

User Guide: Health Survey for England Time-Series Dataset (1991-2009)

Authors: Vanessa Higgins and Alan Marshall,
Centre for Census & Survey Research, University
of Manchester.

Collaborators:

Iain Buchan, University of Manchester

Andrew Renehan, University of Manchester

Matt Sperrin, University of Lancaster

30 April 2012

Health Survey for England: time series dataset (1991-2009)

1. Introduction

This dataset is a time-series dataset of the Health Survey for England (HSE) from 1991-2009. The Stata dataset was compiled by Vanessa Higgins and Alan Marshall from the University of Manchester as part of a research project to examine obesity. The original data was obtained from the Economic and Social Data Service (ESDS). Crown copyright is held jointly with the Economic and Social Data Service. Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. More information about the Health Survey for England is available from the Economic and Social Data Service (ESDS) Government website - <http://www.esds.ac.uk/government/hse/> - including information on the sample design, questionnaire, variables etc. for each year of the survey.

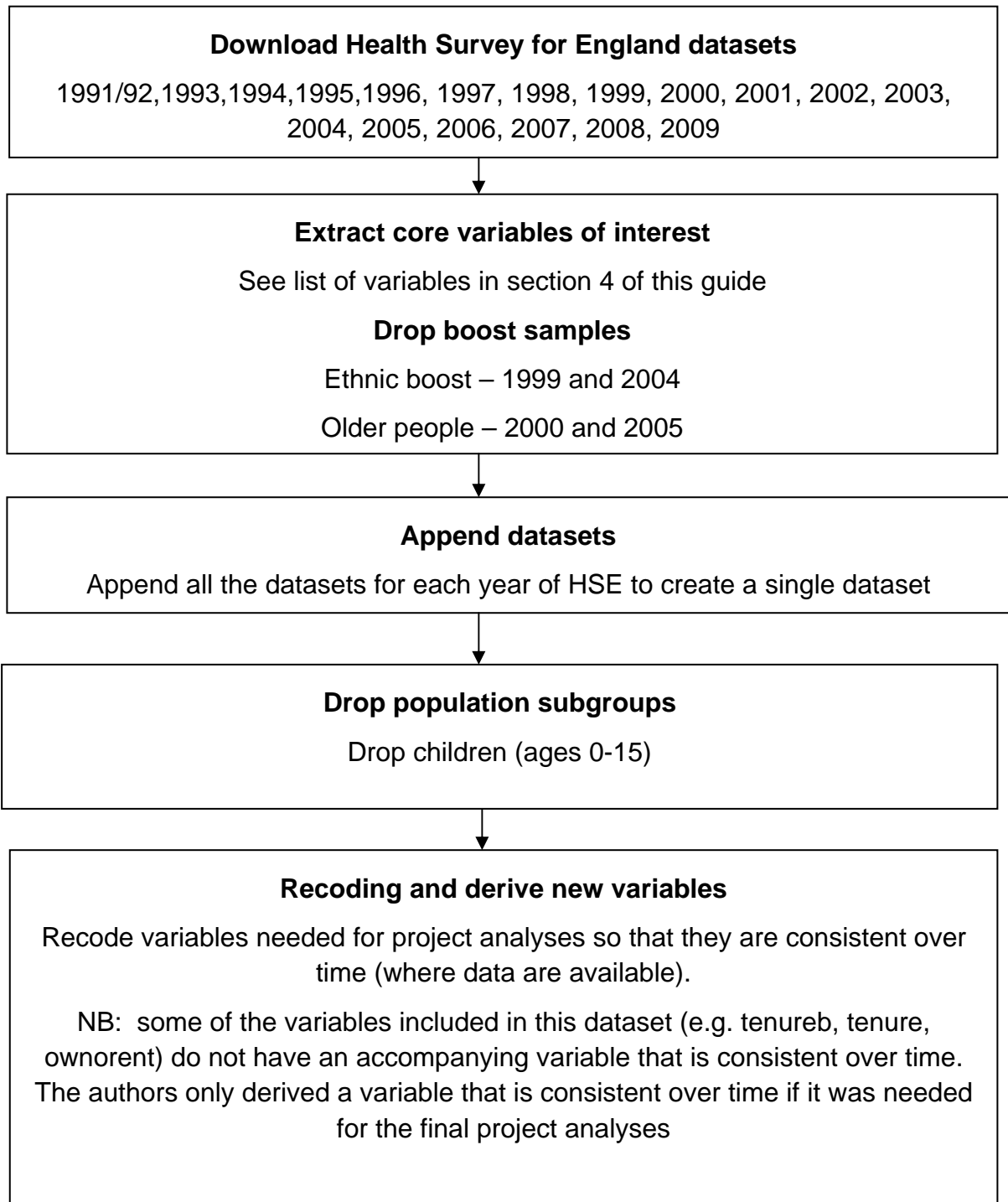
The dataset contains 192 variables– the variables included are those required for the project analyses on obesity. Some variables are available for the whole time period but some are only available for certain years. This is documented in more detail in Section 4 of this user guide. The Stata syntax file which was used to create the dataset ("*data extraction and merging do file.do*") accompanies this guide. The *do* file allows users to look in detail at how the dataset was compiled and how variables were derived. It is also possible to use the syntax to recreate the dataset and to add more variables if required (if users first download the 18 original HSE datasets from the ESDS).

The dataset contains 206,569¹ adult individuals (aged 16+) from the general population samples of the HSE. The HSE boost samples are excluded (ethnic boosts in 1999 and 2004 and elderly boosts in 2000 and 2005). Children aged 15 and younger are excluded from the dataset.

The name of the dataset is *1991-2009HSE.dta*

¹ However of the 206,569 cases there are 398 cases from the 1991/92 original data that have missing values for most of the original HSE variables apart from the variables *pserial* (renamed from *hserno* in 1991/92), *area* and *bmi* (although there are no height and weight variables to calculate BMI). Users may wish to exclude these 398 cases from their analyses.

2. Creating the HSE time-series dataset (1991-2009)



3. Survey weights

Survey sampling weights are available in the original HSE datasets from 2003 onwards. We have centred the survey weight variable (wt_int) so that the mean is equal to 1 (one) for each year from 2003 onwards. Prior to 2003 we have created a wt_int variable for each year which is set to equal 1 for all individuals, this ensures the weights have no effect prior to 2003. The wt_int variable can therefore be applied to all analyses from 1991-2009. There are no sampling weights for disproportionate sampling because boosted samples and children are discarded from this dataset.

4. Variables in the dataset

The variables available in the dataset are listed in the table below, along with the variable label and the years for which the variable is available. The fourth column identifies whether the variable is (a) an original variable from the HSE dataset or (b) a variable derived by the authors for the purpose of their analyses. Please note that some original HSE variables are included in the dataset but no consistent time-series variable has been derived by the authors because it was not needed for the project analyses. For example, the original HSE variables for 'housing tenure' are included (tenureb, tenure & ownorent) but a consistent time-series variable has not been derived. However it is possible for users to derive their own time-series variables from these original HSE variables.

For more information on the original HSE variables go to the original HSE User Guides/documentation available from ESDS: www.esds.ac.uk/government/hse/ .

For more information on the variables derived by the authors go to the accompanying syntax file (stata do file) which gives the detail of how these variables were derived.

Variable name	Variable label	Years	(a) Original HSE variable or (b) derived by authors
Classification variables			
sex	sex	1991-2009	a
age	age last birthday	1991-2009	a
agegp	grouped age	1991-2009	b
tenureb	household tenure	1998-2009	a
tenure	household tenure	1997	a
ownorent	household tenure	1996	a
sclass	social class	1996-2009	a
manual	manual/non manual worker	1996-2009	b
eqv5	equivalised income quintiles	1997-1998 2000-2003 2005-2009	a
eqvpop3	equivalised income tertiles (general population)	1999 & 2004	a
eqvinc	equivalised income	1997-2009	a
origin	origin of individual	2008-2009	a
ethinda	ethnic group	2005-2007	a
ethcind	can i check, to which of the groups on this card do you consider youn	2004	a
ethnici	ethnic group	1999-2003	a
nethnic	ethnic group	1998	a
ethnic	ethnic group	1996-1997	a
topqual2	highest educational qualification - students separate	1991-1994 1997-2009	a
topqual3	highest educational qualification	1993-1994 1997-2009	a
quala1	which qualifications - 1st	1995-1996	a
quala2	which qualifications - 2nd	1995-1996	a
quala3	which qualifications - 3rd	1995-1996	a

quala4	which qualifications - 4th	1995-1996	a
quala5	which qualifications - 5th	1995-1996	a
quala6	which qualifications - 6th	1995-1996	a
quala7	which qualifications - 7th	1995-1996	a
quala8	which qualifications - 8th	1995-1996	a
quala9	which qualifications - 9th	1995-1996	a
quala10	which qualifications - 10th	1995-1996	a
Obesity variables			
bmival	valid bmi measurements only	1997-2009	a
bmiok	whether bmi measure is valid	1991-1996	a
bmi	body mass index	1991-1996	a
bmivalid	BMI 1991-2009	1991-2009	b
bmigrp	grouped BMI 1991-2009	1991-2009	b
htval	valid height (cm)	1997-2009	a
height	respondents height (centimetres)	1991-1996	a
htok	whether height measure is valid	1991-1996	a
htvalid	Height 1991-2009	1991-2009	b
weight	respondent's weight (kg)	1991-1996	a
wtval	valid weight (kg) inc. estimated>130kg	1997-2009	a
wtok	whether weight measure is valid	1991-1996	a
wtvalid	Weight 1991-2009	1991-2009	b
hipval	valid mean hip (cm)	1997-2009	a
hipok	whether hip measurement valid	1993-1994	a
allhip	mean of all hip measurements	1993-1994	a
hipav	mean of two hip measurements	1991/2	a
hipvalid	Hip 1991-2009	1991-2009	b
wstval	valid mean waist (cm)	1997-2009	a
allwaist	allwaist mean of all waist measurements	1993-1994	a
wstok	whether waist measure is valid	1993-1994	a

waistav	mean of two waist measurements	1991-1994	a
waistvalid	Waist 1991-2009	1991-2009	b
whval	valid mean waist/hip ratio	1997-2009	a
alwhipra	mean waist-hip ratio	1993-1994	a
whok	whether valid waist-hip ratio possible	1993-1994	a
whiprat	waist-hip ratio (continuous)	1991/2	a
WHR	Waist-hip ratio 1991-2009	1991-2009	b
Health-related variables			
limitill	limiting longstanding illness	1996-2009	a
dinadias	dinamap diastolic bp (mean 2nd/3rd) inc. invalid (converted from omron)	2003-2009	a
didiaval	dinamap valid mean diastolic bp	2003-2009	a
diaval	valid mean diastolic bp	1997-2002	a
diabp2	mean diastolic bp - grouped (10)	1996	a
newdiast	diastolic bp (mean 2nd/3rd) including invalid	1996-2002	a
diastolic	Diastolic blood pressure 1997-2009	1997-2009	b
dinasyst	dinamap systolic bp (mean 2nd/3rd) inc. invalid (converted from omron)	2003-2009	a
disysval	dinamap valid mean systolic bp	2003-2009	a
sysval	valid mean systolic bp	1997-2002	a
newsyst	systolic bp (mean 2nd/3rd) including invalid	1996-2002	a
sysbp2	mean systolic bp - grouped (8)	1996	a
systolic	Systolic blood pressure 1997-2009	1997-2009	b
hdlval	valid hdl cholesterol result	1998-2000 2003-2006 2008-2009	a
cholest	total cholesterol result (blood data)	2008-2009 2003-2006 1998-2000	a

fldlval	valid ldl cholesterol (fasting) result	1999, 2003, 2004	a
Smoking variables			
cigsta3	cigarette smoking status: current/ex-reg/never-reg	1997-2009	a
cigst2	cigarette smoking status - banded current smokers	1997-2009	a
cot15val	valid cotinine (saliva): 0<15,15+	1997-2009	a
startsmk	age when started smoking	1995-2009	a
endsmoke	how long ago did you stop smoking cigarettes	1995-2009 1991-1993	a
cigst1	cigarette smoking status - never/ex-reg/ex-occ/current	1997-1998	a
newsmok2	cigarette smoking status (7)	1996	a
cigsmk2	cigarette smoking classification - broad groups	1991-1993 1995-1996	a
cigsmk1	cigsmk1 cigarette smoking status	1991-1993	a
cotexnic	grouped cotinine, exc nicotine product users - adults	1996	a
newsmok	extended cigarette smoking status classification	1994-1995	a
smokeVH	Smoking status 1994	1994	b
smokeAM	Smoking status 1997-1998	1997-1998	b
smoke	Smoking status 1991-2009	1991-2009	b
smkqt96	Smoking quantity 1996	1996	b
smkqt9495	Smoking quantity 1994-1995	1994-1995	b
smkqt9293	Smoking quantity 1991-1993	1991-1993	b
smkqt9709	Smoking quantity 1997-2009	1997-2009	b
smkquant	Smoking quantity 1991-2009	1991-2009	b
Drinking variables			
ddrinkof	freq of drinking over last 12 months (all types of alcoholic drinks)	1996	a

dnofit	frequency drank any alcoholic drink last 12 mths	1997-2009	a
Sample design variables			
wt_int	hse 2009 weight for analysis of core interview sample	1991-2009	a (derived for 1991/2-2002)
area	sample point number	1991-2009	a (renamed from 'psu' in 2006 and 2008)
pserial	serial number of individual	1991-2009	a (renamed from 'hserno' in 1993 and 1991/2)
cluster	stratification level	2003-2009	a
psu	sample point number	2008-2009 2006	a
year	Survey year	1991-2009	b
samptype	sample type	2000, 2002, 2005	a
tablewt	scaled weight variable	1997, 2002	a
samtyp	sample type	1997	a
yeargrp	Survey year - grouped	1991-2009	b
year1	-	1991-1993	b
Area/region variables			
gor	government office region	1998-2000 2002-2005 2008-2009	a
gor07	government office region for hse07 report	2007	a
gor06	government office region	2006	a
gora	government office regions	2001	a
gornew06	government office region 2006 – consistent value labels	2006	b
gornew07	government office region 2007 – consistent value labels	2007	b

gornew9800	government office region 1998-2000 – consistent value labels	1998-2000	b
gorVH	Government Office Regions 1998-2009	1998-2009	b
Physical activity variables			
a30t06g	summary moderate + activity level	2006, 2008	a
adt30gp	new summary activity level	1998-1999 2002-2004	a
active	physically active in your job	1991-1994 1997-1999 2002-2004 2008	a
hvyhwkkm	hvyhwkkm any heavy housework listed on this card? (show card k)	2008	a
hevyhwrk	any heavy housework	1991-1994 1997-1999 2002-2004 2006	a
actphy	sports/fitness activities completed in last 4 weeks (show card n)	1998-1999 2002-2004 2006-2008	a
actphys	done different activities in last 4 weeks?	1997	a
actany	any sport or exercise in last 4 weeks (card e)	1991-1994	a
actanyVH	Any sports/exercise in last 4 weeks (1991-2009)	1991-1994 1997-1999 2002-2004 2006-2008	b
oactq	have you done any other sport or exercise not listed on the card?	2008	a
oactq11	any other sport or exercise (1)	2002-2004, 2006	a
oactq12	any other sport or exercise (2)	2002	a

oactq13	any other sport or exercise (3)	2002	a
oactq14	any other sport or exercise (4)	2002	a
oactq15	any other sport or exercise (5)	2002	a
oactq16	any other sport or exercise (6)	2002	a
oactq1	any other sport or exercise? (1)	1997-1999	a
oactq2	any other sport or exercise? (2)	1997	a
oactq3	any other sport or exercise? (3)	1997	a
oactq4	any other sport or exercise? (4)	1997	a
acta	acta other sports activity level (a)	1991-1994	a
actb	actb other sports activity level (b)	1991	a
actc	actc other sports activity level (c)	1991	a
actd	actd other sports activity level (d)	1991	a
Social capital variables			
liveyrs	how many years have you lived in this local area	2001-2006	a
livemths	livemths how many months have you lived in this local area	2001-2006	a
enjoy	this area is a place i enjoy living in	2001-2006	a
neigrs	this area is a place where neighbours look after each other	2001-2006	a
transprt	this area has good local transport	2001-2006	a
leisure	this area has good leisure things for people like me	2001-2006	a
shops	ease of getting to supermarket	2001-2006	a
poffice	ease of getting to post office	2001-2006	a
teens	problem of teenagers hanging around on the streets	2001-2006	a
vandals	problem of vandalism, graffiti or deliberate damage to property	2001-2006	a
orgs01	political parties	2001-2006	a
orgs02	trade unions (including student unions)	2001-2006	a

orgs03	environmental groups	2001-2006	a
orgs04	parent-teacher association or school association	2001-2006	a
orgs05	tenants' or residents' group or neighbourhood watch	2001-2006	a
orgs06	education, arts, music or singing group (including evening classes)	2001-2006	a
orgs07	charity, voluntary or community group	2001-2006	a
orgs09	group for elderly or older people (eg lunch club)	2001-2006	a
orgs10	youth group (eg scouts, guides, youth club)	2001-2006	a
orgs11	women's institute or townswomen's guild or women's group	2001-2006	a
orgs12	social club (including working men's club, rotary club)	2001-2006	a
orgs13	sports club, gym, exercise or dance group	2001-2006	a
orgs14	other group or organisation	2001-2006	a
orgs15	no, don't regularly join in any of these	2001-2006	a
trusted	trusted can people be trusted	2001-2006	a
helpful	helpful do people try to be helpful	2001-2006	a
advntg	advntg people take advantage of you	2001-2006	a
clivysr	how long have you lived in this local area (years)?	2005	a
clivmon	how long have you lived in this local area (months)?	2005	a
cenjoy	cenjoy this area is a place i enjoy living in	2005	a
cneigbrs	cneigbrs this area is a place where neighbours look after each other	2005	a
ctrans	this area has good local transport	2005	a
cleisure	this area has good leisure things for people like me	2005	a
cshops	ease of getting to supermarket	2005	a

cpoffice	ease of getting to post office	2005	a
cteens	problem of teenagers hanging around on the streets	2005	a
cvandals	cvandals problem of vandalism, graffiti or deliberate damage to property	2005	a
cjoin	join in the activities of any of the organisations	2005	a
corgs01	political parties	2005	a
corgs02	trade unions (incl students' union)	2005	a
corgs03	environmental group	2005	a
corgs04	parent/teacher or school associations	2005	a
corgs05	tenants'/residents' group or neighbourhood watch	2005	a
corgs06	education, arts or music group (includes evening classes)	2005	a
corgs07	religious group or church organisation	2005	a
corgs08	charity, voluntary or community group	2005	a
corgs09	group for elderly/older people (eg lunch clubs)	2005	a
corgs10	youth group (e.g. scouts, guides, youth clubs etc)	2005	a
corgs11	women's institute or townswomen's guild or womens' group	2005	a
corgs12	social club (includes working men's club, rotary club)	2005	a
corgs13	other group or organisation	2005	a
ctrust	can people be trusted	2005	a
chelp	do people try to be helpful	2005	a
cadvntg	people take advantage of you	2005	a

*Created by Vanessa Higgins and Alan Marshall, University of Manchester, April 2012

***This do file creates a single 1991-2009 HSE dataset containing a reduced version of each year of the Health Survey for England 1991-2009 (adult data only)

***There are 3 steps:

***Step 1 saving reduced and recoded datasets

*Opening a file

*Selecting appropriate variables

*Conducting the necessary recoding

*Saving a reduced version of the file

***Step 2 - appending each year of the HSE to create a single 1991-2009 file

*Open the reduced version in 2009

*Append each years reduced HSE file

***Step 3 - deriving variables that are consistent over time for the years available

*Notes:

*No cluster variable before 2002. Cluster set to equal . for all years prior to 2002

*No sampling weight variable (wt_int) before 2003. Wt_int set to equal 1 for all years before 2003

*NB: if you wish to run this file you will need to download each year of the HSE data and save into a folder called to C:\Obesity Project\Data

*****Step 1 - saving reduced and recoded datasets*****

***Open 2009 data

clear

set mem 200m

use "C:\Obesity project\Data\2009 UKDA-6732-stata8\stata8\hse09ai.dta", clear

*Select required variables

keep pserial cluster psu wt_int topqual3 topqual2 limitill origin tenureb eqvinc sex age sclass wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke cot15val cigst2 dinadias didiaval dinasyst disysval hdlval cholest eqv5 gor

*Generate a year identifier variables

gen year=2009

*rename psu to area to match other years

rename psu area

*changing gor from string to numeric variable

encode gor , generate (gorNEW)

drop gor

rename gorNEW gor

*Save reduced and recoded file

save "C:\Obesity project\Data\Reduced and recoded HSE files\2009.dta", replace

***Open 2008 data

clear

use "C:\Obesity project\Data\2008 UKDA-6397-stata8_se\stata8_se\hse08ai.dta", clear

*Select required variables

keep pserial cluster psu wt_int topqual3 topqual2 limitill origin tenureb eqvinc sex age sclass wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke cot15val cigst2 dinadias didiaval dinasyst disysval hdlval cholest a30t06g eqv5 gor actphy oactq active hvyhwkxm a30t06g

*Generate a year identifier variables

```

gen year=2008

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2008.dta", replace

***Open 2007 data
clear
use "C:\Obesity project\Data\2007 UKDA-6112-stata8_se\stata8_se\hse07ai.dta", clear

*Select required variables
keep pserial cluster area wt_int topqual3 topqual2 limitill ethinda tenureb eqvinc
sex age sclass wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke
cot15val cigst2 dinadias didiaval dinasyst disysval eqv5 gor07

*Generate a year identifier variables
gen year=2007

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2007.dta", replace

***Open 2006 data
clear
use "C:\Obesity project\Data\2006 UKDA-5809-stata8\stata8\hse06ai.dta", clear

*Select required variables
keep pserial cluster psu wt_int topqual3 topqual2 limitill ethinda tenureb eqvinc sex
age sclass wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke
cot15val cigst2 dinadias didiaval dinasyst disysval hdlval cholest a30t06g eqv5 gor06
liveyrs livemths enjoy neigbrs transprt leisure shops poffice teens vandals orgs01
orgs02 orgs03 orgs04 orgs05 orgs06 orgs07 orgs08 orgs09 orgs10 orgs11 orgs12 orgs13
orgs14 orgs15 trusted helpful advntg actphy oactq11 oactq12 oactq13 oactq14 oactq15
oactq16 hevyrwrk a30t06g

*Generate a year identifier variables
gen year=2006

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2006.dta", replace

***Open 2005 data
clear
use "C:\Obesity project\Data\2005 UKDA-5675-stata8\stata8\hse05ai.dta", clear

*Select required variables
keep pserial cluster area wt_int topqual3 topqual2 limitill ethinda tenureb eqvinc
sex age sclass wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke
cot15val cigst2 dinadias didiaval dinasyst disysval hdlval cholest samptype eqv5 gor
liveyrs livemths enjoy neigbrs transprt leisure shops poffice teens vandals orgs01
orgs02 orgs03 orgs04 orgs05 orgs06 orgs07 orgs08 orgs09 orgs10 orgs11 orgs12 orgs13
orgs14 orgs15 trusted helpful advntg clivyrns clivmon cenjoy cneigbrs ctrans cleisure
cshops cpoffice cteens cvandals cjoin corgs01 corgs02 corgs03 corgs04 corgs05 corgs06
corgs07 corgs08 corgs09 corgs10 corgs11 corgs12 corgs13 ctrust chelp cadvntg

*Generate a year identifier variables
gen year=2005

*Drop boosted older person sample in 2005
drop if samptype==2

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2005.dta", replace

***Open 2004 data
clear
use "C:\Obesity project\Data\2004 UKDA-5439-stata8\stata8\hse04gpa.dta", clear

```



```

*Select required variables
keep pserial cluster area wt_int topqual3 topqual2 limitill ethcind tenureb eqvinc
sex age sclass wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke
cot15val cigst2 dinadias didiaval dinasyt disysval hdlval fldlval cholest adt30gp
eqvpop3 gor liveyrs livemths enjoy neigbrs transprt leisure shops poffice teens
vandals orgs01 orgs02 orgs03 orgs04 orgs05 orgs06 orgs07 orgs08 orgs09 orgs10 orgs11
orgs12 orgs13 orgs14 orgs15 trusted helpful advntg actphy oactq11 oactq12 oactq13
oactq14 oactq15 oactq16 active hevhwrk adt30gp

*dividing wt_int by its mean so it has a mean of 1 (like all other years)
summ wt_int
gen wt_inttemp=wt_int/14.15606
drop wt_int
rename wt_inttemp wt_int
summ wt_int

*Generate a year identifier variables
gen year=2004

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2004.dta", replace

***2003
clear
use "C:\Obesity project\Data\2003 UKDA-5098-stata8\stata8\hse03ai.dta", clear

*Select required variables
keep pserial cluster area int_wt topqual3 topqual2 limitill ethnici tenureb eqvinc
sex age sclass wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke
cot15val cigst2 dinadias didiaval dinasyt disysval hdlval fldlval cholest adt30gp
eqv5 gor liveyrs livemths enjoy neigbrs transprt leisure shops poffice teens vandals
orgs01 orgs02 orgs03 orgs04 orgs05 orgs06 orgs07 orgs08 orgs09 orgs10 orgs11 orgs12
orgs13 orgs14 orgs15 trusted helpful advntg actphy oactq11 oactq12 oactq13 oactq14
oactq15 oactq16 active hevhwrk adt30gp

*rename int_wt to be consistent with previous years
rename int_wt wt_int

*Generate a year identifier variables
gen year=2003

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2003.dta", replace

***2002
clear
use "C:\Obesity project\Data\2002 UKDA-4912-stata8_se\stata8_se\hse02ai.dta", clear

*Select required variables
keep pserial area topqual3 topqual2 limitill ethnici tenureb eqvinc sex age sclass
wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke cot15val
cigst2 newdiast diaval newsyst sysval adt30gp tablewt samptype eqv5 gor liveyrs
livemths enjoy neigbrs transprt leisure shops poffice teens vandals orgs01 orgs02
orgs03 orgs04 orgs05 orgs06 orgs07 orgs08 orgs09 orgs10 orgs11 orgs12 orgs13 orgs14
orgs15 trusted helpful advntg actphy oactq11 oactq12 oactq13 oactq14 oactq15 oactq16
active hevhwrk adt30gp

*Generate a year identifier variables
gen year=2002

*Drop boosted children/youngpeople/mother sample in 2002
drop if samptype==1

*I have generated a weight=1 for all cases in 2002 as we're dropping the boost so not
relevant
gen wt_int=1

```

```

*gen wt_int=tablewt

*changing gor from string to numeric variable
encode gor , generate (gorNEW)
drop gor
rename gorNEW gor

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2002.dta", replace

***2001
clear
use "C:\Obesity project\Data\2001 UKDA-4628-stata8\stata8\hse01ai.dta", clear

*Select required variables
keep pserial area topqual3 topqual2 limitill ethnici tenureb eqvinc sex age sclass
wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke cot15val
cigst2 newdiast diaval newsyst sysval eqv5 gora liveyrs livemths enjoy neighbors
transprt leisure shops poffice teens vandals orgs01 orgs02 orgs03 orgs04 orgs05
orgs06 orgs07 orgs08 orgs09 orgs10 orgs11 orgs12 orgs13 orgs14 orgs15 trusted helpful
advntg

*Generate a year identifier variables
gen year=2001

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*changing gora from string to numeric variable
encode gora , generate (gorNEW)
drop gora
rename gorNEW gora

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2001.dta", replace

***2000
clear
use "C:\Obesity project\Data\2000 UKDA-4487-stata8\stata8\hse00ai.dta", clear

*Select required variables
keep pserial area topqual3 topqual2 limitill ethnici tenureb eqvinc sex age sclass
wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke cot15val
cigst2 newdiast diaval newsyst sysval hdlval cholest samptype eqv5 gor

*Generate a year identifier variables
gen year=2000

*Drop boosted older person sample in 2000
drop if samptype==1

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\2000.dta", replace

```

```

***1999
clear
use "C:\Obesity project\Data\1999 UKDA-4365-stata8\stata8\hse99gp3.dta", clear

*Select required variables
keep pserial area topqual3 topqual2 limitill ethnici tenureb eqvinc sex age sclass
wtval htval bmival wstval hipval whval dnoft cigsta3 startsmk endsmoke cot15val
cigst2 newdiast diaval newsyst sysval hdlval fldlval cholest adt30gp eqvpop3 gor
actphy oactq1 oactq2 oactq3 oactq4 active hevhwrk adt30gp

*Generate a year identifier variables
gen year=1999

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\1999.dta", replace

***1998
clear
use "C:\Obesity project\Data\1998 UKDA-4150-stata8\stata8\hse98ai.dta", clear

*Select required variables
keep pserial area topqual3 topqual2 limitill nethnic tenureb eqvinc sex age sclass
wtval htval bmival wstval hipval whval dnoft startsmk endsmoke cot15val cigst1
newdiast diaval newsyst sysval hdlval cholest adt30gp eqv5 gor actphy oactq1 oactq2
oactq3 oactq4 active hevhwrk adt30gp

*Generate a year identifier variables
gen year=1998

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\1998.dta", replace

***1997
clear
use "C:\Obesity project\Data\1997 UKDA-3979-stata8\stata8\hse97ai.dta", clear

*Select required variables
keep pserial area topqual3 topqual2 limitill ethnic tenure eqvinc sex age sclass
wtval htval bmival wstval hipval whval dnoft startsmk endsmoke cot15val cigst1 cigst2
newdiast diaval newsyst sysval tablewt samtyp eqv5 actphys oactq1 oactq2 oactq3
oactq4 active hevhwrk

*Generate a year identifier variables
gen year=1997

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=tablewt

*No cluster variable before 2002. Create a cluster variable for appending purposes

```

```

gen cluster=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\1997.dta", replace

***1996
clear
use "C:\Obesity project\Data\1996 UKDA-3886-stata8\stata8\hse96bind.dta", clear

*Select required variables
keep pserial area quala1 quala2 quala3 quala4 quala5 quala6 quala7 quala8 quala9
quala10 limitill ethnic ownorent sex age sclass weight wtok height htok bmi bmiok
ddrinkof cigsmk2 startsmk endsmoke cotexnic newsmok2 newdiast diabp2 newsyst sysbp2
*included cigsmk2 in case we decide we need to use it

*Generate a year identifier variables
gen year=1996

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*No equalise household income before 1996
gen eqvinc=.
gen eqv5=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\1996.dta", replace

***1995
clear
use "C:\Obesity project\Data\1995 UKDA-3796-stata8\stata8\hse95ai2nd.dta", clear

*Select required variables
keep pserial area sex age quala1 quala2 quala3 quala4 quala5 quala6 quala7 quala8
quala9 quala10 bmi bmiok weight wtok height htok cigsmk2 startsmk endsmoke newsmok
*to check/amend variable names and add later: quala1 limitill ethnic ownorent sclass
cotexnic ddrinkof newdiast diabp2 newsyst sysbp2
*waist and hip - no waist or hip variables in 1995
*included cigsmk2 in case we decide we need to use it

*Generate a year identifier variables
gen year=1995

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*No equalise household income before 1996
gen eqvinc=.
gen eqv5=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\1995.dta", replace

***1994
clear
use "C:\Obesity project\Data\1994 UKDA-3640-stata8\stata8\hse94ai2nd.dta", clear

*Select required variables

```

```

keep pserial area sex age bmi bmiok weight wtok height htok newsmok allwaist allhip
alwhipra hipok wstok whok topqual2 topqual3 actany acta actb actc actd active
hevyhwrk
*startsmk and endsmk and cigsmk2 not in 1994
*to check/amend variable names and add later: qualal limitill ethnic ownorent sclass
startsmk endsmoke cotexnic ddrinkof newdiast diabp2 newsyst sysbp2

*Generate a year identifier variables
gen year=1994

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*No equalise household income before 1996
gen eqvinc=.
gen eqv5=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\1994.dta", replace

***1993
clear
use "C:\Obesity project\Data\1993 UKDA-3316-stata8\stata8\3316.dta", clear

*Select required variables
keep hserno area sex age bmi bmiok weight wtok height htok cigsmk2 endsmoke cigsmk1
year schedtyp allwaist allhip alwhipra hipok wstok whok topqual2 topqual3 actany acta
actb actc actd active hevyhwrk
*to check/amend variable names and add later: qualal limitill ethnic ownorent sclass
ddrinkof cotexnic newdiast diabp2 newsyst sysbp2
*startsmk not in 1994 so included cigsmk2

*Generate a year identifier variables
gen year1=year
drop year
gen year=1993

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*Drop non responders - these are included in 1993 (but not any other years)
drop if schedtyp==0
drop schedtyp

*rename hserno pserial - consistent with other years
rename hserno pserial

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*No equalise household income before 1996
gen eqvinc=.
gen eqv5=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\1993.dta", replace

***1991/2
clear
use "C:\Obesity project\Data\1991 to1992 UKDA-3238-stata8\stata8\3238", clear

```

```

*Select required variables
keep hserno area sex age weight wtok height htok bmi bmiok cigsmk2 endsmoke cigsmk1
year waistav hipav whiprat whok topqual2 actany acta actb actc actd active hevyrwrk
*to check/amend variable names and add later: qualal limitill ethnic ownorent sclasdd
drinkof cotexnic newdiast diabp2 newsyst sysbp2
*startsmk not in 1991-2 so included cigsmk2

*Generate a year identifier variables
gen year1=year
drop year
gen year=1992

*No wt_int (sampling weights) variable before 2003. Generate a weight variable and
set to equal 1. We need same variables in each file to append them.

gen wt_int=1

*rename hserno pserial - consistent with other years
rename hserno pserial

*No cluster variable before 2002. Create a cluster variable for appending purposes
gen cluster=.

*No equalise household income before 1996
gen eqvinc=.
gen eqv5=.

*Save reduced and recoded file
save "C:\Obesity project\Data\Reduced and recoded HSE files\1991-92.dta", replace

*****Step 2 - appending each year of the HSE to create
one dataset for 1991-2009*****
use "C:\Obesity project\Data\Reduced and recoded HSE files\2009.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2008.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2007.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2006.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2005.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2004.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2003.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2002.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2001.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\2000.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\1999.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\1998.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\1997.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\1996.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\1995.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\1994.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\1993.dta"
append using "C:\Obesity project\Data\Reduced and recoded HSE files\1991-92.dta"

*drop all children (<16 years) from 1991-2009 keep only required ages
drop if age<16

save "C:\Obesity project\Data\Reduced and recoded HSE files\1991_2009HSE.dta",
replace

*****Step 3 - creating derived variables*****

**Creating harmonised variables from 1991-2009 for BMI, weight, height, smoking,
waist and hip**
**BMI
gen bmivalid=bmival if year>=1997
replace bmivalid=bmi if bmiok==1 & year<=1996
codebook bmivalid
mvdecode bmivalid, mv (-1=.)
codebook bmivalid

```

```

**Weight
gen wtvalid=wtval if year>=1997
replace wtvalid=weight if wtok==1 & year<=1996
codebook wtvalid
mvdecode wtvalid, mv (-1=.)
codebook wtvalid

**Height
gen htvalid=htval if year>=1997
replace htvalid=height if htok==1 & year<=1996
codebook htvalid
mvdecode htvalid, mv (-8=.\-1=.)
codebook htvalid

**smoking status
gen smokeVH=newsmok if year==1994
recode smokeVH 7/11=1 2/6=2 1=3 *=.
gen smoke=cigsta3 if year>=1999
replace smoke=cigsmk2 if year==1996 | year==1995 | year==1993 | year==1992
gen smokeAM=cigst1 if year==1997|year==1998
recode smokeAM 4=1 2/3=2 1=3
replace smoke=smokeAM if year==1997|year==1998
replace smoke=smokeVH if year==1994
mvdecode smoke, mv (-9=.\-8=.\-7=.\-6=.\-2=.\-1=.)
label define smoke 1 "current" 2 "ex-regular" 3 "never regular"
label values smoke smoke

*smoking quantity
***recoding into temp variables to match value labels across years
gen smkqt96=newsmok2 if year==1996
recode smkqt96 4=1 5=2 6=3 *=.
gen smkqt9495=newsmok if year==1994|year==1995
recode smkqt9495 8=1 9=2 10=3 *=.
gen smkqt9293=cigsmk1 if year==1992 | year==1993
recode smkqt9293 1=30 2=20 3=10 *=.
recode smkqt9293 30=3 20=2 10=1 *=.
gen smkqt9709=cigst2 if year>=1997
recode smkqt9709 1=1 2=2 3=3 *=.

*generating a consistent smoking quantity variable
gen smkquant=smkqt9709 if year>=1997
replace smkquant=smkqt96 if year==1996
replace smkquant=smkqt9495 if year==1994 | year==1995
replace smkquant=smkqt92 if year==1992 | year==1993
mvdecode smkquant, mv (-9=.\-8=.\-1=.)
ta smkquant
label define smkquant 1 "light <10" 2 "moderate 10-19" 3 "heavy 20+"
label values smkquant smkquant

*change 1994 height to cms
replace htvalid=htvalid*100 if year==1994

*waist
gen waistvalid=wstval if year>=1997
replace waistvalid=. if year==1995 | year==1996
replace waistvalid=allwaist if wstok==1 & (year==1993 | year==1994)
replace waistvalid=waistav if year==1992
codebook waistvalid
mvdecode waistvalid, mv (-1=.\-9=.\-8=.\-7=.)
codebook waistvalid
*1992 doesn't have a wstok variable (it has a whok variable but this is whether whr
is ok not just waist ok)

*hip
gen hipvalid=hipval if year>=1997
replace hipvalid=. if year==1995 | year==1996
replace hipvalid=allhip if hipok==1 & (year==1993 | year==1994)

```

```

replace hipvalid=hipav if year==1992
codebook hipvalid
mvdecode hipvalid, mv (-1=. \-9=. \-8=. \-7=.)
codebook hipvalid
*1992 doesn't have a hipok variable (it has a whok variable but this is whether whr
is ok not just hip ok)

*wasit-hip ratio
gen WHR=whval if year>=1997
replace WHR=. if year==1995 | year==1996
replace WHR=alwhipra if whok==1 & (year==1993 | year==1994)
replace WHR=whiprat if whok==1 & year==1992
codebook WHR
mvdecode WHR, mv (-1=. \-9=. \-8=. \-7=.)
codebook WHR

*****recoding age into agegroups*****
gen agegp=.
replace agegp=1 if age>=16 & age <=34
replace agegp=2 if age>=35 & age <=54
replace agegp=3 if age>=55 & age<102
label define agegp 1 "16-34" 2 "35-54" 3 "55+"
label values agegp agegp
ta agegp

*****setting missing values*****
**setting social class missing values to missing
mvdecode sclass, mv (-1=.)
**recoding armed forces and 'not fully described as missing' (not clear where to put
armed forces and only a handful of cases)
mvdecode sclass, mv (7=. \8=.)
**setting qualifiaction class missing values to missing
mvdecode topqual3, mv (-1=. \-9=. \-8=.)
mvdecode topqual2, mv (-1=. \-9=. \-8=. \-7=.)
**setting tenureb missing values to missing
mvdecode tenureb, mv (-9=. \-8=.)
**setting drinking missing values to missing
mvdecode dnoft, mv (-9=. \-8=. \-6=. \-1=.)
**setting limiting longstanding illness missing values to missing
mvdecode limitill, mv (-9=. \-8=.)
**setting equivalised income dv missing values to missing
mvdecode eqv5, mv (-1=. \-90=.)

*****deriving diastolic blood pressure over time where data
available*****
gen diastolic=diaval if year>=1997 & year<=2002
replace diastolic=didiaval if year>=2003 & year<=2009
codebook diastolic
mvdecode diastolic, mv (-2=. \-1=. \-8=. \-7=.)
codebook diastolic
mean diastolic, over (year)

*****deriving systolic blood pressure over time where data
available*****
gen systolic=sysval if year>=1997 & year<=2002
replace systolic=disysval if year>=2003 & year<=2009
codebook systolic
mvdecode systolic, mv (-2=. \-1=. \-8=. \-7=.)
codebook systolic
mean systolic, over (year)

*****value labels for gor over time (1998-2000 emids and wmids; 2006-07
SW and SE*****
gen gornew06=gor06
recode gornew06 8=10 9=8
recode gornew06 10=9
ta gornew06

```



```

gen gornew07=gor07
recode gornew07 8=10 9=8
recode gornew07 10=9
ta gornew07

gen gornew9800=gor if year>=1998 & year<=2000
recode gornew9800 4=10 5=4
recode gornew9800 10=5
ta gornew9800

*creating gorVH over time
gen gorVH=gor if (year>=2002 & year<=2005) | (year>=2008 & year<=2009)
replace gorVH=gornew9800 if year>=1998 & year<=2000
replace gorVH=gora if year==2001
replace gorVH=gornew06 if year==2006
replace gorVH=gornew07 if year==2007
mvdecode gorVH, mv (10=.)
label define gorVH 1 "NE" 2 "NW" 3 "Yorks and Humberside" 4 "East Midlands" 5 "West
Midlands" 6 "East England" 7 "London" 8 "SE" 9 "SW"
label values gorVH gorVH

*****creating yeargrp*****
gen yeargrp=1 if year>=1992 & year<=1997
replace yeargrp=2 if year>=1998 & year<=2001
replace yeargrp=3 if year>=2002 & year<=2005
replace yeargrp=4 if year>=2006 & year<=2009
label define yeargrp 1 "92-97" 2 "98-01" 3 "02-05" 4 "06-09"
label values yeargrp yeargrp
mvdecode yeargrp, mv (-99=.)

*****creating obesity variable*****
gen bmigrp=bmivalid
replace bmigrp=1 if bmivalid >0 & bmivalid <18.5
replace bmigrp=2 if bmivalid >=18.5 & bmivalid <25
replace bmigrp=3 if bmivalid >=25 & bmivalid <30
replace bmigrp=4 if bmivalid >=30 & bmivalid <=75
label define bmigrp 1 "underweight" 2 "healthy weight" 3 "overweight" 4 "obese"
label values bmigrp bmigrp
1

*****creating manual/non-manual variable*****
gen manual=sclass
replace manual=1 if sclass>=4 & sclass<=6
replace manual=2 if sclass>=1 & sclass<=3
label define manual 1 "manual" 2 "non-manual"
label values manual manual

*****deriving 'any sports/exercise in past 4 weeks' - consistent
variable over time where data are available*****
mvdecode actany, mv (-8=.\-7=.\9=.)
mvdecode actphys, mv (-9=.\-8=.)
mvdecode actphy, mv (-9=.\-8=.\-1=.)
mvdecode acta, mv (-9=.\-8=.\-7=.\-1=.\9=.\4=.)
mvdecode actb, mv (-9=.\-8=.\-7=.\-1=.)
mvdecode actc, mv (-9=.\-8=.\-7=.\-1=.)
mvdecode actd, mv (-9=.\-8=.\-7=.\-1=.)
mvdecode oactq1, mv (-9=.\-8=.)
mvdecode oactq2, mv (-1=.)
mvdecode oactq3, mv (-1=.)
mvdecode oactq4, mv (-1=.)
mvdecode oactq, mv (-9=.\-8=.)
mvdecode oactq11, mv (-9=.\-8=.\-1=.)
mvdecode oactq12, mv (-1=.)
mvdecode oactq13, mv (-1=.)
mvdecode oactq14, mv (-1=.)
mvdecode oactq15, mv (-1=.)

```

```

mvdecode oactq16, mv (-1=.)
gen actanyVH=actany if year==1992 | year==1993 | year==1994
replace actanyVH=actphys if year==1997
replace actanyVH=actphy if year==1998 |year==1999 |year==2002 |year==2003 |year==2004
|year==2006 |year==2006
replace actanyVH=1 if acta==1 | actb==1 | actc==1 | actd==1 | oactq1==1 | oactq2==1 |
oactq3==1 |oactq4==1 |oactq11==1 |oactq12==1 |oactq13==1 |oactq14==1 |oactq15==1 |
|oactq16==1 |oactq==1
label define actanyVH 1"yes" 2 "no"
label values actanyVH actanyVH

label variable smoke "Smoking status 1991-2009"
label variable smokeVH "Smoking status 1994"
label variable smokeAM "Smoking status 1997-1998"
label variable year "Survey year"
label variable yeargrp "Survey year - grouped"
label variable smkqt96 "Smoking quantity 1996"
label variable smkqt9495 "Smoking quantity 1994-1995"
label variable smkqt9293 "Smoking quantity 1991-1993"
label variable smkqt9709 "Smoking quantity 1997-2009"
label variable smkquant "Smoking quantity 1991-2009"
label variable manual "manual/non manual worker"
label variable bmigrp "grouped BMI 1991-2009"
label variable agegp "grouped age"
label variable bmivalid "BMI 1991-2009"
label variable actanyVH "Any sports/exercise in last 4 weeks (1991-2009)"
label variable gornew06 "government office region 2006 - consistent value labels"
label variable gornew07 "government office region 2007 - consistent value labels"
label variable gornew9800 "government office region 1998-2000 - consistent value
labels"
label variable gorVH "Government Office Regions 1998-2009"
label variable htvalid "Height 1991-2009"
label variable wtvalid "Weight 1991-2009"
label variable hipvalid "Hip 1991-2009"
label variable waistvalid "Waist 1991-2009"
label variable WHR "Waist-hip ratio 1991-2009"
label variable diastolic "Diastolic blood pressure 1997-2009"
label variable systolic "Systolic blood pressure 1997-2009"

save "C:\Obesity project\Data\Reduced and recoded HSE files\1991_2009HSE.dta",
replace

```