Codebook

Harmonised ExPoSE England Dataset

V. 1.1

Dataset name:	DATA_EN
Dataset size:	101.6 Mb
Column count:	111
Row count:	168,296
Updated date:	2024-10-29
Changes from version 1.1	Total cholesterol values in 2017 corrected to eliminate an unwarranted adjustment previously applied. Sampling weights for laboratory in 2005 adjusted to take into account thant only subjects 65 years and older had their blood analysed. Description of income quintiles corrected to indicate that they are based on the equivalized income (no change in data). Added new variables: *heartmed*, *heartmed2*, *exercisefreq_e*, *glyhb_h*, *hbA1c*.

1	**Column name:**	**country_ISO**
	Column description:	Country ISO code
	Data type:	Factor
	Unique non-missing value count:	1
	Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
GBR	168,296	168,296	100.00

2	**Column name:**	**country_name**
	Column description:	Country name
	Data type:	Factor
	Unique non-missing value count:	1

Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
England	168,296	168,296	100.00

3	Column name:	source
	Column description:	Data source
	Data type:	Factor
	Unique non-missing value count:	17
	Missing value count:	0

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
HSE 2009	4645	HSE 2006	14142
HSE 2007	6882	HSE 2003	14836
HSE 2002	7393	HSE 2008	15098
HSE 2005	7630	HSE 2001	15634
HSE 2017	7997	HSE 1998	15802

4	Column name:	year
	Column description:	Year of data collection - Survey median
	Data type:	Numeric
	Unique non-missing value count:	17
	Missing value count:	0

Min	Mean	Median	Max	SD
1998.00	2007.53	2008.00	2017.00	5.66

5	Column name:	pid
	Data type:	Numeric
	Source information:	Generated in the harmonised dataset to be unique for each individual in the dataset.
	Unique non-missing value count:	168,296
	Missing value count:	0

Min	Mean	Median	Max	SD
19981.00	45663875.48	20094344.50	200815098.00	65732282.95

6	Column name:	psu
	Column description:	Primary Sampling Unit (PSU)
	Source information:	Generated in the harmonised dataset to be unique. Codes are not linked across sources, i.e. it is possible that different codes are used to indicate the same PSU in two different source datasets.
	Data type:	Numeric
	Unique non-missing value count:	11,162
	Missing value count:	0

Min	Mean	Median	Max	SD
19981.00	5308736.71	2007556.00	20101839.00	7105865.99

7	Column name:	stratum
	Column description:	Sampling stratum
	Source information:	It is recommended to use the Government Office Region (GOR) as the stratum for trend analyses over time to ensure that there are at least 2 PSUs per stratum for calculation of variance. Individuals with a missing GOR were dropped from the analysis (106 individuals in 1998)
	Data type:	Factor
	Unique non-missing value count:	9
	Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
1	11,878	11,878	7.06
2	23,586	35,464	14.01
3	17,038	52,502	10.12
4	16,009	68,511	9.51
5	17,174	85,685	10.20
6	19,068	104,753	11.33

7	19,770	124,523	11.75
8	25,319	149,842	15.04
9	18,454	168,296	10.97

8	Column name:	aweight_en_int
	Column description:	Sampling weight: Interview
	Source information:	Interview weights from the original surveys (wt_int). To be used for analyses involving only variables in the individual questionnaire or height, weight and BMI.
	Data type:	Numeric
	Unique non-missing value count:	77,021
	Missing value count:	0

Min	Mean	Median	Max	SD
0.28	1.00	1.00	9.89	0.31

9	Column name:	aweight_en_int_cvd
	Column description:	Sampling weight: Interview, CVD analysis
	Source information:	Based on the interview weights from the original surveys. However, for 2006, weights are the sample 2 interview weights for those aged 65+ who were asked the CVD questions and exclude those not given the CVD module (sample 1). To be used for analyses of the individual questionnaire that involve CVD module variables such as prior CVD and diabetes diagnoses, etc.
	Data type:	Numeric
	Unique non-missing value count:	60,112
	Missing value count:	64,740

Min	Mean	Median	Max	SD
0.20	1.00	1.00	10.37	0.41

10	Column name:	aweight_en_nonlab
	Column description:	Sampling weight: non-lab risk score
	Source information:	Nurse interview weight (wt_nurse) from the HSE survey. To be used for analyses involving variables

		from the nurse interview such as blood pressure, waist and hip measurements and prescribed medications, including analyses involving the non-laboratory CVD risk scores.
	Data type:	Numeric
	Unique non-missing value count:	82,493
	Missing value count:	39,041

Min	Mean	Median	Max	SD
0.23	1.00	1.00	14.29	0.38

11	**Column name:**	**aweight_en_nonlab_cvd**
	Column description:	Sampling weight: non-lab risk score, CVD analysis
	Source information:	To be used in analyses involving non-lab CVD risk score or nurse interview variables that also include questions from the CVD module in 2006 (see note on aweight_en_nonlab)
	Data type:	Numeric
	Unique non-missing value count:	81,625
	Missing value count:	40,251

Min	Mean	Median	Max	SD
0.23	1.00	1.00	14.29	0.38

12	**Column name:**	**aweight_en_lab**
	Column description:	Sampling weight: laboratory risk score
	Source information:	Blood sample weight (wt_blood) from the HSE survey. To be used for analyses involving variables from the blood sample such as HbA1c and total cholesterol.
	Data type:	Numeric
	Unique non-missing value count:	60,111
	Missing value count:	64,740

Min	Mean	Median	Max	SD
0.20	1.00	1.00	10.37	0.41

13	**Column name:**	**aweight_en_lab_cvd**

	Column description:	Sampling weight: laboratory risk score, CVD analysis
	Source information:	To be used in analyses involving the laboratory CVD risk scores or in analyses of blood sample variables that also include questions from the CVD module in 2006 (see note on aweight_en_lab).
	Data type:	Numeric
	Unique non-missing value count:	59,434
	Missing value count:	65,635

Min	Mean	Median	Max	SD
0.20	1.00	1.00	10.37	0.41

14	**Column name:**	**geolevel1_name**
	Column description:	Administrative level 1 - Name
	Source information:	Government office region. Merseyside was merged with North West in 1998.
	Data type:	Factor
	Unique non-missing value count:	9
	Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
East Midlands	17,174	17,174	10.20
Eastern	19,068	36,242	11.33
London	19,770	56,012	11.75
North East	11,878	67,890	7.06
North West & Merseyside	23,586	91,476	14.01
South East	25,319	116,795	15.04
South West	18,454	135,249	10.97
West Midlands	16,009	151,258	9.51
Yorkshire & The Humberside	17,038	168,296	10.12

15	**Column name:**	**geolevel1_code**

	Column description:	Administrative level 1 - Code
	Source information:	Government office region. Merseyside was merged with North West in 1998.
	Data type:	Factor
	Unique non-missing value count:	9
	Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
E1	11,878	11,878	7.06
E2	23,586	35,464	14.01
E3	17,038	52,502	10.12
E4	16,009	68,511	9.51
E5	17,174	85,685	10.20
E6	19,068	104,753	11.33
E7	19,770	124,523	11.75
E8	25,319	149,842	15.04
E9	18,454	168,296	10.97

16	**Column name:**	**geotype2**
	Column description:	Urban/rural
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	18

Categories	Frequency	Cumulative Frequency	Percent
Non-urban	39,859	39,859	23.68
Urban	128,419	168,278	76.31
Missing	18	168,296	0.01

17	**Column name:**	**intm**
	Column description:	Interview - Month
	Data type:	Factor

Unique non-missing value count:	12
Missing value count:	24,042

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
December	6877	February	12338
August	11568	March	12733
May	11985	January	13363
June	12026	November	15001
September	12028	Missing	24042

18	**Column name:**	**intq**

Column description:	Interview - Quarter
Data type:	Factor
Unique non-missing value count:	4
Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
First quarter of year	45,147	45,147	26.83
Fourth quarter of year	39,969	85,116	23.75
Second quarter of year	41,792	126,908	24.83
Third quarter of year	41,388	168,296	24.59

19	**Column name:**	**inty**

Data type:	Numeric
Source information	Year of interview
Unique non-missing value count:	17
Missing value count:	0

Min	Mean	Median	Max	SD
1998.00	2007.53	2008.00	2017.00	5.66

20	**Column name:**	**Vism**

	Column description:	Anthropometry - Month
	Source information:	Month of anthropometric measurements (nurse visits).
	Data type:	Factor
	Unique non-missing value count:	12
	Missing value count:	67,136

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
December	5088	November	9048
August	7878	February	9113
April	8231	March	9375
July	8374	January	9456
September	8603	Missing	67136

21	**Column name:**	**visq**
	Column description:	Anthropometry - Quarter
	Source information:	Quarter of anthropometric measurements (nurse visit).
	Data type:	Factor
	Unique non-missing value count:	4
	Missing value count:	46,008

Categories	Frequency	Cumulative Frequency	Percent
I	34,240	34,240	20.35
II	30,759	64,999	18.28
III	29,904	94,903	17.77
IV	27,385	122,288	16.27
Missing	46,008	168,296	27.34

22	**Column name:**	**hh_size**
	Column description:	Household size
	Data type:	Numeric
	Unique non-missing value count:	12

	Missing value count:	24,042		

Min	Mean	Median	Max	SD
1.00	2.68	2.00	12.00	1.34

23	Column name:	hh_size_cat
	Column description:	Household size, categorical
	Data type:	Factor
	Unique non-missing value count:	6
	Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
1	29,388	29,388	17.46
2	64,558	93,946	38.36
3	29,872	123,818	17.75
4	29,173	152,991	17.33
5	10,338	163,329	6.14
6+	4,967	168,296	2.95

24	Column name:	hh_ownhome
	Column description:	Dwelling - Ownership
	Source Information	Mortgage, own outright & shared ownership have been considered as ownership (‘Yes’ 1). Renting, living rent free or squatting have been considered as not owning (‘No’ 0).
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	502

Categories	Frequency	Cumulative Frequency	Percent
No	47,668	47,668	28.32
Yes	120,126	167,794	71.38
Missing	502	168,296	0.30

25	Column name:	hh_ass_car_truck
	Column description:	Household assets: Car/Truck
	Source information:	Derived from HSE variable ‘car’ indicating whether the household has access to a car. Those with item not applicable were coded as having no car/truck.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	65

Categories	Frequency	Cumulative Frequency	Percent
No	32,859	32,859	19.52
Yes	135,372	168,231	80.44
Missing	65	168,296	0.04

26	Column name:	hh_carnum
	Column description:	Number of cars available
	Source information:	Derived from HSE variable ‘numcars’.
	Data type:	Factor
	Unique non-missing value count:	4
	Missing value count:	102

Categories	Frequency	Cumulative Frequency	Percent
None	32,859	32,859	19.52
One	70,978	103,837	42.17
Three or more	13,769	117,606	8.18
Two	50,588	168,194	30.06
Missing	102	168,296	0.06

27	Column name:	hh_recgrant
	Column description:	Household receives government support.
	Source Information	Government support includes income from housing benefits, council tax benefits, family credit, child tax credit, other state benefits, job-seekers allowance or employment and support allowance. Those who

		reported receiving at least one of the above benefits in the household were coded as 'Yes'. Missing responses included refusals, 'don't know' responses and where the item was not applicable because the respondent was not the household head or spouse/partner of the household head. All others were coded 'No'.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	6,665

Categories	Frequency	Cumulative Frequency	Percent
No	114,078	114,078	67.78
Yes	47,553	161,631	28.26
Missing	6,665	168,296	3.96

28	**Column name:**	**hh_income**
	Column description:	Household income
	Source information:	In British Pounds [GBP]. This was based on the categorical variable 'totinc' in HSE surveys 1998-2015. Following the method used to derive equivalised income in HSE, the midpoint of each category was taken using the lower and upper values of each category as boundaries to give an estimate or average household income; for the lowest category (<£520) a value of £450 was used and for the highest category of income (£150,000+), the value was estimated at £160,000. Refusals, no answer, unknown incomes and cases where the item or schedule was not applicable (e.g. interviewee was not the household head/respondent or spouse), were coded as missing. Incomes have not been deflated or adjusted for purchasing power parity. For 2015, this variable is only available in the Special Licence dataset.
	Data type:	Numeric
	Unique non-missing value count:	31
	Missing value count:	54,611

Min	Mean	Median	Max	SD
450.00	32660.70	24700.00	160000.00	29230.48

29	**Column name:**	**hh_income_quint**

	Column description:	Household income quintile
	Source information	Based on equivalised household income as derived in the HSE. This is a relative measure specific to each year and does not necessarily have the same meaning across time.
	Data type:	Factor
	Unique non-missing value count:	5
	Missing value count:	31,413

Categories	Frequency	Cumulative Frequency	Percent
I (Lowest)	24,819	24,819	14.75
II	26,329	51,148	15.64
III	27,941	79,089	16.60
IV	28,783	107,872	17.10
V (Highest)	29,011	136,883	17.24
Missing	31,413	168,296	18.67

30	**Column name:**	**hh_income_eq**
	Column description:	Household income, equivalised
	Source information:	In GBP. As defined in the HSE (eqvinc). For 2015, this variable is only available in the Special Licence dataset.
	Data type:	Numeric
	Unique non-missing value count:	4,212
	Missing value count:	50,697

Min	Mean	Median	Max	SD
175.10	28613.91	21017.96	262295.08	26197.68

31	**Column name:**	**sex**
	Column description:	Sex
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
Female	93,173	93,173	55.36
Male	75,123	168,296	44.64

32	Column name:	age
	Column description:	Age
	Source information:	For HSE survey years up to HSE 2014, the actual age in years was used. For 2014, the actual age was used through age 89. For those aged 90+, their age was imputed at age 92 years, which was the nearest integer mean age of all adults aged 90+ years in the previous years of the survey considered (1998, 2001-2003, 2005-2013). For HSE 2015 to 2017, the categorical variable Age16g5 was used to estimate the ages of adults. The midpoint of each category was used to impute age in years. For those ages 16-17 years, their age was estimated at 16.5 years. For those ages 18-19 years, their age was estimated at 18.5 years. For the 5-year age categories beginning at age 20, the midpoint was also taken as the estimated age (e.g., 22 for age group 20-24, 27 for age group 25-29, etc.) For those aged, 90+, their age was imputed at age 92 years.
	Data type:	Numeric
	Unique non-missing value count:	88
	Missing value count:	0

Min	Mean	Median	Max	SD
16.00	49.01	48.00	104.00	18.68

33	Column name:	agecat1
	Column description:	Age category (5 years)
	Data type:	Factor
	Unique non-missing value count:	14
	Missing value count:	0

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
75-79	8029	30-34	13824
16-19	8273	50-54	14201

20-24	9548	45-49	14498
80+	9575	35-39	15055
70-74	9994	40-44	15277

34	**Column name:**	**agecat2**
	Column description:	Age category (10 years)
	Data type:	Factor
	Unique non-missing value count:	8
	Missing value count:	0

Categories	Frequency	Cumulative Frequency	Percent
16-19	8,273	8,273	4.92
20-29	21,157	29,430	12.57
30-39	28,879	58,309	17.16
40-49	29,775	88,084	17.69
50-59	27,620	115,704	16.41
60-69	24,994	140,698	14.85
70-79	18,023	158,721	10.71
80+	9,575	168,296	5.69

35	**Column name:**	**race_e**
	Column description:	Ethnicity
	Source information:	The four categories are based on those used in the HSE and have been recoded to align with the order in the South African dataset.
	Data type:	Factor
	Unique non-missing value count:	5
	Missing value count:	534

Categories	Frequency	Cumulative Frequency	Percent
Asian	8,144	8,144	4.84
Black	4,124	12,268	2.45
Mixed	1,574	13,842	0.94

Other	1,523	15,365	0.90
White	152,397	167,762	90.55
Missing	534	168,296	0.32

36	**Column name:**	**marstatus**
	Column description:	Marital status
	Data type:	Factor
	Unique non-missing value count:	3
	Missing value count:	52

Categories	Frequency	Cumulative Frequency	Percent
Married/living with partner	105,394	105,394	62.62
Never married/single	34,150	139,544	20.29
Widowed/divorced/ separated	28,700	168,244	17.05
Missing	52	168,296	0.03

37	**Column name:**	**edu3**
	Column description:	Education: categorisation 3
	Source information:	Highest level of education. Those for whom the item was not applicable were coded as having no qualification.
	Data type:	Factor
	Unique non-missing value count:	7
	Missing value count:	12

Categories	Frequency	Cumulative Frequency	Percent
No qualification	44,499	66,935	26.44
NVQ1/CSE other grade equiv	8,047	74,982	4.78
NVQ2/GCE O Level equiv	36,986	111,968	21.98
NVQ3/GCE A Level equiv	22,533	134,501	13.39

NVQ4/NVQ5/Degree or equiv	33,783	168,284	20.07
Higher ed below degree	18,115	18,115	10.76
Foreign/Other	4,321	22,436	2.57
Missing	12	168,296	0.01

38	**Column name:**	**emp**
	Column description:	Employment
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	17,261

Categories	Frequency	Cumulative Frequency	Percent
Employed	142,564	142,564	84.71
Unemployed	8,471	151,035	5.03
Missing	17,261	168,296	10.26

39	**Column name:**	**occupation**
	Source information:	Derived from sclass and sclass12 ‘Armed forces’ (sclass) and ‘6’ (sclass12) were grouped with other.
	Data type:	Factor
	Unique non-missing value count:	8
	Missing value count:	17,261

Categories	Frequency	Cumulative Frequency	Percent
Unemployed	8,471	143,299	5.03
Unskilled manual	7,736	151,035	4.60
Semi-skilled manual	25,300	76,563	15.03
Skilled manual	25,124	101,687	14.93
Skilled non-manual	33,141	134,828	19.69
Managerial technical	43,268	43,268	25.71
Professional	7,483	51,263	4.45

Other/not fully described	512	43,780	0.30
Missing	17,261	168,296	10.26

40	Column name:	smokstatus
	Column description:	Smoking status
	Data type:	Factor
	Unique non-missing value count:	3
	Missing value count:	1,092

Categories	Frequency	Cumulative Frequency	Percent
Current smoker	36,493	36,493	21.68
Former smoker	44,087	80,580	26.20
Never smoker	86,624	167,204	51.47
Missing	1,092	168,296	0.65

41	Column name:	currsmok
	Column description:	Current smoker
	Source information:	Non-smokers include former smokers.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	1,092

Categories	Frequency	Cumulative Frequency	Percent
No	130,711	130,711	77.67
Yes	36,493	167,204	21.68
Missing	1,092	168,296	0.65

42	Column name:	alcstatus
	Column description:	Alcohol status
	Source information:	Current drinkers include those who drink very occasionally. Never drinkers are those reporting never drinking and always being a non-drinker. Former

		drinkers are those who reported never drinking nowadays and that they used to drink but stopped.
	Data type:	Factor
	Unique non-missing value count:	3
	Missing value count:	1,091

Categories	Frequency	Cumulative Frequency	Percent
Current drinker	145,726	145,726	86.59
Former drinker	8,730	154,456	5.19
Never drinker	12,749	167,205	7.58
Missing	1,091	168,296	0.65

43	**Column name:**	**curralc**
	Column description:	Current drinker
	Source information:	Current drinkers include those who drink very occasionally. Never drinkers are those reporting never drinking and always being a non-drinker. Former drinkers are those who reported never drinking nowadays and that they used to drink but stopped.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	1,064

Categories	Frequency	Cumulative Frequency	Percent
Current drinker	145,726	145,726	86.59
Non-drinker	21,506	167,232	12.78
Missing	1,064	168,296	0.63

44	**Column name:**	**alcmax**
	Column description:	Alcohol consumption on heaviest day [units]
	Source information:	Number of units of alcohol consumed on heaviest day of drinking in the past 7 days.
	Data type:	Factor
	Unique non-missing value count:	4
	Missing value count:	1,673

Categories	Frequency	Cumulative Frequency	Percent
<=4 units/day (men), <=3 units/day (women)	55,248	55,248	32.83
>4 and <=8 units/day (men), >3 and <=6 units/day (women)	25,851	81,099	15.36
>8 units/day (men), >6 units/day (women)	24,623	105,722	14.63
None	60,901	166,623	36.19
Missing	1,673	168,296	0.99

45	**Column name:**	**fruitveg**
	Column description:	5+ portions of fruit/vegetables eaten yesterday
	Source information:	Only available from 2001 onward.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	32,212

Categories	Frequency	Cumulative Frequency	Percent
No	99,166	99,166	58.92
Yes	36,918	136,084	21.94
Missing	32,212	168,296	19.14

46	**Column name:**	**Exercisefreq_e**
	Column description:	Weekly frequency of 30+ min moderate or more activity
	Data type:	Factor
	Unique non-missing value count:	
	Missing value count:	

Categories	Frequency	Cumulative Frequency	Percent
Never			
Less than once per week			
Once or twice per week			

Three of four times per week	
Five or more times per week	
Missing	

47	Column name:	self_health
	Column description:	Self-rated health
	Source information:	This was originally coded as: very good, good, fair, bad, very bad. It was recoded and categories renamed to align with the categories created for the harmonized South African dataset.
	Data type:	Factor
	Unique non-missing value count:	4
	Missing value count:	63

Categories	Frequency	Cumulative Frequency	Percent
Poor/bad	12,448	114,038	7.40
Average/fair	31,361	31,361	18.63
Good	70,229	101,590	41.73
Very good/excellent	54,195	168,233	32.20
Missing	63	168,296	0.04

48	Column name:	diag_hbp
	Column description:	Diagnosis: Hypertension
	Source information:	Available all years except 2001, 2002, 2007 & 2008. Based on 2 questions, whether you have ever had the condition and whether it was diagnosed by a doctor/(nurse). Those who responded ‘no’, ‘don’t know’ or ‘refused’ to ever having were not asked about doctor diagnosis. Those who responded ‘don’t know’ or ‘refused’ to ever having had the condition were coded as missing here. Those under age 65 in 2005 and those not receiving the CVD module in 2006 were coded as missing.
	Data type:	Factor
	Unique non-missing value count:	2

Missing value count:	53,051	

Categories	Frequency	Cumulative Frequency	Percent
No	83,630	83,630	49.69
Yes	31,615	115,245	18.79
Missing	53,051	168,296	31.52

49	**Column name:**	**diag_isch**
	Column description:	Diagnosis: Heart attack/angina
	Source information:	Only available for the CVD years [1998, 2003, 2005 (65+), 2006 (CVD module) 2011, 2013, 2017]. Based on 2 questions, whether you have ever had the condition and whether it was diagnosed by a doctor/(nurse). Those who responded ‘no’, ‘don’t know’ or ‘refused’ to ever having were not asked about doctor diagnosis. Those who responded ‘don’t know’ or ‘refused’ to ever having had the condition were coded as missing here. Those under age 65 in 2005 and those not receiving the CVD module in 2006 were coded as missing.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	107,054

Categories	Frequency	Cumulative Frequency	Percent
No	57,720	57,720	34.30
Yes	3,522	61,242	2.09
Missing	107,054	168,296	63.61

50	**Column name:**	**diag_stroke**
	Column description:	Diagnosis: Stroke
	Source information:	Only available for the CVD years [1998, 2003, 2005 (65+), 2006 (CVD module) 2011, 2013, 2017]. Based on 2 questions, whether you have ever had the condition and whether it was diagnosed by a doctor/(nurse). Those who responded ‘no’, ‘don’t know’ or ‘refused’ to ever having were not asked about doctor diagnosis. Those who responded ‘don’t know’ or ‘refused’ to ever having had the condition were coded as missing here. Those under age 65 in 2005 and those not receiving the CVD module in 2006 were coded as missing.

Data type:	Factor
Unique non-missing value count:	2
Missing value count:	106,987

Categories	Frequency	Cumulative Frequency	Percent
No	59,754	59,754	35.51
Yes	1,555	61,309	0.92
Missing	106,987	168,296	63.57

51	**Column name:**	**diag_diab**
	Column description:	Diagnosis: Diabetes/hyperglycaemia
	Source information:	Available all years except 2001, 2002, 2007 & 2008. Based on 2 questions, whether you have ever had the condition and whether it was diagnosed by a doctor/(nurse). Those who responded ‘no’, ‘don’t know’ or ‘refused’ to ever having were not asked about doctor diagnosis. Those who responded ‘don’t know’ or ‘refused’ to ever having had the condition were coded as missing here. Those under age 65 in 2005 and those not receiving the CVD module in 2006 were coded as missing.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	52,746

Categories	Frequency	Cumulative Frequency	Percent
No	108,404	108,404	64.41
Yes	7,146	115,550	4.25
Missing	52,746	168,296	31.34

52	**Column name:**	**diag_cancer**
	Column description:	Diagnosis: Cancer
	Source information:	Derived from the list of complaints from the limiting longstanding illnesses reported according to ICD categories. They are available for all years until 2014. For 2015, they are only available in the Special Licence dataset. However, they do not necessarily specify whether they were diagnosed by a doctor/nurse or not, so it is assumed that most were diagnosed by a

		doctor/nurse. In addition, respondents are only able to report up to 6 conditions. Therefore, those who have more conditions would only have the first 6 recorded.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	8,101

Categories	Frequency	Cumulative Frequency	Percent
No	156,931	156,931	93.25
Yes	3,264	160,195	1.94
Missing	8,101	168,296	4.81

53	**Column name:**	**diag_heart**
	Column description:	Diagnosis: Heart problems
	Source information:	Heart and circulatory conditions. Derived from the list of complaints from the limiting longstanding illnesses reported according to ICD categories. They are available for all years until 2014. For 2015, they are only available in the Special Licence dataset. However, they do not necessarily specify whether they were diagnosed by a doctor/nurse or not, so it is assumed that most were diagnosed by a doctor/nurse. In addition, respondents are only able to report up to 6 conditions. Therefore, those who have more conditions would only have the first 6 recorded.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	8,101

Categories	Frequency	Cumulative Frequency	Percent
No	139,659	139,659	82.98
Yes	20,536	160,195	12.20
Missing	8,101	168,296	4.81

54	**Column name:**	**diag_diab2**
	Column description:	Diagnosis: diabetes, excluding pregnancy
	Source information:	Based on the derived HSE variable diabete2. Those who responded ‘no’, ‘don’t know’ or ‘refused’ to ever having were not asked about doctor diagnosis. Those

		who responded ‘don’t know’ or ‘refused’ to ever having had the condition were coded as not having a doctor diagnosis, following the method of HSE. This differs from our derived variables above of diagnosis including in pregnancy.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	52,644

Categories	Frequency	Cumulative Frequency	Percent
No	108,990	108,990	64.76
Yes	6,662	115,652	3.96
Missing	52,644	168,296	31.28

55	**Column name:**	**diag_hbp2**
	Column description:	Diagnosis: hypertension, excluding pregnancy
	Source information:	Based on the derived HSE variable bp1. Those who responded ‘no’, ‘don’t know’ or ‘refused’ to ever having were not asked about doctor diagnosis. Those who responded ‘don’t know’ or ‘refused’ to ever having had the condition were coded as not having a doctor diagnosis, following the method of HSE. This differs from our derived variables above of diagnosis including in pregnancy.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	52,712

Categories	Frequency	Cumulative Frequency	Percent
No	86,925	86,925	51.65
Yes	28,659	115,584	17.03
Missing	52,712	168,296	31.32

56	**Column name:**	**diag_angi**
	Column description:	Diagnosis: angina
	Source information:	Only available for the CVD years [1998, 2003, 2005 (65+), 2006 (CVD module) 2011, 2013, 2017]. Based on 2 questions, whether you have ever had the condition and whether it was diagnosed by a doctor/(nurse). Those

		who responded ‘no’, ‘don’t know’ or ‘refused’ to ever having were not asked about doctor diagnosis. Those who responded ‘don’t know’ or ‘refused’ to ever having had the condition were coded as missing here. Those under age 65 in 2005 and those not receiving the CVD module in 2006 were coded as missing.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	107,053

Categories	Frequency	Cumulative Frequency	Percent
No	58,582	58,582	34.81
Yes	2,661	61,243	1.58
Missing	107,053	168,296	63.61

57	**Column name:**	**diag_mi**
	Column description:	Diagnosis: heart attack (myocardial infarction)
	Source information:	Only available for the CVD years [1998, 2003, 2005 (65+), 2006 (CVD module) 2011, 2013, 2017]. Based on 2 questions, whether you have ever had the condition and whether it was diagnosed by a doctor/(nurse). Those who responded ‘no’, ‘don’t know’ or ‘refused’ to ever having were not asked about doctor diagnosis. Those who responded ‘don’t know’ or ‘refused’ to ever having had the condition were coded as missing here. Those under age 65 in 2005 and those not receiving the CVD module in 2006 were coded as missing.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	106,981

Categories	Frequency	Cumulative Frequency	Percent
No	59,408	59,408	35.30
Yes	1,907	61,315	1.13
Missing	106,981	168,296	63.57

58	**Column name:**	**diag_lung**
	Column description:	Diagnosis: Respiratory condition

	Source information:	Derived from the list of complaints from the limiting longstanding illnesses reported according to ICD categories. They are available for all years until 2014. For 2015, they are only available in the Special Licence dataset. However, they do not necessarily specify whether they were diagnosed by a doctor/nurse or not, so it is assumed that most were diagnosed by a doctor/nurse. In addition, respondents are only able to report up to 6 conditions. Therefore, those who have more conditions would only have the first 6 recorded.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	8,101

Categories	Frequency	Cumulative Frequency	Percent
No	146,194	146,194	86.87
Yes	14,001	160,195	8.32
Missing	8,101	168,296	4.81

59	**Column name:**	**diag_mental**
	Column description:	Diagnosis: Mental disorder
	Source information:	Derived from the list of complaints from the limiting longstanding illnesses reported according to ICD categories. They are available for all years until 2014. For 2015, they are only available in the Special Licence dataset. However, they do not necessarily specify whether they were diagnosed by a doctor/nurse or not, so it is assumed that most were diagnosed by a doctor/nurse. In addition, respondents are only able to report up to 6 conditions. Therefore, those who have more conditions would only have the first 6 recorded.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	8,101

Categories	Frequency	Cumulative Frequency	Percent
No	153,269	153,269	91.07
Yes	6,926	160,195	4.12
Missing	8,101	168,296	4.81

60	Column name:	diag_infectious
	Column description:	Diagnosis: Infectious disorder
	Source information:	Derived from the list of complaints from the limiting longstanding illnesses reported according to ICD categories. They are available for all years until 2014. For 2015, they are only available in the Special Licence dataset. However, they do not necessarily specify whether they were diagnosed by a doctor/nurse or not, so it is assumed that most were diagnosed by a doctor/nurse. In addition, respondents are only able to report up to 6 conditions. Therefore, those who have more conditions would only have the first 6 recorded.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	8,101

Categories	Frequency	Cumulative Frequency	Percent
No	159,892	159,892	95.01
Yes	303	160,195	0.18
Missing	8,101	168,296	4.81

61	Column name:	diag_metabolic
	Column description:	Diagnosis: Endocrine/metabolic disorder
	Source information:	Derived from the list of complaints from the limiting longstanding illnesses reported according to ICD categories. They are available for all years until 2014. For 2015, they are only available in the Special Licence dataset. However, they do not necessarily specify whether they were diagnosed by a doctor/nurse or not, so it is assumed that most were diagnosed by a doctor/nurse. In addition, respondents are only able to report up to 6 conditions. Therefore, those who have more conditions would only have the first 6 recorded.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	8,101

Categories	Frequency	Cumulative Frequency	Percent
No	147,445	147,445	87.61
Yes	12,750	160,195	7.58

Missing	8,101	168,296	4.81

62	Column name:	diag_nerve
	Column description:	Diagnosis: Nervous system disorder
	Source information:	Derived from the list of complaints from the limiting longstanding illnesses reported according to ICD categories. They are available for all years until 2014. For 2015, they are only available in the Special Licence dataset. However, they do not necessarily specify whether they were diagnosed by a doctor/nurse or not, so it is assumed that most were diagnosed by a doctor/nurse. In addition, respondents are only able to report up to 6 conditions. Therefore, those who have more conditions would only have the first 6 recorded.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	8,101

Categories	Frequency	Cumulative Frequency	Percent
No	153,707	153,707	91.33
Yes	6,488	160,195	3.86
Missing	8,101	168,296	4.81

63	Column name:	diag_blood
	Column description:	Diagnosis: Blood disorder
	Source information:	Derived from the list of complaints from the limiting longstanding illnesses reported according to ICD categories. They are available for all years until 2014. For 2015, they are only available in the Special Licence dataset. However, they do not necessarily specify whether they were diagnosed by a doctor/nurse or not, so it is assumed that most were diagnosed by a doctor/nurse. In addition, respondents are only able to report up to 6 conditions. Therefore, those who have more conditions would only have the first 6 recorded.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	8,101

Categories	Frequency	Cumulative Frequency	Percent

No	158,869	158,869	94.40
Yes	1,326	160,195	0.79
Missing	8,101	168,296	4.81

64	Column name:	bpmed
	Column description:	Medication: Hypertension
	Source information:	Based on only those diagnosed with high BP apart from when pregnant. Not available for 2001, 2002 and 2016.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	60,699

Categories	Frequency	Cumulative Frequency	Percent
No	89,594	89,594	53.24
Yes	18,003	107,597	10.70
Missing	60,699	168,296	36.07

65	Column name:	bpmed_coded
	Column description:	Medication: Hypertension, coded
	Source information:	Those with no answer/refused/unable to code (331) are coded as missing.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	46,853

Categories	Frequency	Cumulative Frequency	Percent
No	102,597	102,597	60.96
Yes	18,846	121,443	11.20
Missing	46,853	168,296	27.84

66	Column name:	diabmed
	Column description:	Medication: Diabetes/hyperglicaemia
	Data type:	Factor

Unique non-missing value count:	2	
Missing value count:	46,611	

Categories	Frequency	Cumulative Frequency	Percent
No	117,113	117,113	69.59
Yes	4,572	121,685	2.72
Missing	46,611	168,296	27.70

67	**Column name:**	**cholmed**
	Column description:	Medication: Hypercholesterolaemia
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	47,457

Categories	Frequency	Cumulative Frequency	Percent
No	106,423	106,423	63.24
Yes	14,416	120,839	8.57
Missing	47,457	168,296	28.20

68	**Column name:**	**heartmed**
	Column description:	Medication: Cardiovascular
	Source information:	Cardiovascular medications taken
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	46079

Categories	Frequency	Cumulative Frequency	Percent
No	91387	91,387	54.3%
Yes	30830	122,217	18.3%
Missing	46079	168,296	27.4%

69	**Column name:**	**heartmed2**

	Column description:	Medication: Heart condition or stroke
	Source information:	Only available in CVD years (1998, 2003, 2005 (65+), 2006, 2011, 2013, 2017)
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	

Categories	Frequency	Cumulative Frequency	Percent
No	62,436	62,436	37.1%
Yes	4947	67,383	2.9%
Missing	100,886	168,269	59.9%

70	**Column name:**	**contraceptives**
	Column description:	Medication: Contraceptives
	Source information:	Females only. Item not applicable coded as 'No' (0), Schedule not obtained coded as missing.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	115,924

Categories	Frequency	Cumulative Frequency	Percent
No	44,923	44,923	26.69
Yes	7,449	52,372	4.43
Missing	115,924	168,296	68.88

71	**Column name:**	**currpreg**
	Column description:	Currently pregnant
	Source information:	This refers to pregnancy at the time of the nurse interview and is valid for females only. Females over age 49 were not asked this question and were given a response of 'No' (0). Females under age 50 who did not participate in the nurse schedule would have a missing

		response. Item not applicable coded as ‘No’ (0), Schedule not obtained coded as missing.
	Data type:	Factor
	Unique non-missing value count:	2
	Missing value count:	79,096

Categories	Frequency	Cumulative Frequency	Percent
No/don't know	87,938	87,938	52.25
Yes	1,262	89,200	0.75
Missing	79,096	168,296	47.00

72	**Column name:**	**height1**
	Column description:	Height [cm] - First reading
	Source Information	Implausible values: Height<120 cm or height > 220 cm
	Data type:	Numeric
	Unique non-missing value count:	706
	Missing value count:	19,576

Min	Mean	Median	Max	SD
121.00	167.45	167.00	210.30	9.68

73	**Column name:**	**height**
	Column description:	Height [cm] - Average of available readings
	Data type:	Numeric
	Unique non-missing value count:	706
	Missing value count:	19,576

Min	Mean	Median	Max	SD
121.00	167.45	167.00	210.30	9.68

74	**Column name:**	**weight1**
	Column description:	Weight [Kg] - First reading

	Source Information	Weights above 130 kg were estimated (self-reported) in the HSE as the scales did not go above those values. From 2014, weights above 200 kg were estimated (self-reported). Implausible values: Females: Weight<25 Kg; Weight > 250 Kg Males: Weight <35 Kg; Weight > 250 Kg [From the 2nd South African Comparative Risk Assessment (SACRA) study]
	Data type:	Numeric
	Unique non-missing value count:	1,240
	Missing value count:	21,586

Min	Mean	Median	Max	SD
26.00	76.46	74.60	205.00	16.84

75	**Column name:**	**Weight**
	Column description:	Weight [Kg] - Average of available readings
	Data type:	Numeric
	Unique non-missing value count:	1,240
	Missing value count:	21,586

Min	Mean	Median	Max	SD
26.00	76.46	74.60	205.00	16.84

76	**Column name:**	**waist1**
	Column description:	Waist circumference [cm] - First reading
	Source Information	Implausible values considered: waist < 30 cm or waist > 220 cm.
	Data type:	Numeric
	Unique non-missing value count:	969
	Missing value count:	49,628

Min	Mean	Median	Max	SD
34.00	92.04	91.40	172.20	14.33

77	Column name:	waist2
	Column description:	Waist circumference [cm] - Second reading
	Data type:	Numeric
	Unique non-missing value count:	972
	Missing value count:	49,842

Min	Mean	Median	Max	SD
34.10	92.14	91.50	173.50	14.34

78	Column name:	waist3
	Column description:	Waist circumference [cm] - Third reading
	Data type:	Numeric
	Unique non-missing value count:	297
	Missing value count:	167,870

Min	Mean	Median	Max	SD
64.80	97.37	95.85	153.50	15.02

79	Column name:	waist
	Column description:	Waist circumference [cm] - Average of available readings
	Data type:	Numeric
	Unique non-missing value count:	2,082
	Missing value count:	49,627

Min	Mean	Median	Max	SD
34.05	92.10	91.50	172.85	14.33

80	Column name:	hip1
	Column description:	Hip circumference [cm] - First reading
	Source Information	Implausible values: (Based on the distribution in the sample) Hip circumference < 40 cm or hip circumference > 230 cm

	Data type:	Numeric
	Unique non-missing value count:	890
	Missing value count:	49,712

Min	Mean	Median	Max	SD
50.00	105.44	104.20	181.20	10.20

81	**Column name:**	**hip2**
	Column description:	Hip circumference [cm] - Second reading
	Data type:	Numeric
	Unique non-missing value count:	887
	Missing value count:	49,924

Min	Mean	Median	Max	SD
50.10	105.46	104.20	190.00	10.21

82	**Column name:**	**hip3**
	Column description:	Hip circumference [cm] - Third reading
	Data type:	Numeric
	Unique non-missing value count:	140
	Missing value count:	168,123

Min	Mean	Median	Max	SD
84.80	111.10	109.90	190.00	14.34

83	**Column name:**	**hip**
	Column description:	Hip circumference [cm] - Average of available readings
	Data type:	Numeric
	Unique non-missing value count:	1,711
	Missing value count:	49,710

Min	Mean	Median	Max	SD
50.05	105.46	104.25	181.25	10.21

84	Column name:	sbp1
	Column description:	Systolic Blood Pressure [mmHg] - First reading
	Source Information	Based on the original HSE readings (including invalid measurements) Implausible values excluded: SBP < 60 mmHg or SBP > 270 mmHg SBP readings were set to missing if they were less than 15 mmHg greater than the corresponding DBP reading.
	Data type:	Numeric
	Unique non-missing value count:	314
	Missing value count:	52,791

Min	Mean	Median	Max	SD
61.00	130.69	128.11	246.00	18.84

85	Column name:	sbp2
	Column description:	Systolic Blood Pressure [mmHg] - Second reading
	Source Information	Based on the original HSE readings (including invalid measurements) Implausible values excluded: SBP < 60 mmHg or SBP > 270 mmHg SBP readings were set to missing if they were less than 15 mmHg greater than the corresponding DBP reading.
	Data type:	Numeric
	Unique non-missing value count:	323
	Missing value count:	53,113

Min	Mean	Median	Max	SD
60.00	128.35	126.00	247.00	18.23

86	Column name:	sbp3
	Column description:	Systolic Blood Pressure [mmHg] - Third reading
	Source Information	Based on the original HSE readings (including invalid measurements) Implausible values excluded: SBP < 60 mmHg or SBP > 270 mmHg

		SBP readings were set to missing if they were less than 15 mmHg greater than the corresponding DBP reading.
	Data type:	Numeric
	Unique non-missing value count:	321
	Missing value count:	53,559

Min	Mean	Median	Max	SD
62.00	127.50	125.38	240.00	17.77

87	**Column name:**	**sbp_mean1**
	Column description:	Systolic Blood Pressure [mmHg] - Average of available readings
	Data type:	Numeric
	Unique non-missing value count:	1,096
	Missing value count:	52,737

Min	Mean	Median	Max	SD
70.00	128.87	126.67	243.67	17.81

88	**Column name:**	**sbp_mean2**
	Column description:	Systolic Blood Pressure [mmHg] - Average of available readings excluding the first
	Data type:	Numeric
	Unique non-missing value count:	688
	Missing value count:	53,069

Min	Mean	Median	Max	SD
69.50	127.93	125.83	242.50	17.70

89	**Column name:**	**dbp1**
	Column description:	Diastolic Blood Pressure [mmHg] - First reading
	Source Information	Based on the original HSE readings (including invalid measurements) Implausible values excluded: DBP < 30 mmHg; DBP > 150 mmHg.

		DBP readings were set to missing if they were less than 15 mmHg lower than the corresponding SBP reading.
	Data type:	Numeric
	Unique non-missing value count:	168
	Missing value count:	61,999

Min	Mean	Median	Max	SD
60.00	76.27	75.00	160.00	9.99

90	**Column name:**	**dbp2**
	Column description:	Diastolic Blood Pressure [mmHg] - Second reading
	Source Information	Based on the original HSE readings (including invalid measurements) Implausible values excluded: DBP < 30 mmHg; DBP > 150 mmHg. DBP readings were set to missing if they were less than 15 mmHg lower than the corresponding SBP reading.
	Data type:	Numeric
	Unique non-missing value count:	160
	Missing value count:	63,468

Min	Mean	Median	Max	SD
60.00	75.34	74.00	162.00	9.58

91	**Column name:**	**dbp3**
	Column description:	Diastolic Blood Pressure [mmHg] - Third reading
	Source Information	Based on the original HSE readings (including invalid measurements) Implausible values excluded: DBP < 30 mmHg; DBP > 150 mmHg. DBP readings were set to missing if they were less than 15 mmHg lower than the corresponding SBP reading.
	Data type:	Numeric
	Unique non-missing value count:	161
	Missing value count:	64,701

Min	Mean	Median	Max	SD
60.00	75.01	74.00	158.00	9.49

92	Column name:	dbp_mean1
	Column description:	Diastolic Blood Pressure [mmHg] - Average of available readings
	Data type:	Numeric
	Unique non-missing value count:	595
	Missing value count:	58,383

Min	Mean	Median	Max	SD
60.00	75.07	74.00	160.00	9.42

93	Column name:	dbp_mean2
	Column description:	Diastolic Blood Pressure [mmHg] - Average of available readings excluding the first
	Data type:	Numeric
	Unique non-missing value count:	353
	Missing value count:	60,714

Min	Mean	Median	Max	SD
60.00	74.86	73.80	160.00	9.37

94	Column name:	rhr1
	Column description:	Resting Heart Rate [ppm] - First reading
	Source Information	Biologically implausible values: RHR < 20 bpm; RHR > 250 bpm. (Informed by clinical opinion)
	Data type:	Numeric
	Unique non-missing value count:	112
	Missing value count:	48,855

Min	Mean	Median	Max	SD
31.00	70.13	69.00	160.00	11.59

95	Column name:	rhr2
	Column description:	Resting Heart Rate [ppm] - Second reading
	Data type:	Numeric
	Unique non-missing value count:	109
	Missing value count:	49,249

Min	Mean	Median	Max	SD
32.00	70.36	70.00	159.00	11.41

96	Column name:	rhr3
	Column description:	Resting Heart Rate [ppm] - Third reading
	Data type:	Numeric
	Unique non-missing value count:	108
	Missing value count:	59,880

Min	Mean	Median	Max	SD
31.00	70.66	70.00	148.00	11.33

97	Column name:	rhr_mean1
	Column description:	Resting Heart Rate [ppm] - Average of available readings
	Data type:	Numeric
	Unique non-missing value count:	369
	Missing value count:	48,792

Min	Mean	Median	Max	SD
34.00	70.39	69.67	147.67	11.15

98	Column name:	rhr_mean2
	Column description:	Resting Heart Rate [ppm] - Average of available readings excluding the first
	Data type:	Numeric
	Unique non-missing value count:	203

	Missing value count:	49,212		
Min	Mean	Median	Max	SD
32.00	70.52	70.00	146.50	11.20

99	**Column name:**	**Airtemp**
	Column description:	Air temperature during blood pressure measurement [Degrees Celsius]
	Data type:	Numeric
	Unique non-missing value count:	265
	Missing value count:	53,547

Min	Mean	Median	Max	SD
0.00	20.70	20.60	38.70	2.38

100	**Column name:**	**bmi**
	Column description:	BMI [kg/m^2]
	Source information:	Calculated using the height and weight measurements. Implausible values removed from the dataset: BMI < 10 kg/m^2 or BMI>131 kg/m^2. Ref: Iyen, B., Weng, S., Vinogradova, Y. et al. Long-term body mass index changes in overweight and obese adults and the risk of heart failure, cardiovascular disease and mortality: a cohort study of over 260,000 adults in the UK. BMC Public Health 21, 576 (2021). https://doi.org/10.1186/s12889-021-10606-1.
	Data type:	Numeric
	Unique non-missing value count:	96,331
	Missing value count:	24,757

Min	Mean	Median	Max	SD
11.81	27.17	26.47	69.14	5.23

101	**Column name:**	**bmicat**
	Column description:	BMI category
	Source information:	Underweight = BMI <18.5 kg/m^2. Healthy weight = 18.5 kg/m2 < BMI < 25 kg/m^2. Overweight = 25 kg/m^2 < BMI < 30 kg/m^2. Obesity I = 30 kg/m^2 < BMI < 35

		kg/m^2. Obesity II = 35 kg/m^2 < BMI < 40 kg/m^2. Obesity III = 40 kg/m^2 < BMI
	Data type:	Factor
	Unique non-missing value count:	6
	Missing value count:	24,757

Categories	Frequency	Cumulative Frequency	Percent
Healthy weight	51,413	51,413	30.55
Obesity I	24,345	75,758	14.47
Obesity II	7,915	83,673	4.70
Obesity III	3,271	86,944	1.94
Overweight	54,383	141,327	32.31
Underweight	2,212	143,539	1.31
Missing	24,757	168,296	14.71

102	**Column name:**	**chol_tot**
	Column description:	Total cholesterol [mmol/l] Includes those on lipid lowering drugs.
	Source information:	Implausible values were considered to be those with total cholesterol <1.75 mmol/L (7 individuals), >20 mmol/L (0 individuals), or where total cholesterol was lower than HDL cholesterol (0 individuals), following the method used by the NCD Risk Factor Collaboration. NCD Risk Factor Collaboration (NCD-RisC). (2020). Repositioning of the global epicentre of non-optimal cholesterol. Nature, 582(7810), 73–77. https://doi.org/10.1038/s41586-020-2338-1.
	Data type:	Numeric
	Unique non-missing value count:	105
	Missing value count:	99,947

Min	Mean	Median	Max	SD
1.80	5.35	5.30	15.30	1.16

103	**Column name:**	**chol_hdl**

	Column description:	High-density lipoprotein (HDL) cholesterol [mmol/l] Includes those on lipid lowering drugs.
	Source information:	Implausible values were considered to be those with HDL cholesterol <0.40 mmol/L (9 individuals), >5.00 mmol/L (0 individuals), or where total cholesterol was lower than HDL cholesterol (0 individuals) following the method used by NCD- RisC. NCD Risk Factor Collaboration (NCD-RisC). (2020). Repositioning of the global epicentre of non-optimal cholesterol. Nature, 582(7810), 73–77. https://doi.org/10.1038/s41586-020-2338-1
	Data type:	Numeric
	Unique non-missing value count:	44
	Missing value count:	99,970

Min	Mean	Median	Max	SD
0.40	1.56	1.50	4.90	0.43

104	**Column name:**	**glyhb_h**
	Column description:	Glycated haemoglobin (HbA1c) (%) Based on the valid glycated haemoglobin variables in the HSE.
	Source information:	HbA1c< 2.5% or > 25% (none) were considered implausible (Nicholas et al., 2013: https://doi.org/10.1371/journal.pone.0068008)
	Data type:	Numeric
	Unique non-missing value count:	147
	Missing value count:	110,980

Min	Mean	Median	Max	SD
2.90	5.63	5.50	16.90	0.74

105	**Column name:**	**hbA1c**
	Column description:	HbA1c (mmol/mol)

	Source information:	Glycated haemoglobin in mmol/mol are only available from 2012 onward (variables iffcvala IFCCA1, harmonised in the present dataset as glyhb2_h). For consistency across years, HbA1c was derived for all years from plausible values of the glyhb variable above and converted to mmol/mol using the following equation: mmol/mol = 10.93 * % – 23.5 mmol/mol Australian Diabetes Society. 2023. HbA1c Conversion Table. Available online: https://diabetessociety.com.au/documents/HbA1cConversionTable.pdf (Accessed 25 Oct 2024)
	Data type:	Numeric
	Unique non-missing value count:	
	Missing value count:	

Min	Mean	Median	Max	SD
8.20	38.02	36.62	161.22	8.14

106	**Column name:**	**globorisk_nonlab**
	Column description:	Globorisk CVD non laboratory risk score
	Source information:	10 years risk of developing a fatal or non-fatal cardiovascular event. Function of age, sex, smoking status, systolic blood pressure and BMI. Ref: Ueda P et al. Laboratory-Based and Office-Based Risk Scores and Charts to Predict 10-Year Risk of Cardiovascular Disease in 182 Countries: A Pooled Analysis of Prospective Cohorts and Health Surveys. The Lancet Diabetes & Endocrinology 5, no. 3 (March 1, 2017): 196–213. Calculated using sbp_mean2 variable for systolic blood pressure. Defined for age between 40 and 74 years.
	Data type:	Numeric
	Unique non-missing value count:	61,593
	Missing value count:	106,703

Min	Mean	Median	Max	SD
0.27	8.11	6.03	85.64	7.26

107	**Column name:**	**globorisk_lab**

	Column description:	Globorisk CVD laboratory risk score
	Source information:	10-year risk of developing a fatal or non-fatal cardiovascular event. Function of age, sex, smoking status, systolic blood pressure, total cholesterol and diabetes mellitus status. Ref: Ueda P et al. Laboratory-Based and Office-Based Risk Scores and Charts to Predict 10-Year Risk of Cardiovascular Disease in 182 Countries: A Pooled Analysis of Prospective Cohorts and Health Surveys. The Lancet Diabetes & Endocrinology 5, no. 3 (March 1, 2017): 196–213. Calculated using sbp_mean2 variable for systolic blood pressure. Defined for age between 40 and 74 years.
	Data type:	Numeric
	Unique non-missing value count:	35,625
	Missing value count:	132,443

Min	Mean	Median	Max	SD
0.24	7.49	5.33	89.64	7.22

108	**Column name:**	**globorisk_lab_fatal**
	Column description:	Globorisk CVD fatal risk score
	Source information:	10 years risk of dying from a cardiovascular event. Function of age, sex, smoking status, systolic blood pressure, total cholesterol and diabetes mellitus status. Ref: Ueda P et al. Laboratory-Based and Office-Based Risk Scores and Charts to Predict 10-Year Risk of Cardiovascular Disease in 182 Countries: A Pooled Analysis of Prospective Cohorts and Health Surveys. The Lancet Diabetes & Endocrinology 5, no. 3 (March 1, 2017): 196–213. Calculated using sbp_mean2 variable for systolic blood pressure. Defined for age between 40 and 74 years.
	Data type:	Numeric
	Unique non-missing value count:	35,413
	Missing value count:	132,443

Min	Mean	Median	Max	SD
0.03	2.86	1.38	59.85	4.05

109	**Column name:**	**who_nonlab**
	Column description:	WHO/ISH CVD non-laboratory risk score

	Source information:	10 years risk of developing a fatal or non-fatal cardiovascular event. Function of age, sex, smoking status, systolic blood pressure and BMI. Ref: Kaptoge S et Al. (2019). World Health Organization cardiovascular disease risk charts: Revised models to estimate risk in 21 global regions. The Lancet Global Health, 7(10), e1332–e1345. https://doi.org/10.1016/S2214-109X(19)30318-3. Calculated using sbp_mean2 variable for systolic blood pressure. Defined for age between 40 and 80 years.
	Data type:	Numeric
	Unique non-missing value count:	66,439
	Missing value count:	101,812

Min	Mean	Median	Max	SD
0.19	8.55	5.97	63.36	7.82

110	**Column name:**	**who_lab**
	Column description:	WHO/ISH CVD non-laboratory risk score
	Source information:	10-year risk of developing a fatal or non-fatal cardiovascular event. Function of age, sex, smoking status, systolic blood pressure, total cholesterol and diabetes mellitus status. Ref: Kaptoge S et Al. (2019). World Health Organization cardiovascular disease risk charts: Revised models to estimate risk in 21 global regions. The Lancet Global Health, 7(10), e1332–e1345. https://doi.org/10.1016/S2214-109X(19)30318-3. Calculated using sbp_mean2 variable for systolic blood pressure. Defined for age between 40 and 80 years.
	Data type:	Numeric
	Unique non-missing value count:	37,219
	Missing value count:	129,637

Min	Mean	Median	Max	SD
0.20	8.27	5.83	85.05	7.69

111	**Column name:**	**fhs_nonlab**
	Column description:	Framingham CVD non-laboratory risk score
	Source information:	10-year risk of developing a fatal or non-fatal cardiovascular event. Function of age, sex, smoking

	status, systolic blood pressure, BMI, diabetes status and use of antihypertensive medication. Ref. D'Agostino R et Al. General Cardiovascular Risk Profile for Use in Primary Care. Circulation 117(6);2008:743–53.. Calculated using sbp_mean2 variable for systolic blood pressure. Defined for age between 30 and 74 years.
Data type:	Numeric
Unique non-missing value count:	2,895
Missing value count:	117,028

Min	Mean	Median	Max	SD
1.00	11.53	8.16	30.00	9.53