/other	0.99	[0.51-1.94]	0.978
No qualification	0.95	[0.68-1.31]	0.739
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.23	[0.89-1.71]	0.202
Intermediate occupations	1.16	[0.8-1.68]	0.439
Small employers and own account workers	1.26	[0.83-1.92]	0.284
Lower supervisory and technical	1.67	[1.08-2.58]	0.021
Semi-routine occ.	1.25	[0.85-1.82]	0.253
Routine occ.	1.41	[0.94-2.1]	0.095

	OR	95% CI	p-value
Never worked and long term unemployed	0.99	[0.42-2.34]	0.987
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.13	[0.92-1.39]	0.241
30 – 39 obese	1.14	[0.91-1.44]	0.26
>40 obese	1.39	[0.85-2.27]	0.189
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.15	[0.89-1.49]	0.295
Never regular	0.95	[0.75-1.21]	0.684
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	1.68	[1.39-2.05]	<0.001
Extremely anxious or depressed	0.88	[0.52-1.47]	0.613
Economic status			
In employment	1.00		
ILO Unemployed	0.88	[0.56-1.39]	0.581
Retired	0.65	[0.47-0.91]	0.013
Other economically inactive	0.80	[0.58-1.09]	0.15
Activities limited due to illness			
Yes	1.00		
No	0.13	[0.11-0.17]	<0.001
Self-assessed general health			
Very good	1.00		
Good	1.72	[1.38-2.15]	<0.001
Fair	2.22	[1.67-2.95]	<0.001
Bad	8.95	[5.37-14.91]	<0.001
Very bad	10.08	[4.36-23.33]	<0.001

Table 37 Multivariate model including *limitact2* (excluding any related to back pain) (M8)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.39	[1.79-3.18]	<0.001
45-54	2.92	[2.19-3.89]	<0.001
55-64	3.53	[2.6-4.79]	<0.001
65-74	4.09	[2.75-6.09]	<0.001
Over 75	4.35	[2.81-6.74]	<0.001
Gender			
Male	1.00		
Female	1.44	[1.21-1.71]	<0.001
Ethnicity			
White	1.00		
Mixed	0.45	[0.16-1.33]	0.149
Asian	1.08	[0.74-1.56]	0.694
Black/Black Caribbean	0.86	[0.46-1.58]	0.622
Other	1.84	[0.49-6.89]	0.368
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.99	[0.75-1.33]	0.971
NVQ3/GCE A level equivalent	0.89	[0.66-1.18]	0.405
NVQ2/GCE O level equivalent	0.86	[0.66-1.12]	0.26
NVQ1/CSE other grade equivalent	0.94	[0.61-1.45]	0.775
Foreign/other	1.02	[0.56-1.86]	0.953
No qualification	1.05	[0.78-1.41]	0.754
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.35	[1-1.82]	0.052
Intermediate occupations	1.29	[0.92-1.82]	0.142
Small employers and own account workers	1.45	[0.99-2.13]	0.058
Lower supervisory and technical	1.89	[1.26-2.81]	0.002
Semi-routine occ.	1.37	[0.97-1.94]	0.076
Routine occ.	1.65	[1.15-2.37]	0.007
Never worked and long term unemployed	0.92	[0.42-1.99]	0.828
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.16	[0.96-1.41]	0.116
30 – 39 obese	1.42	[1.15-1.75]	0.001
>40 obese	2.35	[1.53-3.61]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00	[0.79-1.27]	0.987
Never regular	0.80	[0.65-1]	0.047
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.44	[2.06-2.9]	<0.001

	OR	95% CI	p-value
Extremely anxious or depressed	2.20	[1.39-3.48]	0.001
Economic status			
In employment	1.00		
ILO Unemployed	0.93	[0.61-1.41]	0.718
Retired	1.00	[0.74-1.35]	0.988
Other economically inactive	1.39	[1.07-1.81]	0.014
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.25	[0.21-0.31]	<0.001

Table 38 Multivariate analysis (including *comorbid*) (M9)

	OR	95% CI	p-value
Total number of respondents	5,056		
Age (agegrp2)			
<34	1.00		
35-44	2.25	[1.68-3.02]	<0.001
45-54	2.50	[1.87-3.35]	<0.001
55-64	2.79	[2.05-3.81]	<0.001
65-74	3.01	[2.01-4.52]	<0.001
Over 75	3.63	[2.33-5.65]	<0.001
Gender			
Male	1.00		
Female	1.51	[1.27-1.8]	<0.001
Ethnicity			
White	1.00		
Mixed	0.48	[0.16-1.41]	0.182
Asian	1.11	[0.76-1.62]	0.58
Black/Black Caribbean	0.95	[0.51-1.75]	0.86
Other	1.87	[0.49-7.1]	0.356
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.99	[0.74-1.33]	0.947
NVQ3/GCE A level equivalent	0.92	[0.69-1.23]	0.583
NVQ2/GCE O level equivalent	0.83	[0.63-1.09]	0.188
NVQ1/CSE other grade equivalent	0.94	[0.6-1.47]	0.787
Foreign/other	1.07	[0.58-1.98]	0.827
No qualification	1.07	[0.79-1.45]	0.667
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.37	[1.01-1.86]	0.042
Intermediate occupations	1.28	[0.9-1.81]	0.171
Small employers and own account workers	1.41	[0.96-2.09]	0.084
Lower supervisory and technical	1.96	[1.3-2.95]	0.001
Semi-routine occ.	1.40	[0.98-1.99]	0.063
Routine occ.	1.73	[1.19-2.51]	0.004
Never worked and long term unemployed	1.09	[0.5-2.36]	0.835
BMI (bmigrp2)			

	OR	95% CI	p-value
<24 normal and underweight	1.00		
25 – 29 overweight	1.08	[0.89-1.31]	0.44
30 – 39 obese	1.25	[1.01-1.54]	0.041
>40 obese	2.07	[1.34-3.19]	0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.99	[0.78-1.25]	0.91
Never regular	0.83	[0.66-1.03]	0.092
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.28	[1.92-2.71]	<0.001
Extremely anxious or depressed	2.20	[1.41-3.43]	0.001
Economic status			
In employment	1.00		
ILO Unemployed	0.97	[0.64-1.49]	0.894
Retired	0.98	[0.73-1.33]	0.919
Other economically inactive	1.35	[1.04-1.75]	0.026
Comorbidities (comorbid)			
No comorbidity	1.00		
Comorbidity present	4.66	[3.93-5.54]	<0.001

Table 39 Multivariate analysis (including *comorbid2*) (M10)

	OR	95% CI	p-value
Total number of respondents	5,056		
Age (agegrp2)			
<34	1.00		
35-44	2.44	[1.84-3.24]	<0.001
45-54	3.00	[2.26-3.98]	<0.001
55-64	3.81	[2.82-5.14]	<0.001
65-74	4.26	[2.89-6.29]	<0.001
Over 75	5.26	[3.43-8.06]	<0.001
Gender			
Male	1.00		
Female	1.46	[1.23-1.73]	<0.001
Ethnicity			
White	1.00		
Mixed	0.46	[0.16-1.33]	0.152
Asian	1.12	[0.78-1.6]	0.556
Black/Black Caribbean	0.83	[0.46-1.51]	0.543
Other	1.58	[0.42-5.91]	0.498
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.98	[0.74-1.3]	0.892
NVQ3/GCE A level equivalent	0.91	[0.69-1.21]	0.522
NVQ2/GCE O level equivalent	0.87	[0.67-1.13]	0.288
NVQ1/CSE other grade equivalent	0.93	[0.61-1.42]	0.731
Foreign/other	1.00	[0.55-1.8]	0.993
No qualification	1.06	[0.79-1.41]	0.694
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.36	[1.01-1.83]	0.04
Intermediate occupations	1.26	[0.9-1.76]	0.172
Small employers and own account workers	1.45	[0.99-2.11]	0.054
Lower supervisory and technical	1.86	[1.25-2.75]	0.002
Semi-routine occ.	1.34	[0.95-1.89]	0.091
Routine occ.	1.68	[1.18-2.4]	0.004
Never worked and long term unemployed	0.93	[0.44-1.97]	0.858
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.15	[0.96-1.39]	0.139
30 – 39 obese	1.51	[1.24-1.86]	<0.001
>40 obese	2.84	[1.88-4.29]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.95	[0.76-1.19]	0.652
Never regular	0.77	[0.62-0.95]	0.015
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.76	[2.34-3.25]	<0.001

	OR	95% CI	p-value
Extremely anxious or depressed	3.43	[2.21-5.3]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	0.98	[0.65-1.47]	0.91
Retired	1.18	[0.88-1.57]	0.27
Other economically inactive	1.68	[1.31-2.16]	<0.001
Comorbidities (comorbid2)			
No comorbidity	1.00		
Comorbidity present	1.31	[1.12-1.55]	0.001

Table 40 Multivariate analysis (including comorbidity count) (M11)

	OR	95% CI	p-value
Total number of respondents	4,909		
Age (agegrp2)			
<34	1.00		
35-44	2.18	[1.62-2.92]	<0.001
45-54	2.29	[1.71-3.08]	<0.001
55-64	2.52	[1.84-3.45]	<0.001
65-74	2.68	[1.77-4.05]	<0.001
Over 75	3.03	[1.92-4.77]	<0.001
Gender			
Male	1.00		
Female	1.52	[1.27-1.82]	<0.001
Ethnicity			
White	1.00		
Mixed	0.56	[0.19-1.61]	0.28
Asian	1.05	[0.72-1.54]	0.81
Black/Black Caribbean	0.92	[0.5-1.72]	0.804
Other	1.89	[0.48-7.42]	0.359
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.98	[0.73-1.33]	0.921
NVQ3/GCE A level equivalent	0.90	[0.67-1.22]	0.509
NVQ2/GCE O level equivalent	0.84	[0.64-1.11]	0.233
NVQ1/CSE other grade equivalent	0.96	[0.61-1.51]	0.85
Foreign/other	1.04	[0.55-1.97]	0.908
No qualification	1.10	[0.81-1.5]	0.53
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.31	[0.96-1.78]	0.085
Intermediate occupations	1.20	[0.84-1.7]	0.315
Small employers and own account workers	1.31	[0.88-1.95]	0.19
Lower supervisory and technical	1.88	[1.24-2.85]	0.003
Semi-routine occ.	1.30	[0.91-1.87]	0.147
Routine occ.	1.57	[1.07-2.29]	0.02
Never worked and long term unemployed	1.00	[0.45-2.22]	0.997
BMI (bmigrp2)			

	OR	95% CI	p-value
<24 normal and underweight	1.00		
25 – 29 overweight	1.07	[0.88-1.3]	0.518
30 – 39 obese	1.14	[0.92-1.42]	0.228
>40 obese	1.65	[1.05-2.6]	0.031
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.99	[0.78-1.26]	0.939
Never regular	0.84	[0.67-1.05]	0.12
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.11	[1.77-2.52]	<0.001
Extremely anxious or depressed	1.55	[0.96-2.51]	0.072
Economic status			
In employment	1.00		
ILO Unemployed	0.99	[0.64-1.51]	0.945
Retired	0.87	[0.64-1.19]	0.398
Other economically inactive	1.28	[0.98-1.68]	0.068
Comorbidity count categories			
No comorbidity	1.00		
1	3.29	[2.71-3.99]	<0.001
2	6.68	[5.26-8.5]	<0.001
3	11.38	[7.89-16.41]	<0.001
4 or more	24.31	[14.03-42.1]	<0.001

Table 41 Multivariate analysis (including *comorbidcountcat2*) (M12)

	OR	95% CI	p-value
Total number of respondents	5,071		
Age (agegrp2)			
<34	1.00		
35-44	2.39	[1.8-3.17]	<0.001
45-54	2.86	[2.16-3.8]	<0.001
55-64	3.63	[2.69-4.91]	<0.001
65-74	4.00	[2.7-5.93]	<0.001
Over 75	4.76	[3.09-7.33]	<0.001
Gender			
Male	1.00		
Female	1.46	[1.23-1.73]	<0.001
Ethnicity			
White	1.00		
Mixed	0.51	[0.18-1.44]	0.201
Asian	1.08	[0.75-1.55]	0.695
Black/Black Caribbean	0.81	[0.44-1.47]	0.486
Other	1.47	[0.38-5.72]	0.576
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.96	[0.72-1.27]	0.764
NVQ3/GCE A level equivalent	0.89	[0.67-1.18]	0.419
NVQ2/GCE O level equivalent	0.88	[0.67-1.14]	0.32
NVQ1/CSE other grade equivalent	0.94	[0.61-1.45]	0.773
Foreign/other	0.96	[0.53-1.75]	0.891
No qualification	1.08	[0.8-1.44]	0.623
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.32	[0.98-1.77]	0.065
Intermediate occupations	1.22	[0.87-1.71]	0.243
Small employers and own account workers	1.38	[0.94-2.02]	0.097
Lower supervisory and technical	1.79	[1.21-2.66]	0.004
Semi-routine occ.	1.29	[0.91-1.81]	0.149
Routine occ.	1.58	[1.1-2.26]	0.013
Never worked and long term unemployed	0.85	[0.4-1.81]	0.677
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.15	[0.95-1.39]	0.141
30 – 39 obese	1.45	[1.18-1.78]	<0.001
>40 obese	2.40	[1.56-3.68]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.95	[0.75-1.19]	0.654
Never regular	0.78	[0.63-0.96]	0.02
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.65	[2.24-3.13]	<0.001

	OR	95% CI	p-value
Extremely anxious or depressed	2.84	[1.8-4.49]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	0.99	[0.66-1.5]	0.976
Retired	1.11	[0.83-1.49]	0.466
Other economically inactive	1.66	[1.29-2.14]	<0.001
Comorbidity count categories			
No comorbidity	1.00		
1	0.93	[0.76-1.14]	0.495
2	1.72	[1.33-2.21]	<0.001
3	3.09	[2.07-4.6]	<0.001
4 or more	5.42	[2.96-9.92]	<0.001

Table 42 Multivariate analysis including *qofdis* (M13)

	OR	95% CI	p-value
Total number of respondents	4,910		
Age (agegrp2)			
<34	1.00		
35-44	2.44	[1.84-3.25]	<0.001
45-54	3.00	[2.26-3.98]	<0.001
55-64	3.86	[2.86-5.2]	<0.001
65-74	4.33	[2.94-6.39]	<0.001
Over 75	5.36	[3.5-8.2]	<0.001
Gender			
Male	1.00		
Female	1.47	[1.24-1.74]	<0.001
Ethnicity			
White	1.00		
Mixed	0.47	[0.16-1.35]	0.161
Asian	1.11	[0.77-1.6]	0.565
Black/Black Caribbean	0.73	[0.39-1.35]	0.312
Other	1.53	[0.41-5.78]	0.529
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.95	[0.72-1.26]	0.723
NVQ3/GCE A level equivalent	0.90	[0.68-1.2]	0.479
NVQ2/GCE O level equivalent	0.86	[0.67-1.12]	0.275
NVQ1/CSE other grade equivalent	0.90	[0.58-1.38]	0.625
Foreign/other	0.99	[0.55-1.78]	0.975
No qualification	1.06	[0.79-1.41]	0.714
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.37	[1.02-1.84]	0.036
Intermediate occupations	1.28	[0.92-1.79]	0.144
Small employers and own account workers	1.46	[1.01-2.13]	0.047
Lower supervisory and technical	1.84	[1.24-2.72]	0.002
Semi-routine occ.	1.34	[0.95-1.88]	0.094
Routine occ.	1.68	[1.17-2.39]	0.005

	OR	95% CI	p-value
Never worked and long term unemployed	0.93	[0.44-1.96]	0.844
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.16	[0.96-1.4]	0.114
30 – 39 obese	1.52	[1.24-1.86]	<0.001
>40 obese	2.83	[1.86-4.28]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.96	[0.76-1.2]	0.7
Never regular	0.77	[0.62-0.95]	0.016
Anxiety/depression			
Not anxious or depressed			
Moderately anxious or depressed	2.78	[2.35-3.28]	<0.001
Extremely anxious or depressed	3.36	[2.16-5.24]	<0.001
Economic status			
In employment			
ILO Unemployed	1.00	[0.66-1.5]	0.986
Retired	1.20	[0.9-1.6]	0.207
Other economically inactive	1.69	[1.32-2.18]	<0.001
QOF diseases			
Absent			
Present	1.28	[1.07-1.53]	0.008

Table 43 Multivariate analysis including obesity (M14)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.52	[1.9-3.34]	<0.001
45-54	3.12	[2.35-4.13]	<0.001
55-64	4.06	[3.02-5.45]	<0.001
65-74	4.59	[3.13-6.74]	<0.001
Over 75	5.67	[3.72-8.63]	<0.001
Gender			
Male	1.00		
Female	1.45	[1.23-1.72]	<0.001
Ethnicity			
White	1.00		
Mixed	0.47	[0.16-1.35]	0.161
Asian	1.10	[0.76-1.58]	0.615
Black/Black Caribbean	0.72	[0.39-1.33]	0.299
Other	1.52	[0.41-5.67]	0.533
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	0.96	[0.73-1.28]	0.789
NVQ3/GCE A level equivalent	0.92	[0.69-1.21]	0.543
NVQ2/GCE O level equivalent	0.87	[0.67-1.13]	0.308
NVQ1/CSE other grade equivalent	0.90	[0.59-1.39]	0.64
Foreign/other	0.97	[0.54-1.75]	0.924
No qualification	1.06	[0.8-1.42]	0.682
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.37	[1.02-1.83]	0.036
Intermediate occupations	1.28	[0.92-1.79]	0.145
Small employers and own account workers	1.47	[1.01-2.14]	0.045
Lower supervisory and technical	1.83	[1.24-2.7]	0.002
Semi-routine occ.	1.34	[0.95-1.88]	0.095
Routine occ.	1.67	[1.17-2.39]	0.004
Never worked and long term unemployed	0.95	[0.45-1.99]	0.884
Obesity			
No	1.00		
Yes	1.55	[1.32-1.83]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.96	[0.76-1.2]	0.708
Never regular	0.77	[0.62-0.95]	0.014
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.86	[2.43-3.37]	<0.001
Extremely anxious or depressed	3.82	[2.48-5.89]	<0.001
Economic status			

	OR	95% CI	p-value
In employment	1.00		
ILO Unemployed	0.99	[0.66-1.49]	0.978
Retired	1.23	[0.92-1.63]	0.159
Other economically inactive	1.75	[1.37-2.24]	<0.001

Table 44 Multivariate analysis (including *obesity* and *qofdis*) (M15)

	OR	95% CI	p-value
Total number of respondents	4,910		
Age (agegrp2)			
<34	1.00		
35-44	2.50	[1.88-3.31]	<0.001
45-54	3.04	[2.3-4.03]	<0.001
55-64	3.91	[2.91-5.26]	<0.001
65-74	4.37	[2.97-6.42]	<0.001
Over 75	5.37	[3.52-8.19]	<0.001
Gender			
Male	1.00		
Female	1.47	[1.24-1.73]	<0.001
Ethnicity			
White	1.00		
Mixed	0.46	[0.16-1.33]	0.154
Asian	1.09	[0.76-1.57]	0.633
Black/Black Caribbean	0.72	[0.39-1.34]	0.299
Other	1.50	[0.4-5.66]	0.546
Not stated	(omitted)		
Education			
NVQ4/NVQ5/Degree or equivalent			
Higher education below degree	0.96	[0.72-1.27]	0.766
NVQ3/GCE A level equivalent	0.92	[0.69-1.21]	0.545
NVQ2/GCE O level equivalent	0.88	[0.68-1.14]	0.32
NVQ1/CSE other grade equivalent	0.91	[0.59-1.41]	0.683
Foreign/other	0.99	[0.55-1.78]	0.976
No qualification	1.07	[0.8-1.43]	0.634
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.38	[1.03-1.85]	0.033
Intermediate occupations	1.28	[0.92-1.79]	0.142
Small employers and own account workers	1.47	[1.01-2.14]	0.045
Lower supervisory and technical	1.85	[1.25-2.73]	0.002
Semi-routine occ.	1.33	[0.95-1.87]	0.098
Routine occ.	1.67	[1.17-2.38]	0.005
Never worked and long term unemployed	0.94	[0.45-1.98]	0.87
Obesity			
Absent	1.00		
Present	1.49	[1.26-1.76]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.97	[0.77-1.21]	0.76

	OR	95% CI	p-value
Never regular	0.78	[0.63-0.96]	0.02
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.78	[2.36-3.28]	<0.001
Extremely anxious or depressed	3.36	[2.16-5.22]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00	[0.66-1.5]	0.994
Retired	1.19	[0.9-1.59]	0.223
Other economically inactive	1.69	[1.31-2.17]	<0.001
QOF diseases			
Absent	1.00		
Present	1.31	[1.09-1.57]	0.003

7.2 Variable selection: overall back pain vs no back pain

In order to obtain the most parsimonious models we then applied stepwise backward and forward variable selection using the *stepwise* command in Stata.

7.2.1 Overall back pain versus no back pain

Table 45 Automatic stepwise forward model (M1)

	OR	95% CI	p-value
Total number of respondents	5,072		
Age (agegrp2)			
<34	1.00		
35-44	2.49	[1.88-3.29]	<0.001
45-54	3.06	[2.32-4.03]	<0.001
55-64	4.01	[3.01-5.33]	<0.001
65-74	4.56	[3.14-6.61]	<0.001
Over 75	5.64	[3.76-8.48]	<0.001
Gender			
Male	1.00		
Female	1.41	[1.21-1.65]	<0.001
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level equivalent	1.00		
NVQ2/GCE O level equivalent	1.00		
NVQ1/CSE other grade equivalent	1.00		
Foreign/other	1.00		
No qualification	1.25	[1.03-1.51]	0.021
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.44	[1.21-1.71]	<0.001
>40 obese	2.79	[1.88-4.14]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.77	[0.66-0.89]	0.001
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.87	[2.44-3.38]	<0.001
Extremely anxious or depressed	3.83	[2.5-5.89]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.22	[0.92-1.62]	0.167
Other economically inactive	1.72	[1.36-2.19]	<0.001

Table 46 Automatic stepwise backward model (M2)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.48	[1.88-3.28]	<0.001
45-54	3.11	[2.36-4.09]	<0.001
55-64	4.45	[3.39-5.84]	<0.001
65-74	5.79	[4.34-7.72]	<0.001
Over 75	7.40	[5.45-10.05]	<0.001
Gender			
Male	1.00		
Female	1.48	[1.26-1.73]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.42	[1.06-1.91]	0.019
Semi-routine occ.	1.00		
Routine occ.	1.35	[1.09-1.67]	0.006
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.44	[1.21-1.7]	<0.001
>40 obese	2.80	[1.89-4.16]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.77	[0.66-0.9]	0.001
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.87	[2.44-3.38]	<0.001
Extremely anxious or depressed	3.94	[2.57-6.04]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.65	[1.31-2.08]	<0.001

Table 47 Automatic stepwise forward and backward model (M3)

	OR	95% CI	p-value
Total number of respondents	5,072		
Age (agegrp2)			
<34	1.00		
35-44	2.48	[1.88-3.28]	<0.001
45-54	3.11	[2.36-4.09]	<0.001
55-64	4.45	[3.39-5.84]	<0.001
65-74	5.79	[4.34-7.72]	<0.001
Over 75	7.40	[5.45-10.05]	<0.001
Gender			
Male	1.00		
Female	1.48	[1.26-1.73]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.42	[1.06-1.91]	0.019
Semi-routine occ.	1.00		
Routine occ.	1.35	[1.09-1.67]	0.006
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.44	[1.21-1.7]	<0.001
>40 obese	2.80	[1.89-4.16]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.77	[0.66-0.9]	0.001
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.87	[2.44-3.38]	<0.001
Extremely anxious or depressed	3.94	[2.57-6.04]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.65	[1.31-2.08]	<0.001

Table 48 Auto stepwise forward and backward model including limitact (M5)

	OR	95% CI	p-value
Total number of respondents	5,071		
Age (agegrp2)			
<34	1.00		
35-44	2.24	[1.66-3.01]	<0.001
45-54	2.54	[1.9-3.41]	<0.001
55-64	2.99	[2.23-4.01]	<0.001
65-74	3.30	[2.42-4.5]	<0.001
Over 75	3.07	[2.19-4.3]	<0.001
Gender			
Male	1.00		
Female	1.47	[1.24-1.75]	<0.001
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.00		
>40 obese	1.65	[1.06-2.58]	0.028
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.66-0.93]	0.005
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	1.97	[1.64-2.37]	<0.001
Extremely anxious or depressed	1.00		
Activities limited due to illness			
Yes	1.00		
No	0.09	[0.07-0.11]	<0.001

Table 49 Auto stepwise forward and backward including genhealth variable (M6)

	OR	95% CI	p-value
Total number of respondents	5,071		
Age (agegrp2)			
<34	1.00		
35-44	2.42	[1.82-3.21]	<0.001
45-54	2.65	[2-3.51]	<0.001
55-64	3.27	[2.47-4.34]	<0.001
65-74	3.77	[2.8-5.06]	<0.001
Over 75	3.93	[2.86-5.41]	<0.001
Gender			
Male	1.00		
Female	1.61	[1.37-1.9]	<0.001
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	1.91	[1.6-2.27]	<0.001
Extremely anxious or depressed	1.00		
Self-assessed general health			
Very good	1.00		
Good	1.00		

	OR	95% CI	p-value
Fair	5.61	[4.42-7.1]	<0.001
Bad	1.00		
Very bad	1.00		

Table 50 Automatic stepwise forward and backward models (with limitact and genhealth) (M7)

	OR	95% CI	p-value
Total number of respondents	5,071		
Age (agegrp2)			
<34	1.00		
35-44	2.20	[1.63-2.97]	<0.001
45-54	2.41	[1.79-3.25]	<0.001
55-64	2.93	[2.14-4.02]	<0.001
65-74	3.95	[2.61-5.99]	<0.001
Over 75	3.50	[2.21-5.54]	<0.001
Gender			
Male	1.00		
Female	1.57	[1.32-1.87]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.23	[1.01-1.48]	0.036
Never regular	1.00		
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	1.71	[1.41-2.06]	<0.001
Extremely anxious or depressed	1.00		
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	0.70	[0.51-0.97]	0.031
Other economically inactive	1.00		
Activities limited due to illness			
Yes	1.00		
No	0.14	[0.11-0.17]	<0.001
Self-assessed general health			
Very good	1.00		
Good	1.75	[1.41-2.17]	<0.001
Fair	2.27	[1.73-2.98]	<0.001
Bad	8.71	[5.36-14.16]	<0.001
Very bad	9.32	[4.12-21.04]	<0.001

Table 51 Automatic forward and backward model (including limitact2) (M8)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.44	[1.84-3.24]	<0.001
45-54	3.02	[2.28-3.99]	<0.001
55-64	3.78	[2.86-4.99]	<0.001
65-74	4.55	[3.38-6.12]	<0.001
Over 75	4.89	[3.55-6.74]	<0.001
Gender			
Male	1.00		
Female	1.45	[1.23-1.71]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.08-1.98]	0.015
Semi-routine occ.			
Routine occ.	1.30	[1.04-1.62]	0.019
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.31	[1.1-1.57]	0.003
>40 obese	2.17	[1.44-3.29]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.79	[0.67-0.92]	0.003
Anxiety/depression			
Not anxious or depressed			
Moderately anxious or depressed	2.46	[2.07-2.91]	<0.001
Extremely anxious or depressed	2.26	[1.43-3.55]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.37	[1.07-1.75]	0.012
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.25	[0.21-0.31]	<0.001

Table 52 Stepwise forward and backward including comorbid (M9)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.24	[1.68-2.98]	<0.001
45-54	2.50	[1.88-3.33]	<0.001
55-64	2.89	[2.17-3.84]	<0.001
65-74	3.22	[2.38-4.36]	<0.001
Over 75	3.90	[2.83-5.38]	<0.001
Gender			
Male	1.00		
Female	1.54	[1.3-1.81]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.51	[1.11-2.07]	0.009
Semi-routine occ.			
Routine occ.	1.36	[1.09-1.7]	0.007
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.20	[1.01-1.44]	0.043
>40 obese	2.00	[1.32-3.03]	0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.82	[0.7-0.97]	0.017
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.30	[1.94-2.73]	<0.001
Extremely anxious or depressed	2.25	[1.45-3.49]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.35	[1.05-1.72]	0.017
Comorbidities (comorbid)			
No comorbidity	1.00		
Comorbidity present	4.64	[3.91-5.51]	<0.001

Table 53 Stepwise forward and backward including comorbid2 (M10)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.46	[1.86-3.25]	<0.001
45-54	3.05	[2.31-4.02]	<0.001
55-64	4.18	[3.18-5.51]	<0.001
65-74	5.25	[3.91-7.04]	<0.001
Over 75	6.61	[4.84-9.05]	<0.001
Gender			
Male	1.00		
Female	1.48	[1.27-1.73]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.09-1.97]	0.012
Semi-routine occ.	1.00		
Routine occ.	1.35	[1.09-1.67]	0.005
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.40	[1.18-1.66]	<0.001
>40 obese	2.64	[1.77-3.93]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.67-0.91]	0.002
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.78	[2.35-3.27]	<0.001
Extremely anxious or depressed	3.55	[2.31-5.47]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.60	[1.27-2.02]	<0.001
Comorbidities (comorbid)			
No comorbidity	1.00		
Comorbidity present	1.32	[1.12-1.55]	0.001

Table 54 Stepwise forward and backward including comorbidity count (comorbidcountcat) (M11)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.20	[1.65-2.93]	<0.001
45-54	2.29	[1.72-3.05]	<0.001
55-64	2.40	[1.8-3.21]	<0.001
65-74	2.33	[1.69-3.2]	<0.001
Over 75	2.55	[1.8-3.6]	<0.001
Gender			
Male	1.00		
Female	1.52	[1.28-1.79]	<0.001
Education			
NVQ4/NVQ5/Degree or equivalent	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level equivalent	1.00		
NVQ2/GCE O level equivalent	1.00		
NVQ1/CSE other grade equivalent	1.00		
Foreign/other	1.00		
No qualification	1.31	[1.07-1.61]	0.008
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.07-2]	0.017
Semi-routine occ.	1.00		
Routine occ.	1.00		
Never worked and long term unemployed	1.00		
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.83	[0.7-0.97]	0.023
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.07	[1.74-2.47]	<0.001
Extremely anxious or depressed	1.00		
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.35	[1.06-1.73]	0.016
Comorbidity count categories			
No comorbidity	1.00		
1	3.33	[2.75-4.04]	<0.001
2	7.18	[5.68-9.08]	<0.001
3	12.19	[8.53-17.42]	<0.001
4 or more	27.61	[16.15-47.22]	<0.001

Table 55 Stepwise forward and backward model including comorbidcountcat2 (M12)

	OR	95% CI	p-value
Total number of respondents	5,071		
Age (agegrp2)			
<34	1.00		
35-44	2.42	[1.83-3.19]	<0.001
45-54	2.92	[2.22-3.85]	<0.001
55-64	3.91	[2.97-5.15]	<0.001
65-74	4.69	[3.49-6.3]	<0.001
Over 75	5.65	[4.12-7.75]	<0.001
Gender			
Male	1.00		
Female	1.48	[1.26-1.73]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.47	[1.09-1.98]	0.012
Semi-routine occ.			
Routine occ.	1.33	[1.07-1.65]	0.01
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight			
25 – 29 overweight			
30 – 39 obese	1.33	[1.12-1.58]	0.001
>40 obese	2.22	[1.47-3.35]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.79	[0.68-0.92]	0.003
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.66	[2.25-3.14]	<0.001
Extremely anxious or depressed	2.92	[1.86-4.59]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.58	[1.25-2.01]	<0.001
Comorbidity count categories			
No comorbidity	1.00		
1	1.00		
2	1.78	[1.39-2.26]	<0.001
3	3.25	[2.21-4.79]	<0.001
4 or more	5.65	[3.1-10.3]	<0.001

Table 56 Stepwise forward and backward model including qofdis (M13)

	OR	95% CI	p-value
Total number of respondents	4,910		
Age (agegrp2)			
<34	1.00		
35-44	2.47	[1.87-3.26]	<0.001
45-54	3.05	[2.31-4.02]	<0.001
55-64	4.28	[3.26-5.63]	<0.001
65-74	5.45	[4.07-7.29]	<0.001
Over 75	6.91	[5.06-9.42]	<0.001
Gender			
Male	1.00		
Female	1.49	[1.27-1.75]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.44	[1.07-1.93]	0.016
Semi-routine occ.	1.00		
Routine occ.	1.34	[1.09-1.66]	0.006
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.39	[1.17-1.65]	<0.001
>40 obese	2.61	[1.75-3.89]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.67-0.91]	0.002
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.80	[2.38-3.3]	<0.001
Extremely anxious or depressed	3.50	[2.26-5.42]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.60	[1.26-2.02]	<0.001
QOF disease			
Absent	1.00		
Present	1.29	[1.08-1.55]	0.006

Table 57 Stepwise backward and forward model including obesity (instead of bmi) (M14)

	OR	95% CI	p-value
Total number of respondents	4,910		
Age (agegrp2)			
<34	1.00		
35-44	2.49	[1.89-3.29]	<0.001
45-54	3.11	[2.36-4.09]	<0.001
55-64	4.43	[3.37-5.81]	<0.001
65-74	5.73	[4.29-7.63]	<0.001
Over 75	7.27	[5.35-9.87]	<0.001
Gender			
Male	1.00		
Female	1.50	[1.28-1.75]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.44	[1.08-1.94]	0.014
Semi-routine occ.	1.00		
Routine occ.	1.34	[1.09-1.66]	0.006
Never worked and long term unemployed	1.00		
Obesity			
Absent	1.00		
Present	1.55	[1.32-1.83]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.67-0.9]	0.001
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.89	[2.46-3.41]	<0.001
Extremely anxious or depressed	3.99	[2.6-6.11]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.65	[1.31-2.08]	<0.001

Table 58 Stepwise backward and forward model including obesity and qofdis (M15)

	OR	95% CI	p-value
Total number of respondents	4,910		
Age (agegrp2)			
<34	1.00		
35-44	2.47	[1.87-3.27]	<0.001
45-54	3.04	[2.31-4.01]	<0.001
55-64	4.25	[3.23-5.58]	<0.001
65-74	5.36	[4.01-7.17]	<0.001
Over 75	6.75	[4.95-9.21]	<0.001
Gender			
Male	1.00		
Female	1.51	[1.29-1.77]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.09-1.96]	0.012
Semi-routine occ.	1.00		
Routine occ.	1.34	[1.08-1.66]	0.007
Never worked and long term unemployed	1.00		
Obesity			
Absent	1.00		
Present	1.49	[1.27-1.76]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.67-0.91]	0.002
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.81	[2.38-3.31]	<0.001
Extremely anxious or depressed	3.50	[2.26-5.41]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.60	[1.27-2.02]	<0.001
QoF disease			
Absent	1.00		
Present	1.31	[1.1-1.57]	0.003

Table 59 Stepwise backward and forward model including comorbid2 and limitact2 (M16)

	OR	95% CI	p-value
Total number of respondents	4,894		
Age (agegrp2)			
<34	1.00		
35-44	2.47	[1.86-3.27]	<0.001
45-54	3.13	[2.37-4.15]	<0.001
55-64	4.16	[3.14-5.51]	<0.001
65-74	5.25	[3.88-7.09]	<0.001
Over 75	5.51	[3.99-7.63]	<0.001
Gender			
Male	1.00		
Female	1.45	[1.23-1.7]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.42	[1.05-1.93]	0.023
Semi-routine occ.	1.00		
Routine occ.	1.27	[1.02-1.59]	0.033
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight			
25 – 29 overweight			
30 – 39 obese	1.37	[1.14-1.63]	0.001
>40 obese	2.29	[1.51-3.48]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.67-0.92]	0.002
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.55	[2.14-3.03]	<0.001
Extremely anxious or depressed	2.39	[1.51-3.78]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.37	[1.07-1.75]	0.012
Comorbidity (excl. related to back pain)			
Absent	1.00		
Present	0.48	[0.38-0.61]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.14	[0.11-0.19]	<0.001

Table 60 Stepwise forward and backward model including qofdis and limitact2 (M17)

	OR	95% CI	p-value
Total number of respondents	4,909		
Age (agegrp2)			
<34	1.00		
35-44	2.44	[1.84-3.24]	<0.001
45-54	3.02	[2.28-4]	<0.001
55-64	3.78	[2.86-5]	<0.001
65-74	4.56	[3.39-6.13]	<0.001
Over 75	4.90	[3.56-6.75]	<0.001
Gender			
Male	1.00		
Female	1.45	[1.23-1.71]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.08-1.98]	0.015
Semi-routine occ.	1.00		
Routine occ.	1.30	[1.04-1.62]	0.019
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.31	[1.1-1.57]	0.002
>40 obese	2.18	[1.44-3.29]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.79	[0.68-0.93]	0.004
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.46	[2.08-2.92]	<0.001
Extremely anxious or depressed	2.27	[1.44-3.57]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.36	[1.07-1.74]	0.013
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.25	[0.21-0.31]	<0.001

Table 61 Stepwise backward and forward including qofdis and limitact2 BUT excluding econact (M18)

	OR	95% CI	p-value
Total number of respondents	4,911		
Age (agegrp2)			
<34	1.00		
35-44	2.41	[1.82-3.2]	<0.001
45-54	2.96	[2.24-3.92]	<0.001
55-64	3.71	[2.81-4.9]	<0.001
65-74	4.32	[3.23-5.79]	<0.001
Over 75	4.63	[3.37-6.35]	<0.001
Gender			
Male	1.00		
Female	1.48	[1.26-1.74]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.08-1.98]	0.014
Semi-routine occ.	1.00		
Routine occ.	1.35	[1.09-1.68]	0.006
Never worked and long term unemployed	1.00		
BMI (bmigrp2)			
<24 normal and underweight	1.00		
25 – 29 overweight	1.00		
30 – 39 obese	1.32	[1.11-1.57]	0.002
>40 obese	2.18	[1.44-3.3]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.79	[0.68-0.93]	0.003
Anxiety/depression			
Not anxious or depressed			
Moderately anxious or depressed	1.00		
Extremely anxious or depressed	2.46	[1.58-3.85]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.25	[0.2-0.3]	<0.001

Table 62 Stepwise forward and backward including qofdis, limitact2 and bmigrp3 BUT excluding econact

	OR	95% CI	p-value
Total number of respondents	4,911		
Age (agegrp2)			
<34	1.00		
35-44	2.41	[1.82-3.2]	<0.001
45-54	2.95	[2.23-3.9]	<0.001
55-64	3.68	[2.79-4.86]	<0.001
65-74	4.27	[3.18-5.72]	<0.001
Over 75	4.54	[3.31-6.23]	<0.001
Gender			
Male	1.00		
Female	1.49	[1.27-1.75]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.48	[1.09-2]	0.012
Semi-routine occ.	1.00		
Routine occ.	1.35	[1.08-1.67]	0.007
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 25 normal weight	1.00		
25 – 29 overweight	1.00		
>30 obese	1.40	[1.18-1.65]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.79	[0.68-0.93]	0.004
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.50	[2.11-2.97]	<0.001
Extremely anxious or depressed	2.47	[1.58-3.85]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.24	[0.2-0.29]	<0.001

Table 63 Stepwise forward and backward model including qofdis and limitact2 and bmigrp3 (M20)

	OR	95% CI	p-value
Total number of respondents	4,909		
Age (agegrp2)			
<34	1.00		
35-44	2.45	[1.84-3.24]	<0.001
45-54	3.01	[2.27-3.98]	<0.001
55-64	3.75	[2.84-4.96]	<0.001
65-74	4.50	[3.35-6.05]	<0.001
Over 75	4.81	[3.49-6.62]	<0.001
Gender			
Male	1.00		
Female	1.46	[1.25-1.72]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.47	[1.09-1.99]	0.012
Semi-routine occ.	1.00		
Routine occ.	1.30	[1.04-1.62]	0.021
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.39	[1.18-1.65]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.79	[0.68-0.93]	0.004
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.36	[1.07-1.74]	0.013
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.47	[2.08-2.93]	<0.001
Extremely anxious or depressed	2.27	[1.44-3.58]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.25	[0.21-0.3]	<0.001

Table 64 Autostepwise forward and backward model including qofdis and bmigrp3 (excluded limitact2) (M21)

	OR	95% CI	p-value
Total number of respondents	4,910		
Age (agegrp2)			
<34	1.00		
35-44	2.47	[1.87-3.27]	<0.001
45-54	3.04	[2.31-4.01]	<0.001
55-64	4.25	[3.23-5.58]	<0.001
65-74	5.36	[4.01-7.17]	<0.001
Over 75	6.75	[4.95-9.21]	<0.001
Gender			
Male	1.00		
Female	1.51	[1.29-1.77]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.09-1.96]	0.012
Semi-routine occ.	1.00		
Routine occ.	1.34	[1.08-1.66]	0.007
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.49	[1.27-1.76]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.67-0.91]	0.002
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.60	[1.27-2.02]	<0.001
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.81	[2.38-3.31]	<0.001
Extremely anxious or depressed	3.50	[2.26-5.41]	<0.001
Any of QOF diseases			
Absent	1.00		
Present	1.31	[1.1-1.57]	0.003

Table 65 Autostepwise forward and backward (bmigrp3) (M22)

	OR	95% CI	p-value
Total number of respondents	4,910		
Age (agegrp2)			
<34	1.0		
35-44	2.49	[1.89-3.29]	<0.001
45-54	3.11	[2.36-4.09]	<0.001
55-64	4.43	[3.37-5.81]	<0.001
65-74	5.73	[4.29-7.63]	<0.001
Over 75	7.27	[5.35-9.87]	<0.001
Gender			
Male	1.00		
Female	1.50	[1.28-1.75]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.44	[1.08-1.94]	0.014
Semi-routine occ.	1.00		
Routine occ.	1.34	[1.09-1.66]	0.006
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.55	[1.32-1.83]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.67-0.9]	0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.65	[1.31-2.08]	<0.001
Anxiety/depression			
Not anxious or depressed	1.00		
Moderately anxious or depressed	2.89	[2.46-3.41]	<0.001
Extremely anxious or depressed	3.99	[2.6-6.11]	<0.001

Table 66 Autostepwise forward and backward model (bmigrp3) BUT excluding anxiety/depression (M23)

	OR	95% CI	p-value
Total number of respondents	5,267		
Age (agegrp2)			
<34	1.00		
35-44	2.56	[1.96-3.34]	<0.001
45-54	3.17	[2.44-4.12]	<0.001
55-64	4.43	[3.41-5.74]	<0.001
65-74	5.34	[4.06-7.01]	<0.001
Over 75	6.71	[5.04-8.94]	<0.001
Gender			
Male	1.00		
Female	1.56	[1.34-1.82]	<0.001
Ethnicity			
White	1.00		
Mixed	1.00		
Asian	1.00		
Black/Black Caribbean	1.00		
Other	3.34	[1.41-7.89]	0.006
Not stated	1.00		
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.34	[1.03-1.74]	0.031
Intermediate occupations	1.37	[1.03-1.82]	0.031
Small employers and own account workers	1.54	[1.11-2.14]	0.009
Lower supervisory and technical	1.92	[1.37-2.7]	<0.001
Semi-routine occ.	1.44	[1.09-1.9]	0.01
Routine occ.	1.95	[1.48-2.57]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.54	[1.33-1.8]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.75	[0.65-0.86]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.95	[1.58-2.42]	<0.001

Table 67 Autostepwise backward and forward excluding depression BUT including limitact2 (M24)

	OR	95% CI	p-value
Total number of respondents	5,265		
Age (agegrp2)			
<34	1.00		
35-44	2.47	[1.88-3.24]	<0.001
45-54	2.97	[2.27-3.88]	<0.001
55-64	3.64	[2.78-4.76]	<0.001
65-74	4.10	[3.09-5.44]	<0.001
Over 75	4.13	[3.06-5.59]	<0.001
Gender			
Male	1.00		
Female	1.51	[1.28-1.77]	<0.001
Ethnicity			
White	1.00		
Mixed	1.00		
Asian	1.00		
Black/Black Caribbean	1.00		
Other	3.99	[1.67-9.53]	0.002
Not stated	1.00		
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.32	[1.01-1.73]	0.046
Intermediate occupations	1.34	[1-1.8]	0.048
Small employers and own account workers	1.55	[1.1-2.17]	0.011
Lower supervisory and technical	1.92	[1.35-2.72]	<0.001
Semi-routine occ.	1.44	[1.08-1.91]	0.013
Routine occ.	1.83	[1.37-2.44]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.37	[1.17-1.61]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.78	[0.67-0.91]	0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.51	[1.2-1.9]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.23	[0.19-0.27]	<0.001

Table 68 Auto-stepwise forward and backward model/excluded ethnicity (M24b)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.42	[1.85-3.17]	<0.001
45-54	2.97	[2.28-3.88]	<0.001
55-64	3.63	[2.78-4.74]	<0.001
65-74	4.12	[3.11-5.46]	<0.001
Over 75	4.19	[3.1-5.66]	<0.001
Gender			
Male	1.00		
Female	1.54	[1.32-1.8]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.1-1.95]	0.009
Semi-routine occ.	1.00		
Routine occ.	1.39	[1.14-1.71]	0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.38	[1.18-1.62]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.77	[0.66-0.89]	0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.54	[1.23-1.93]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.23	[0.19-0.27]	<0.001

Table 69 Autostepwise forward and backward excluding depression and including limitact2 and genhealth (M25)

	OR	95% CI	p-value
Total number of respondents	5,265		
Age (agegrp2)			
<34	1.00		
35-44	2.25	[1.7-2.98]	<0.001
45-54	2.41	[1.82-3.18]	<0.001
55-64	3.13	[2.33-4.19]	<0.001
65-74	4.24	[2.91-6.17]	<0.001
Over 75	3.72	[2.46-5.63]	<0.001
Gender			
Male	1.00		
Female	1.69	[1.44-1.99]	<0.001
Ethnicity			
White	1.00		
Mixed	1.00		
Asian	1.00		
Black/Black Caribbean	1.00		
Other	4.78	[1.96-11.63]	0.001
Not stated	1.00		
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.00		
Semi-routine occ.	1.00		
Routine occ.	1.00		
Never worked and long term unemployed	0.50	[0.25-1]	0.049
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	3.79	[1.32-10.91]	0.014
25 – 29 overweight	4.11	[1.43-11.83]	0.009
>30 obese	4.42	[1.54-12.71]	0.006
Smoking	Automatically dropped		
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	0.73	[0.54-0.97]	0.032
Other economically inactive	1.00		
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.46	[0.38-0.57]	<0.001
Self-assessed general health			
Very good	1.00		
Good	2.11	[1.72-2.59]	<0.001

	OR	95% CI	p-value
Fair	4.96	[3.89-6.32]	<0.001
Bad	28.32	[18.48-43.4]	<0.001
Very bad	36.99	[17.97-76.15]	<0.001

Table 70 Autostepwise excluding depression (M26)

	OR	95% CI	p-value
Total number of respondents	5,265		
Age (agegrp2)			
<34	1.00		
35-44	2.56	[1.96-3.34]	<0.001
45-54	3.17	[2.44-4.12]	<0.001
55-64	4.43	[3.41-5.74]	<0.001
65-74	5.34	[4.06-7.01]	<0.001
Over 75	6.71	[5.04-8.94]	<0.001
Gender			
Male	1.00		
Female	1.56	[1.34-1.82]	<0.001
Ethnicity			
White	1.00		
Mixed	1.00		
Asian	1.00		
Black/Black Caribbean	1.00		
Other	3.34	[1.41-7.89]	0.006
Not stated	1.00		
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.34	[1.03-1.74]	0.031
Intermediate occupations	1.37	[1.03-1.82]	0.031
Small employers and own account workers	1.54	[1.11-2.14]	0.009
Lower supervisory and technical	1.92	[1.37-2.7]	<0.001
Semi-routine occ.	1.44	[1.09-1.9]	0.01
Routine occ.	1.95	[1.48-2.57]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.54	[1.33-1.8]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.75	[0.65-0.86]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.95	[1.58-2.42]	<0.001

Table 71 Stepwise backward and forward including depression (derived) (M27)

	OR	95% CI	p-value
Total number of respondents	5,265		
Age (agegrp2)			
<34	1.00		
35-44	2.49	[1.91-3.24]	<0.001
45-54	3.12	[2.4-4.05]	<0.001
55-64	4.39	[3.38-5.69]	<0.001
65-74	5.35	[4.07-7.02]	<0.001
Over 75	6.76	[5.08-9]	<0.001
Gender			
Male	1.00		
Female	1.59	[1.37-1.85]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.46	[1.11-1.93]	0.007
Semi-routine occ.	1.00		
Routine occ.	1.46	[1.2-1.77]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.54	[1.32-1.8]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.74	[0.64-0.86]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.00		
Other economically inactive	1.85	[1.49-2.3]	<0.001
Depression			
Absent	1.00		
Present	1.80	[1.31-2.48]	<0.001

7.4 Variable selection: severe back pain

The tables below show the details of the models fitted for severe back pain.

Table 72 Stepwise model excluding depression but including limitact2 (1A)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.04	[1.36-3.05]	0.001
45-54	2.69	[1.82-3.96]	<0.001
55-64	2.82	[1.88-4.22]	<0.001
65-74	2.52	[1.52-4.17]	<0.001
Over 75	2.57	[1.51-4.37]	0.001
Gender			
Male	1.00		
Female	1.55	[1.25-1.92]	<0.001
Ethnicity			
White	1.00		
Mixed	1.00		
Asian	1.00		
Black/Black Caribbean	1.00		
Other	3.45	[1.21-9.89]	0.021
Not stated	1.00		
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.00		
Semi-routine occ.	1.44	[1.11-1.86]	0.006
Routine occ.	1.57	[1.19-2.07]	0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.00		
>30 obese	1.43	[1.16-1.77]	0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.66	[0.54-0.81]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.53	[1.06-2.22]	0.023
Other economically inactive	2.05	[1.53-2.75]	<0.001

	OR	95% CI	p-value
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.21	[0.17-0.26]	<0.001
Education			
NVQ4/NVQ5 degree	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level	1.00		
NVQ2/GCE O level	1.40	[1.07-1.82]	0.013
NVQ1/CSE other grade	1.00		
Foreign/other	1.00		
No qualification	1.65	[1.26-2.16]	<0.001

Table 73 Stepwise model excluding depression but including limitact2 (1B)

	OR	95% CI	p-value
Total number of respondents			
Age (agegrp2)			
<34	1.00		
35-44	2.37	[1.65-3.41]	<0.001
45-54	3.03	[2.13-4.32]	<0.001
55-64	3.07	[2.12-4.46]	<0.001
65-74	2.94	[1.85-4.67]	<0.001
Over 75	2.91	[1.78-4.77]	<0.001
Gender			
Male	1.00		
Female	1.79	[1.47-2.18]	<0.001
Ethnicity			
White	1.00		
Mixed	1.00		
Asian	1.00		
Black/Black Caribbean	1.00		
Other	6.21	[2.4-16.04]	<0.001
Not stated	1.00		
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.94	[1.37-2.76]	<0.001
Semi-routine occ.	1.48	[1.16-1.89]	0.002
Routine occ.	1.71	[1.31-2.23]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.30	[1.03-1.64]	0.027
>30 obese	1.76	[1.39-2.23]	<0.001

	OR	95% CI	p-value
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.68	[0.56-0.82]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.46	[1.03-2.05]	0.031
Other economically inactive	2.34	[1.79-3.06]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.18	[0.15-0.23]	<0.001
Education			
NVQ4/NVQ5 degree	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level	1.00		
NVQ2/GCE O level	1.00		
NVQ1/CSE other grade	1.00		
Foreign/other	1.00		
No qualification	1.46	[1.16-1.84]	0.001

Table 74 Stepwise model excluding depression and ethnicity but including limitact2 (2A)

	OR	95% CI	p-value
Total number of respondents	5,260		
Age (agegrp2)			
<34	1.00		
35-44	2.04	[1.36-3.05]	0.001
45-54	2.69	[1.83-3.98]	<0.001
55-64	2.82	[1.88-4.22]	<0.001
65-74	2.51	[1.52-4.15]	<0.001
Over 75	2.57	[1.51-4.38]	<0.001
Gender			
Male	1.00		
Female	1.55	[1.25-1.91]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.00		
Semi-routine occ.	1.42	[1.1-1.84]	0.008
Routine occ.	1.56	[1.19-2.06]	0.002
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		

	OR	95% CI	p-value
25 – 29 overweight	1.00		
>30 obese	1.44	[1.17-1.77]	0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.67	[0.54-0.82]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.53	[1.06-2.22]	0.024
Other economically inactive	2.11	[1.58-2.82]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.21	[0.17-0.26]	<0.001
Education			
NVQ4/NVQ5 degree	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level	1.00		
NVQ2/GCE O level	1.39	[1.06-1.81]	0.016
NVQ1/CSE other grade	1.00		
Foreign/other	1.00		
No qualification	1.66	[1.27-2.17]	<0.001

Table 75 Stepwise model excluding depression and ethnicity but including limitact2 (2B)

	OR	95% CI	p-value
Total number of respondents	5,260		
Age (agegrp2)			
<34	1.00		
35-44	2.36	[1.64-3.39]	<0.001
45-54	3.03	[2.13-4.32]	<0.001
55-64	3.05	[2.1-4.42]	<0.001
65-74	2.91	[1.83-4.62]	<0.001
Over 75	2.91	[1.78-4.76]	<0.001
Gender			
Male	1.00		
Female	1.78	[1.46-2.17]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		
Lower supervisory and technical	1.92	[1.36-2.73]	<0.001
Semi-routine occ.	1.46	[1.14-1.86]	0.003
Routine occ.	1.70	[1.31-2.21]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			

	OR	95% CI	p-value
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.30	[1.03-1.64]	0.025
>30 obese	1.78	[1.41-2.25]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.69	[0.57-0.83]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.46	[1.04-2.05]	0.03
Other economically inactive	2.43	[1.87-3.17]	<0.001
Activities limited due to illness (excluding if comorbidity due to back pain is present)			
Yes	1.00		
No	0.19	[0.15-0.23]	<0.001
Education			
NVQ4/NVQ5 degree	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level	1.00		
NVQ2/GCE O level	1.00		
NVQ1/CSE other grade	1.00		
Foreign/other	1.00		
No qualification	1.47	[1.17-1.85]	0.001

Table 76 Autostepwise excluding depression (4B)

	OR	95% CI	p-value
Total number of respondents	5,261		
Age (agegrp2)			
<34	1.00		
35-44	2.59	[1.82-3.7]	<0.001
45-54	3.41	[2.41-4.82]	<0.001
55-64	3.83	[2.66-5.5]	<0.001
65-74	3.37	[2.15-5.3]	<0.001
Over 75	4.34	[2.69-6.99]	<0.001
Gender			
Male	1.00		
Female	1.75	[1.44-2.12]	<0.001
Ethnicity			
White	1.00		
Mixed	1.00		
Asian	1.00		
Black/Black Caribbean	1.00		
Other	5.21	[2.04-13.29]	0.001
Not stated	1.00		
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.45	[1-2.09]	0.048

	OR	95% CI	p-value
Intermediate occupations	1.65	[1.13-2.42]	0.01
Small employers and own account workers	1.67	[1.07-2.61]	0.025
Lower supervisory and technical	2.59	[1.67-4]	<0.001
Semi-routine occ.	1.99	[1.38-2.88]	<0.001
Routine occ.	2.46	[1.69-3.59]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.32	[1.06-1.65]	0.014
>30 obese	2.08	[1.66-2.61]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	0.77	[0.6-0.99]	0.045
Never regular	0.57	[0.45-0.71]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.99	[1.44-2.76]	<0.001
Other economically inactive	3.36	[2.62-4.33]	<0.001
Education			
NVQ4/NVQ5 degree	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level	1.00		
NVQ2/GCE O level	1.00		
NVQ1/CSE other grade	1.00		
Foreign/other	1.00		
No qualification	1.45	[1.16-1.8]	0.001

Table 77 Stepwise model excluding ethnicity but including depression (derived) (6)

	OR	95% CI	p-value
Total number of respondents	5,260		
Age (agegrp2)			
<34	1.00		
35-44	2.48	[1.74-3.54]	<0.001
45-54	3.28	[2.32-4.64]	<0.001
55-64	3.70	[2.58-5.3]	<0.001
65-74	3.30	[2.11-5.16]	<0.001
Over 75	4.18	[2.61-6.7]	<0.001
Gender			
Male	1.00		
Female	1.81	[1.5-2.19]	<0.001
Socioeconomic status			
Higher managerial and professional	1.00		
Lower managerial and professional	1.00		
Intermediate occupations	1.00		
Small employers and own account workers	1.00		

	OR	95% CI	p-value
Lower supervisory and technical	1.84	[1.31-2.57]	<0.001
Semi-routine occ.	1.38	[1.09-1.74]	0.008
Routine occ.	1.72	[1.34-2.21]	<0.001
Never worked and long term unemployed	1.00		
BMI (bmigrp3)			
<18.4	1.00		
18.5 - 24	1.00		
25 – 29 overweight	1.29	[1.03-1.61]	0.025
>30 obese	2.00	[1.6-2.51]	<0.001
Smoking			
Current smoker	1.00		
Ex-regular smoker	1.00		
Never regular	0.67	[0.56-0.8]	<0.001
Economic status			
In employment	1.00		
ILO Unemployed	1.00		
Retired	1.91	[1.38-2.63]	<0.001
Other economically inactive	3.11	[2.42-4.01]	<0.001
Education			
NVQ4/NVQ5 degree	1.00		
Higher education below degree	1.00		
NVQ3/GCE A level	1.00		
NVQ2/GCE O level	1.00		
NVQ1/CSE other grade	1.00		
Foreign/other	1.00		
No qualification	1.48	[1.19-1.84]	<0.001
Depression (derived)			
Absent	1.00		
Present	2.10	[1.49-2.98]	<0.001

8 Appendix 3: synthetic estimation using Stata

8.1 Synthetic estimates

The proportion of our population according to age and sex is known. The proportion by educational status can be applied to these numbers, taking account of the fact that the distribution by educational status differs by age group. This gives estimated proportion by age, sex and educational status. This information is reflected in the variables below (variable names starting m_noed_, m_othed, m_nqv_, f_noed, f_othed, f_nqv).

Within stata, a new set of variables is created, one for each combination of these risk factors pertinent to the logistic regression model for the chosen disease. For instance, if there are two binary variables for age group included in the regression model, then there are three relevant age groups (those the first variable=1, those with the second variable=1, and those where both variables=0 – it is not possible to have both variables=1 since this would imply being in two separate age groups at the same time). With a binary variable for gender included, we would need groups for each gender – but some models don't include gender, like the one we are using in this illustration. With one binary variable for educational status included in our predictive model, there are 2 categories for education (those with and those without this specified educational status, which here is no qualifications). The total number of combinations of age/ sex/ education groups then becomes 3x1x2=6. Corresponding to these 6 categories we have 6 variables as follows, which are created by summing sub categories (the categories that have equivalent risk within the model in question):

```
gen agegp_23_edu7_1_0=m_noed_4564+ f_noed_4564
gen agegp_23_edu7_1_1=m_noed_6574+ f_noed_6574
gen agegp_23_edu7_1_2=m_noed_75p+ f_noed_75p
gen agegp_23_edu7_1_3=f_othed_4564 +f_nvq_4564+ m_othed_4564+ m_nvq_4564
gen agegp_23_edu7_1_4=f_othed_6574 +f_nvq_6574+ m_othed_6574+ m_nvq_6574
gen agegp_23_edu7_1_5=f_othed_75p +f_nvq_75p+ m_othed_75p+ m_nvq_75p
```

These are calculated based on 3 initial education groups – those with no education (_noed_ variables), those with NQVs only (_nvq_ variables) and those with other education (_othed_), although in this model there is no distinction between those with NVQs and other education (since only the binary variable for no education is included in the model). There is also no distinction between males (variables names starting m_) and females (starting f_). There is distinction between each of the three age groups (45-64, 65-74 and 75 plus), since both binary variables for age categories are included in this model.

Of course, they could be calculated in any way convenient, provided the result is the anticipated proportion in each age/ sex/ educational group, pertinent to the model in hand. They can be named in any convenient way, providing each has the same name, apart from having a different number at the end. This allows use of the reshape command in Stata.

In practice, we do not want to find a synthetic estimate on just one population, but rather on many populations, for instance on each local authority separately. We have a data set containing information on the risk factors in all the different local authorities (LAs) and also other regions, with one line of data per region. The above variables give the proportions for each specified combination of age/ sex/ education categories. There are other variables giving the proportions by each additional risk factor separately (e.g. the proportion of non-smokers, current smokers and ex-smokers).

	la_code	agegp_~0	agegp_~1	agegp_~2	agegp_~3	agegp_~4	agegp_~5
1.	E06000001	.1508832	.1071422	.0981529	.4574982	.098431	.0878925
2.	E06000002	.1188522	.0935188	.0856422	.4995597	.1080905	.0943367
3.	E06000003	.1720217	.1538003	.1252882	.4062224	.0801852	.0624822
4.	E06000004	.0819255	.0819462	.0705499	.5388392	.1242868	.1024525
5.	E06000005	.1694909	.1373918	.1226214	.425029	.0791328	.066334

A reshape long command on the set of 6 agegp_23_edu7_1_ variables (as defined above) is used as follows:

```
reshape long agegp_23_edu7_1_, i(ccg_code) j(agegp_23_edu7_1)
```

this gives 6 lines of data per region (since in this example there are 6 categories of age/ sex/ educational status, and 6 corresponding variables) from the starting place of one line of data per region. As well as a variable defining the categories (agegp_23_edu7_1 as named in the j() part of the above command), we now have a variable giving the proportion in each row of data (called variable agegp_23_edu7_1_ note that this name ends in _). These proportions were originally 6 variables on each line, and now we have 6 separate lines for each region. If you look at the data listing above, the row of proportions turns into a column of 6 proportions, then the second row becomes a column of another 6 proportions below the first six, against the second LA code. The i() part of the command gives a unique identifier for each line of data.

	la_code	agegp_~1	agegp~1_	paleve~0	paleve~1	paleve~2	paleve~3
1.	E06000001	0	.1508832	.3476278	.0815955	.0734759	.4973008
2.	E06000001	1	.1071422	.3476278	.0815955	.0734759	.4973008
3.	E06000001	2	.0981529	.3476278	.0815955	.0734759	.4973008
4.	E06000001	3	.4574982	.3476278	.0815955	.0734759	.4973008
5.	E06000001	4	.098431	.3476278	.0815955	.0734759	.4973008
6.	E06000001	5	.0878925	.3476278	.0815955	.0734759	.4973008
7.	E06000002	0	.1188522	.3012445	.1088107	.067936	.5220087
8.	E06000002	1	.0935188	.3012445	.1088107	.067936	.5220087
9.	E06000002	2	.0856422	.3012445	.1088107	.067936	.5220087
10.	E06000002	3	.4995597	.3012445	.1088107	.067936	.5220087
11.	E06000002	4	.1080905	.3012445	.1088107	.067936	.5220087
12.	E06000002	5	.0943367	.3012445	.1088107	.067936	.5220087

For a risk factor, such as smoking status, where the number by age, sex and other risk factors is not known, the proportion of smokers and of ex-smokers in the population is applied to each age/ sex/ educational status group. Another such variable is physical activity (PA) level (palevel), which is in 4 categories, so has 3 corresponding binary variables, all of which are included in this predictive logistic regression model. This is the next one dealt with in practice.

Four relevant variables are created as follows for PA level, with the requirement that they all have the same name, except for the different numbers at the end, as follows:

```
gen palevelf_0=1-pa_low-pa_mod - pa_high
gen palevelf_1=pa_low
gen palevelf_2=pa_mod
gen palevelf_3=pa_high
```

(derived from pa_ variables for low, moderate and high physical activity levels).

With those 4 variables, a further reshape long command can be applied. Note that we already have 6 lines of data per region. This gives 4 lines of data (one for each PA level) from each line, which gives 6x4=24 lines of data per region now. The i() part of the command that gives the unique identifier now needs to include the age/sex/ education categories variable (agegp_23_edu7_1) as well as the region coding variable (ccg_code). The j() part tells Stata to name the newly created categorical variable palevelf, which represents the different PA level categories. The palevelf_ variable (note _ at end of this name) gives the proportion within each PA level category (these add to one for each la_code/ agegp_23_edu7_1 combination, i.e. for each set of 4 lines – again the 4 values that are listed horizontally above are now listed vertically into this palevelf_ column).

```
reshape long palevelf_, i(ccg_code agegp_23_edu7_1) j(palevelf)
```

	la_code	agegp_~1	palevelf	agegp~1_	paleve~_	bmicat~0	bmicat~2	bmicat~3
1.	E06000001	0	0	.1508832	.3476278	.315303	.3785597	.3061373
2.	E06000001	0	1	.1508832	.0815955	.315303	.3785597	.3061373
3.	E06000001	0	2	.1508832	.0734759	.315303	.3785597	.3061373
4.	E06000001	0	3	.1508832	.4973008	.315303	.3785597	.3061373
5.	E06000001	1	0	.1071422	.3476278	.315303	.3785597	.3061373
6.	E06000001	1	1	.1071422	.0815955	.315303	.3785597	.3061373
7.	E06000001	1	2	.1071422	.0734759	.315303	.3785597	.3061373
8.	E06000001	1	3	.1071422	.4973008	.315303	.3785597	.3061373
9.	E06000001	2	0	.0981529	.3476278	.315303	.3785597	.3061373
10.	E06000001	2	1	.0981529	.0815955	.315303	.3785597	.3061373
11.	E06000001	2	2	.0981529	.0734759	.315303	.3785597	.3061373
12.	E06000001	2	3	.0981529	.4973008	.315303	.3785597	.3061373
13.	E06000001	3	0	.4574982	.3476278	.315303	.3785597	.3061373
14.	E06000001	3	1	.4574982	.0815955	.315303	.3785597	.3061373
15.	E06000001	3	2	.4574982	.0734759	.315303	.3785597	.3061373
16.	E06000001	3	3	.4574982	.4973008	.315303	.3785597	.3061373
17.	E06000001	4	0	.098431	.3476278	.315303	.3785597	.3061373
18.	E06000001	4	1	.098431	.0815955	.315303	.3785597	.3061373
19.	E06000001	4	2	.098431	.0734759	.315303	.3785597	.3061373
20.	E06000001	4	3	.098431	.4973008	.315303	.3785597	.3061373
21.	E06000001	5	0	.0878925	.3476278	.315303	.3785597	.3061373
22.	E06000001	5	1	.0878925	.0815955	.315303	.3785597	.3061373
23.	E06000001	5	2	.0878925	.0734759	.315303	.3785597	.3061373
24.	E06000001	5	3	.0878925	.4973008	.315303	.3785597	.3061373

Similarly for other risk factors, for this model, the other risk factors are BMI (obese, overweight and not overweight categories), smoking (where only ex-smoking is relevant, smokers and non-smokers are combined), gym membership and socio-economic status (with 3 relevant binary variables, giving 4 categories). Therefore for this model, there are
6 x 4 x 3 x 2 x 2 x 4 = 1152 different combinations of predictor variables. With 6 different “reshape long” commands in total, we end up with 1152 lines of data per region.

The weights for each region can be obtained by multiplying the relevant proportions together.
Weight = (proportion in specified age/ sex/ education category) x (proportion by PA level) x (proportion by BMI group) x (proportion by smoking status) x (proportion by gym membership) x (proportion by relevant socio-economic status group).

gen xyz= agegp_23_edu7_1_ * palevelf_* bmicatf2_* smokef2_* hobby1_* ssec8_

These weights (“xyz”) will sum to one for each region. It is a good idea to check that they do so in practice.

For practical purposes, so that we can use Stata efficiently, it is also necessary to create all the binary variables used in the original logistic regression modelling, used to derive our preferred predictive model with regression coefficients. The names and coding of these variables must be identical to those used in the original data set.

The most complex task is recreating age, education and sex variables, since they are combined above for the purposes of the reshape command. For the model in our example, we do not need a sex variable, but we do need the following variables (check with the above commands which define them initially to make sure the appropriate codings are used – the tab2 command below also allows for some checking):

```
gen agegp2=agegp_23_edu7_1==1 | agegp_23_edu7_1==4
gen agegp3=agegp_23_edu7_1==2 | agegp_23_edu7_1==5
gen educ7=agegp_23_edu7_1==0 | agegp_23_edu7_1==1 | agegp_23_edu7_1==2
tab2 agegp_23_edu7_1 agegp2 agegp3 educ7, missing
```

For other variables, such as PA level, it is straight forward to create the required binary variables (the tab2 command again allows for some checking):

```
gen palevelf1=palevelf==1
gen palevelf2=palevelf==2
gen palevelf3=palevelf==3
tab2 palevelf palevelf1 palevelf2 palevelf3, missing
```

Note, on creation of above variables: the right hand-side are expressions, such as palevelf==1 – the variable is coded as =1 when this is true and =0 when this is false and including for missing values of palevelf (here we excluded any data with missing values earlier so this does not apply).

With our dataset set up in this way, we can now use Stata's "predict" command to give us the predicted log odds. For this to work, the last regression that we have undertaken in Stata must be the definitive predictive logistic regression equation for the chosen disease, which requires the dataset used to derive that to be in Stata's memory at the time. When we use the "predict" command, we need the dataset described above (after all the above described transformations) to be in Stata's memory, since that gives the characteristics of the regions on which we want the synthetic estimates. It would also be possible to programme in the linear equation from the logistic regression manually, but I have not done that, since there is then more scope for errors.

The predict commands gives predicted log odds, and we then find prevalences as follows: exp(log odds) / [1+ exp(log odds)]
Then we find the weighted average of these, averaged across all possible combinations of risk factors, using the weights calculated as above (stored in variable named xyz). Using Stata, the weighted average can be found using the "collapse" command as follows, which results in one line of data per region (using a region identifier as the by() variable).

```
predict pred_values, xb
gen pred_OR=exp(pred_values)
gen pred_prev=pred_OR/(1+pred_OR)
gen wt_pred_prev=pred_prev*xyz
collapse (sum) wt_pred_prev xyz, by(ccg_code)
```

Thus the region is a data set with one line of data per region, with an estimate of prevalence against each region, based on the definitive logistic regression equation.

8.2 Calculating confidence intervals for prevalence estimates using bootstrap procedures

There is uncertainty in these synthetic estimates of prevalence based on the imprecision in the estimated coefficients from the logistic regression equations. A bootstrap procedure can be used to construct confidence intervals on these synthetic estimates of prevalence, based on the imprecision in these logistic regression coefficients.

Boot-strap procedures

The philosophy underlying the boot-strap procedure is to consider that the people included in the data set used to derive the logistic regression equation represent the whole population of possible people. However, the whole population is effectively considered to contain thousands of copies of each of these people.

Bootstrap samples are taken from our initial populations (the subsets of the ELSA population that has complete data on appropriate risk factors). The first person to be included in our new bootstrap data set is chosen at random from our starting (ELSA) dataset, with each person being equally likely to be chosen. Then the second person to be included in this bootstrap data set is chosen at random in the same way, again with each person being equally likely to be chosen. It is noteworthy that the second person to be chosen could be

the same person as the person selected first (with probability 1/n where n=sample size, the same probability that any individual will be selected). We then select a 3rd person for our bootstrap sample, then a 4th, 5th, 6th, and so on, up to an nth person (where n is the size of our starting dataset). We are effectively selecting at random "with replacement", which means that the same person can be selected twice, or indeed many times. (This is why I say that the population is effectively considered to have many copies of each person in it).

Therefore the bootstrap data are the same size (same number of people in it) as the original dataset used to derive the logistic regression model. It is theoretically possible (though extremely unlikely) that a bootstrap data set could be identical to that original dataset. However, it is far more likely that there will be differences, since some people will be included in the bootstrap data set twice or more, and many are not included at all, although many would also be included just once.

Logistic regression of the same risk factors can then be applied to this bootstrap sample, i.e. we rerun the logistic regression that gave us our chosen predictive model. However, we get slightly different regression coefficients, because of the modified sample. Prevalence estimates are then derived for each combination of risk factors, based on these new regression equations.

This process is repeated 1,000 times, to find 1,000 different bootstrap samples, by random sampling processes, and to then fit logistic regression equations on each. The prevalence estimates are calculated for each combination of risk factors, for each of these 1,000 bootstrap samples. For each region, a synthetic estimate is calculated for each bootstrap sample, by appropriately weighting the prevalence estimates on each combination of risk factors (with the same weights as described above which reflect the anticipated prevalence of each combination of risk factors in the region). From these 1,000 synthetic estimates of prevalence of each region, a 95% confidence interval is calculated as the 2.5th to 97.5th centiles. Given that the estimates are distributed normally, these are taken to be mean +/- 1.96 SD (taking mean and SD of the 1,000 bootstrap synthetic prevalence estimates for each specified region).

The following commands describe how this is done in Stata:

```
forvalues j=1/1000 {
use bootstrap, clear
                                (NB line above reads in original version of the data use used for logistic regrn eqn)
gen howmany=0
forvalues i=1/11516 {
local nn=floor(uniform()*11516)+1
quietly replace howmany=howmany+1 if nnn==`nn'
}
```

nn is a random variable, derived from a uniform random variable which takes values between 0 and 1, to give a random variable between 1 and the total sample size.
The variable "howmany" records how many times each individual has been selected (for the specific bootstrap sample)

```
drop if howmany==0
expand howmany
```

The above 2 lines drop any people that have not been selected in our sample, and then repeat lines (twice or more) of any that have been selected twice or more.

```
quietly logit kneecategory2 agegp2 agegp3 palevelf1 palevelf2 palevelf3 smokef2 bmicatf22 bmicatf23 educ7
ssec8_5 ssec8_6 ssec8_7 hobby1 [pweight=10*probwtks]
```

The above lines run the chosen logistic regression on this bootstrap sample of data, to get new estimates of regression coefficients.

use temp0, clear

The above reads in data set of all possible combination of risk factors, for purposes of calculating confidence intervals
*** the saved data set has 1 extra variable, so storing the extra bootstrapped estimate
predict est`j', xb
save temp0, replace

}

To get bootstrap confidence intervals on specific regions, we need to firstly find predicted prevalences from these predicted log odds (by taking exp(log odds)/ [1+exp(log odds)] for each bootstrap estimate.

forvalues j=0/1000 {
gen prev`j'=exp(est`j')/(1+exp(est`j'))
}

Remember we are working on a data set with one line of data for each combination of risk factors. We then need to merge this data set, with the data set which gives appropriate weighted for each combination of risk factors for each region (which has many lines of data per region, 1152 for severe knee OA model).

merge agegp2 agegp3 palevelf1 palevelf2 palevelf3 smokef2 bmicatf22 bmicatf23 educ7 ssec8_5 ssec8_6 ssec8_7 hobby1 using prevalences0

(This above commands lists each risk factor binary variable in the model as a variable that we are merging on).

For each bootstrap sample, the synthetic prevalence estimate in any population is found by applying the same weights as above, according to the expected proportion of that population with any specified combination of risk factors (as follows – use of collapse command means that we conveniently end up with one line of data per patient).

forvalues j=0/1000 {
gen wt_prev`j'=xyz*prev`j'
}
collapse (sum) xyz wt_prev* (mean) c_pt45p c_tot_mf_ages, by(ccg_code)

This gives 1,000 different synthetic estimates of prevalence for each population, one for each of the bootstrapped samples of data. The confidence interval is found on these by taking the 2.5^{th} and 97.5^{th} centiles. Alternatively, the confidence interval can be found by taking the mean and SD of these prevalence estimates, and taking the mean +/- 1.96 SDs. [In practice, for estimates of severe knee OA, both these sets of estimates agreed very well, suggesting that the distribution of these estimates approximates very closely to the normal distribution – therefore the second method, using mean +/-1.96 SD, is a bit more precise]

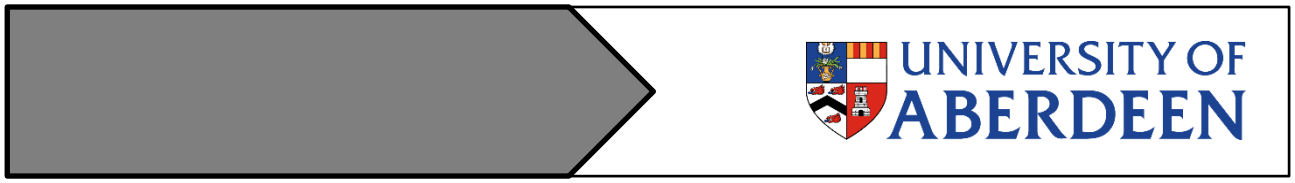
egen meanr=rowmean(wt_prev1-wt_prev1000)
egen p2_5r=rowpctile(wt_prev1-wt_prev1000), p(2.5)
egen p97_5r=rowpctile(wt_prev1-wt_prev1000), p(97.5)
egen medianr=rowpctile(wt_prev1-wt_prev1000), p(50)
egen sdr=rowsd(wt_prev1-wt_prev1000)

Why is it not possible to put confidence intervals separately on each combination of risk factors? It is possible, but then averaging these would not agree to finding confidence intervals directly on appropriately weighted average prevalences of these, appropriate to specific populations. So that would not be a possible way forward with our objectives here.

Why is it necessary to divide the data into different groups by each combination of risk factors, rather simply taking account of the overall distribution of risk factors in the population? The weighted average of prevalences for a person with "average" risk factors is not the same as the weighted average prevalence, across all combinations of risk factors (appropriately weighted). The latter is what we want, and what we calculate directly.

This approach would work to find the appropriately weighted averaged log odds, since this a linear combination of risk factors. However, there is then a change of scale, taking the exponential to get the odds ratio, and then transforming again to get the prevalences.

9 Appendix 4: external validation of Health Survey for England 2011 models for low back pain: data from UK-BioBank Scotland



EPIDEMIOLOGY RESEARCH REPORT 2016b

Report title	**Validation of Arthritis Research UK 'Health Survey for England 2011' models of risk factors for low back pain: data from UK-BioBank**	
Date	September 2016	
Authors	Dr Gareth Jones and Professor Gary Macfarlane	
Contact	Professor Gary Macfarlane	Epidemiology Group School of Medicine, Medical Sciences and Nutrition University of Aberdeen Polwarth Building, Foresterhill Aberdeen, AB25 2ZD
	E-mail	g.j.macfarlane@abdn.ac.uk
	Tel	01224 437 143

9.1 BACKGROUND

Imperial College London have produced models describing risk factors for back pain, using data from the Health Survey for England 2011. In these data, a general population sample was taken, designed to be representative of the whole population (adults and children). In 2011, from approximately 9,000 addresses randomly selected, 5,715 adults and 1,257 children provided data from questionnaire plus a nurse visit. Two final models were produced, one for back pain overall, and one for severe back pain. The model for back pain overall is shown in Table 78. . However, this model needed to be validated in an external dataset.

Table 78: Health Survey for England – Risk factors for back pain in adults (N = 5,269)

		OR[a]	95%CI[a]	P value
Age (years)				
	<34	1.00		
	35 to 44	2.44	1.88-3.18	<0.001
	45 to 54	3.00	2.32-3.89	<0.001
	55 to 64	4.07	3.14-5.27	<0.001
	65 to 74	4.34	3.29-5.72	<0.001
	>75	5.36	3.99-7.2	<0.001
Gender				
	Male	1.00		
	Female	1.66	1.44-1.93	<0.001
Socioeconomic status				
	Higher managerial and professional	1.00		
	Lower managerial and professional	1.00		
	Intermediate occupations	1.00		
	Small employers and own account workers	1.00		
	Lower supervisory and technical	1.00		
	Semi-routine occupation	1.42	1.07-1.87	0.014
	Routine occupation	1.00		
	Never worked and long term unemployed	1.47	1.2-1.8	<0.001
BMI (kgm^{-2})				
	<18.4	1.00		
	18.5 to 24	1.00		
	25 to 29	1.00		
	>30	1.57	1.34-1.82	<0.001
Smoking				
	Current smoker	1.00		
	Ex-regular smoker	1.00		
	Never regular	0.74	0.64-0.86	<0.001
Education				
	Any qualification	1.00		
	No qualification	1.27	1.05-1.52	0.012
a	Odds ratios and confidence intervals are presented to the degree of precision provided in the original e-mails to the University of Aberdeen.			

The aim of the current analysis is to validate the model(s) from the Health Survey for England, above, using data from UK-BioBank.

9.2 Methods

Detailed methods used by UK Biobank are available elsewhere.[1] In brief, between 2006 and 2010 the study recruited 503,325 people, aged 40-69yrs, living within ~25 miles of one of 22 assessment centres, across Great Britain. Participants completed touch-screen questionnaires, which included information on pain, lifestyle and environment factors.

9.2.1 Outcome=back pain

Participants were asked: 'In the last month have you experienced any of the following that interfered with your usual activities?' and were provided with a list of possible pain sites:

(i) Headache
(ii) Facial pain
(iii) Neck or shoulder pain
(iv) Back pain
(v) Stomach or abdominal pain
(vi) Hip pain
(vii) Knee pain
(viii) Pain all over the body
(ix) None of the above
(x) Prefer not to answer

Responses (viii) to (x) were exclusive, such that only these responses could be recorded, or participants could record one, or more than one, responses (i) to (vii) For each site for which pain was reported, participants were asked: 'Have you had <<this pain>> for more than three months?', where 'this pain' was substituted for each individual pain, in turn.

For the current analysis, participants with chronic back pain were defined as those reporting either:

(i) Back pain in the last month, or
(ii) Pain all over the body (which, by definition, includes back pain)

which has persisted for more than three months.

9.2.2 Independent variables

All independent variables were categorised in a manner most similar to the way specified in the Health Survey for England Data (Table 79). However, due to differences in data collection, this was not always possible. Table 79 details how UK-BioBank questionnaire responses have been mapped to Health Survey for England responses, for analysis.

Table 79: categorisation of age, BMI, smoking and education from Health Survey for England and working definition employed in UK-BioBank

Health Survey for England	UK-BioBank [a]
Age group	
<34 yrs	[11]
35 to 44 yrs	<45 yrs
45 to 54 yrs	>=45 to <55 yrs
55 to 64 yrs	>=55 to <65 yrs
65 to 74 yrs	>=65 yrs
>75 yrs	[a]

[11] UK-BioBank, by design, has no participants in these categories.

Health Survey for England	UK-BioBank [a]
BMI[12]	
<18.4 kgm^{-2}	<18.5 kgm^{-2}
18.5 to 24 kgm^{-2}	>=18.5 to <25.0 kgm^{-2}
25 to 29 kgm^{-2}	>=25.0 to <30.0 kgm^{-2}
>30 kgm^{-2}	>=30.0 kgm^{-2}
Smoking	
Current smoker	Current smoker
Ex-smoker	Ex-regular smoker
Never regular	Never smoker, plus (a) Ex 'tried it' smoker, and (b) Ex 'occasional' smoker
Qualifications	
Any qualification	One, or more, of: (a) College or university degree, or (b) NVQ / HND / HNC, or equivalent (c) A / AS-level, or equivalent (d) O-level, or equivalent (e) CSE, or equivalent (f) Other professional qualification (e.g. nursing, teaching)
No qualification	None of the above

In the Health Survey for England, socioeconomic status was assessed using the National Statistics Socio-economic Classification.[2] This classifies individuals into eight categories, based on aspects of their employment and occupation:

(i) Higher managerial and professional
(ii) Lower managerial and professional
(iii) Intermediate occupations
(iv) Small employers and own account workers
(v) Lower supervisory and technical
(vi) Semi-routine occupation
(vii) Routine occupation
(viii) Never worked and long term unemployed

This information was not available in UK-BioBank. Thus, in the current analysis, socioeconomic status was determined using a combination of employment status and Townsend Index. For the former, participants were asked: 'Which of the following describes your current situation?' and were provided with a list of seven possible responses:

(i) In paid employment or self-employed
(ii) Retired
(iii) Looking after home and/or family
(iv) Unable to work because of sickness or disability
(v) Unemployed
(vi) Doing unpaid or voluntary work
(vii) Full or part-time student

The latter, the Townsend Index[3], an area-based index of deprivation, based upon unemployment; car ownership; home ownership; and household overcrowding, and divided in deciles for analysis.

[12] We have interpreted BMI categories we believe they would have been operationalised

Analysis

All independent variables were cross-tabulated against chronic low back pain. Data are presented as simple frequencies, with percentages of back pain (positive versus negative) for each category of exposure. These associations were then quantified using logistic regression. Thus, effects are presented as odds ratios (OR), with 95% confidence intervals (95%CI).
All independent variables were then entered into a simple (i.e. non-stepwise) multiple logistic regression model, specified as closely as possible to Table 1, notwithstanding variations in coding already described. Again, effect estimates are presented as odds ratios, with 95% confidence intervals. Finally, the independent contribution of each individual variable to the final model was assessed using a Wald statistic.

9.3 Results

Data was available on 500,434 individuals (45.6% male; mean age 56.5yrs (SD=8.1)). Only 39.4% of participants (N=197,146) reported having no pain in the past month that interfered with usual activities. Of the remainder, the majority (N=294,443; 97.1%) reported one or more regional pains and 8,845 (2.9%) reported pain all over the body. In total, including those with pain all over the body, 138,963 (27.8%) participants reported back pain, of whom 95,454 (68.7%) reported this pain to have lasted for than three months. Thus, the prevalence of for chronic low back pain – the primary outcome for the current analysis – is shown in Table 80.

Table 80: prevalence of chronic low back pain in UK-BioBank

Chronic low back pain	N	%
Yes	95,454	19.1%
No	404,980	80.9%

All exposures of interest were associated with the occurrence of chronic low back pain (Table 81). Chronic low back pain was more prevalent among older age groups (OR>=65 versus <45 yrs: 1.18; 95%CI: 1.15-1.21) and in women (1.07; 1.06-1.09). Those with no formal educational qualifications were more likely to report chronic low back pain (1.69; 1.66-1.72) and there was evidence of an increase in prevalence with increasing deprivation: those living in the most deprived areas were nearly twice as likely to report chronic low back pain than those in the most affluent areas (1.89; 1.83-1.95).

Table 81: factors associated with chronic low back pain – univariate associations

		Chronic low back pain [a]		OR	95%CI
		Yes	No		
Age (years)					
	<45 yrs	9,093	42,427	1.00	–
	>=45 to <55 yrs	26,302	115,527	1.06	1.03-1.09
	>=55 to <65 yrs	40,812	170,738	1.12	1.09-1.14
	>=65 yrs	19,247	76,288	1.18	1.15-1.21
Gender					
	Male	42,194	185,933	1.00	–
	Female	53,260	219,047	1.07	1.06-1.09
Socioeconomic status – Employment status					
	In paid employment or self-employed	46,001	240,769	1.00	–
	Retired	34,057	132,517	1.35	1.32-1.37

		Chronic low back pain [a]		OR	95%CI
		Yes	**No**		
	Looking after home and/or family	2,865	10,933	1.37	1.31-1.43
	Unable to work because of sickness or disability	9,068	7,659	6.20	6.00-6.40
	Unemployed	1,762	6,450	1.43	1.36-1.51
	Doing unpaid or voluntary work	393	1,920	1.07	0.96-1.19
	Full or part-time student	252	1,084	1.22	1.06-1.40
Socio-economic status – Townsend Index (deciles)					
	1 (most affluent)	8,071	42,059	1.00	–
	2	8,394	41,950	1.04	1.008-1.08
	3	8,411	41,422	1.06	1.02-1.09
	4	8,685	41,402	1.09	1.06-1.13
	5	8,951	41,085	1.14	1.10-1.17
	6	9,173	40,846	1.17	1.13-1.21
	7	9,269	40,753	1.19	1.15-1.22
	8	9,989	39,930	1.30	1.26-1.35
	9	11,196	38,679	1.51	1.46-1.56
	10 (most deprived)	13,182	36,364	1.89	1.83-1.95
Body mass index					
	<18.5 kgm^{-2}	411	2,200	1.00	–
	>=18.5 to <25.0 kgm^{-2}	24,383	137,578	1.05	0.85-1.17
	>=25.0 to <30.0 kgm^{-2}	39,311	172,152	1.29	1.27-1.31
	>=30.0 kgm^{-2}	31,349	93,050	1.90	1.87-1.94
Smoking status					
	Current smoker	10,563	28,536	1.00	–
	Ex-smoker	25,841	94,929	0.74	0.72-0.75
	Never regular	58,584	280,183	0.56	0.55-0.58
Education					
	Any qualification	22,372	62,500	1.00	–
	No qualification	71,035	335,528	1.69	1.66-1.72
a	Total N = 500,434. However, N varies in some exposures, due to missing values				

All independent variables remained important predictors of chronic low back pain in the final multivariable model (all variables significant at p<0.001). The final model, comprising data from 482,232 participants, is shown in Table 82.

Table 82: factors associated with chronic low back pain – multivariable model

				OR [a]	95%CI
Age (years)					
	<45 yrs			1.00	–
	>=45 to <55 yrs			1.00	0.98-1.03
	>=55 to <65 yrs			0.93	0.90-0.95
	>=65 to <75 yrs			0.92	0.89-0.95
Gender					
	Male			1.00	–

	Female			1.16	1.14-1.18
Socioeconomic status – Employment status					
	In paid employment or self-employed			1.00	–
	Retired			1.31	1.29-1.34
	Looking after home and/or family			1.29	1.24-1.35
	Unable to work because of sickness or disability			4.76	4.60-4.92
	Unemployed			1.22	1.15-2.28
	Doing unpaid or voluntary work			1.06	0.95-1.19
	Full or part-time student			1.15	1.003-1.33
Socio-economic status – Townsend Index (deciles)					
	1 (most affluent)			1.00	–
	2			1.02	0.99-1.06
	3			1.03	0.995-1.07
	4			1.04	1.01-1.08
	5			1.06	1.03-1.11
	6			1.08	1.05-1.12
	7			1.08	1.04-1.12
	8			1.14	1.10-1.18
	9			1.23	1.19-1.27
	10 (most deprived)			1.33	1.29-1.38
Body mass index					
	<18.5 kgm^{-2}			1.00	–
	>=18.5 to <25.0 kgm^{-2}			0.84	0.75-0.94
	>=25.0 to <30.0 kgm^{-2}			1.29	1.26-1.31
	>=30.0 kgm^{-2}			1.72	1.69-1.76
Smoking status					
	Current smoker			1.00	–
	Ex-smoker			0.84	0.79-0.84
	Never regular			0.69	0.67-0.70
Education					
	Any qualification			1.00	–
	No qualification			1.32	1.29-1.34
a	Total N of final model = 482,232.				

REFERENCES

1. UK Biobank: Protocol for a large-scale prospective epidemiological resource. Protocol No: UKBB-PROT-09-06 (Main Phase). http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf (2007, accessed 23 September 2016).
2. Rose and Pevalin (with O'Reilly). The National Statistics Socio-economic Classification: origins, development and use. Palgrave MacMillan, Basingstoke, UK. ISBN 1-4039-9648-2.
3. Townsend, P., Phillimore, P. and Beattie, A. (1988) Health and Deprivation: Inequality and the North. Routledge, London. ISBN 0-7099-4351-2.