Codebook

Harmonised ExPoSE South Africa Dataset

V. 1.0

Dataset name:	DATA_SA
Dataset size:	133.8 Mb
Column count:	210
Row count:	124,472
Updated date:	2024-09-12

Column Attributes:

1	**Column name:**	**country_ISO**
	Column description:	Country ISO code
	Data type:	Factor
2	**Column name:**	**country_name**
	Column description:	Country name
	Data type:	Factor
3	**Column name:**	**source**
	Column description:	Data source
	Data type:	Factor
4	**Column name:**	**year**
	Column description:	Year of data collection - Survey median
	Data type:	Numeric
5	**Column name:**	**hhid**
	Column description:	Household identifier
	Source information:	Generated in the harmonised dataset to be unique for each household.
	Data type:	Numeric

6	Column name:	pid
	Column description:	Individual identifier [Person ID (PID)]
	Source information:	Generated in the harmonised dataset to be unique for each individual. It takes into account the panel nature of the NIDS survey: i.e. individuals keep the same PID across waves.
	Data type:	Numeric

7	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

8	Column name:	stratum
	Column description:	Sampling stratum
	Source information:	Generated in the harmonised dataset to be unique.
	Data type:	Factor

9	Column name:	aweight_sa_int_base
	Column description:	Sampling weight: Interview (original)
	Source information:	Interview sampling weights as provided in the source datasets.
	Data type:	Numeric

10	Column name:	aweight_sa_phys_base
	Column description:	Sampling weight: Physical examination (original)
	Source information:	Physical examination sampling weights as provided in the source datasets.
	Data type:	Numeric

11	Column name:	aweight_sa_lab_base
	Column description:	Sampling weight: Laboratory (original)
	Source information:	Laboratory sampling weights as provided in the source datasets.
	Data type:	Numeric

12	Column name:	aweight_sa_int
	Column description:	Sampling weight: Interview
	Source information:	Interview sampling weights as provided in the source datasets. Recalibrated, rescaled to sample size.
	Data type:	Numeric

13	Column name:	aweight_sa_phys
	Column description:	Sampling weight: Physical examination
	Source information:	Physical examination sampling weights as provided in the source datasets. Recalibrated, rescaled to sample size.
	Data type:	Numeric

14	Column name:	aweight_sa_lab
	Column description:	Sampling weight: Laboratory
	Source information:	Laboratory sampling weights as provided in the source datasets. Recalibrated, rescaled to sample size.
	Data type:	Numeric

15	Column name:	geolevel1_name
	Column description:	Administrative level 1 - Name
	Source information:	Corresponds to the provincial administrative boundaries in 2001. Where 2001 province was not available and other smaller geographies were available (e.g., municipality or enumeration areas), these were linked to the 2001 Province using the 2001 Census Spatial Geography files from Statistics South Africa. For the DHS 2016, the geocoded centre of the enumeration areas was mapped onto the provincial boundaries.

	Data type:	Factor

16	Column name:	geolevel1_code
	Column description:	Administrative level 1 - Code
	Source information:	Corresponds to the provincial administrative boundaries in 2001. Where 2001 province was not available and other smaller geographies were available (e.g., municipality or enumeration areas), these were linked to the 2001 Province using the 2001 Census Spatial Geography files from Statistics South Africa. For the DHS 2016, the geocoded centre of the enumeration areas was mapped onto the provincial boundaries.
	Data type:	Factor

17	Column name:	geolevel2_name
	Column description:	Administrative level 2 - Name
	Source information:	Corresponds to the district administrative boundaries in 2001. Where 2001 district council was not available and other smaller geographies were available (e.g., municipality or enumeration areas), these were linked to the 2001 District Council using the 2001 Census Spatial Geography files from Statistics South Africa. For the DHS 2016, the geocoded centre of the enumeration areas was mapped onto the district boundaries.
	Data type:	Factor

18	Column name:	geolevel2_code
	Column description:	Administrative level 2 - Code
	Source information:	Corresponds to the district administrative boundaries in 2001. Where 2001 district council was not available and other smaller geographies were available (e.g., municipality or enumeration areas), these were linked to the 2001 District Council using the 2001 Census Spatial Geography files from Statistics South Africa. For the DHS 2016, the geocoded centre of the enumeration areas was mapped onto the district boundaries.
	Data type:	Factor

19	Column name:	geotype2
	Column description:	Urban/rural

	Data type:	Factor

20	Column name:	intm
	Column description:	Interview - Month
	Data type:	Factor

21	Column name:	inty
	Column description:	Interview - Year
	Data type:	Numeric

22	Column name:	vism
	Column description:	Anthropometry - Month
	Source information:	Month of anthropometric measurements.
	Data type:	Factor

23	Column name:	visq
	Column description:	Anthropometry - Quarter
	Source information:	Quarter of anthropometric measurements.
	Data type:	Factor

24	Column name:	hh_size
	Column description:	Household size
	Data type:	Numeric

25	Column name:	hh_size_cat
	Column description:	Household size, categorical
	Data type:	Factor

26	Column name:	hh_ownhome
	Column description:	Dwelling - Ownership

	Data type:	Factor

27	Column name:	hh_totrooms
	Column description:	Number of rooms in dwelling
	Data type:	Numeric

28	Column name:	hh_sleeprooms
	Column description:	Number of rooms for sleeping in dwelling
	Data type:	Numeric

29	Column name:	hh_ptotrooms
	Column description:	Number of rooms in dwelling, per household member
	Data type:	Numeric

30	Column name:	hh_psleeprooms
	Column description:	Number of rooms used for sleeping, per household member
	Data type:	Numeric

31	Column name:	hh_dwellingtype
	Column description:	Dwelling - Type
	Data type:	Factor

32	Column name:	hh_wallmaterial
	Column description:	Dwelling - Wall material
	Data type:	Factor

33	Column name:	hh_floormaterial
	Column description:	Dwelling - Floor material
	Source information:	Wood includes both finished/parquet and unfinished wood floors which were not separated in the NIDS. Caution should be used in interpretation.

	Data type:	Factor

34	**Column name:**	**hh_roofmaterial**
	Column description:	Dwelling - Wall material
	Data type:	Factor

35	**Column name:**	**hh_roof_wall_1**
	Column description:	Dwelling - roof/wall material: mud/thatching/wattle and daub
	Data type:	Factor

36	**Column name:**	**hh_roof_wall_2**
	Column description:	Dwelling - roof/wall material: mud and cement mix
	Data type:	Factor

37	**Column name:**	**hh_roof_wall_3**
	Column description:	Dwelling - roof/wall material: corrugated iron/zinc
	Data type:	Factor

38	**Column name:**	**hh_roof_wall_4**
	Column description:	Dwelling - roof/wall material: plastic/cardboard
	Data type:	Factor

39	**Column name:**	**hh_roof_wall_5**
	Column description:	Dwelling - roof/wall material: brick/cement/prefab/plaster
	Data type:	Factor

40	**Column name:**	**hh_roof_wall_9999**
	Column description:	Dwelling - roof/wall material: other
	Data type:	Factor

41	Column name:	hh_water
	Column description:	Source of drinking water
	Source information:	‘Borehole/well’ includes both protected and unprotected since some surveys (i.e., NIDS) made no distinction between the two. ‘Other’ includes water from a neighbour’s residence where this was separate in the original variable (e.g. NIDS)
	Data type:	Factor

42	Column name:	hh_toilet
	Column description:	Toilet type
	Source information:	In DHS 1998 the ventilated improved pit latrine responses appear to have been swapped with traditional latrine and therefore the data was inconsistent over time. This was corrected in the harmonised dataset.
	Data type:	Factor

43	Column name:	hh_sharedtoilet
	Column description:	Shared toilet
	Source information:	An indicator was made for shared toilet; however, this is only available for flush toilets in DHS 1998. For respondents whose households used other types of toilet facilities in DHS 1998, the indicator of whether it is a shared toilet or not was set to missing, since this was unknown. Caution therefore needs to be taken with these indicators.
	Data type:	Factor

44	Column name:	hh_refuseremoved
	Column description:	Refuse removal
	Data type:	Factor

45	Column name:	hh_cookingfuel
	Column description:	Main cooking fuel
	Source information:	This is unavailable for DHS 1998 where respondents were allowed to indicate multiple cooking fuels (See indicator variables below).

	Data type:	Factor

46	**Column name:**	**hh_heatingfuel**
	Column description:	Main heating fuel
	Data type:	Factor

47	**Column name:**	**hh_cook_elec**
	Column description:	Cooking fuel: Electricity
	Source information:	For survey years other than 1998, they were coded as either ‘yes’ (1) or missing since respondents were only allowed to select the main cooking fuel and not all cooking fuels used. For these surveys, the indicator was only coded ‘no’ (0) when it was reported that the household did not use a cooking fuel. Caution therefore needs to be taken when using these indicators since many of those with missing values are likely to be ‘no’ and not included in the numerator or denominator.
	Data type:	Factor

48	**Column name:**	**hh_cook_gas**
	Column description:	Cooking fuel: Gas
	Source information:	For survey years other than 1998, they were coded as either ‘yes’ (1) or missing since respondents were only allowed to select the main cooking fuel and not all cooking fuels used. For these surveys, the indicator was only coded ‘no’ (0) when it was reported that the household did not use a cooking fuel. Caution therefore needs to be taken when using these indicators since many of those with missing values are likely to be ‘no’ and not included in the numerator or denominator.
	Data type:	Factor

49	**Column name:**	**hh_cook_par**
	Column description:	Cooking fuel: Paraffin
	Source information:	For survey years other than 1998, they were coded as either ‘yes’ (1) or missing since respondents were only allowed to select the main cooking fuel and not all cooking fuels used. For these surveys, the indicator was only coded ‘no’ (0) when it was reported that the household did not use a cooking fuel. Caution therefore

		needs to be taken when using these indicators since many of those with missing values are likely to be 'no' and not included in the numerator or denominator.
	Data type:	Factor
50	**Column name:**	**hh_cook_wood**
	Column description:	Cooking fuel: Wood
	Source information:	For survey years other than 1998, they were coded as either 'yes' (1) or missing since respondents were only allowed to select the main cooking fuel and not all cooking fuels used. For these surveys, the indicator was only coded 'no' (0) when it was reported that the household did not use a cooking fuel. Caution therefore needs to be taken when using these indicators since many of those with missing values are likely to be 'no' and not included in the numerator or denominator.
	Data type:	Factor
51	**Column name:**	**hh_cook_coal**
	Column description:	Cooking fuel: Coal
	Source information:	For survey years other than 1998, they were coded as either 'yes' (1) or missing since respondents were only allowed to select the main cooking fuel and not all cooking fuels used. For these surveys, the indicator was only coded 'no' (0) when it was reported that the household did not use a cooking fuel. Caution therefore needs to be taken when using these indicators since many of those with missing values are likely to be 'no' and not included in the numerator or denominator.
	Data type:	Factor
52	**Column name:**	**hh_cook_dung**
	Column description:	Cooking fuel: Dung
	Source information:	For survey years other than 1998, they were coded as either 'yes' (1) or missing since respondents were only allowed to select the main cooking fuel and not all cooking fuels used. For these surveys, the indicator was only coded 'no' (0) when it was reported that the household did not use a cooking fuel. Caution therefore needs to be taken when using these indicators since many of those with missing values are likely to be 'no' and not included in the numerator or denominator.

	Data type:	Factor

53	Column name:	hh_cook_other
	Column description:	Cooking fuel: Other
	Source information:	For survey years other than 1998, they were coded as either ‘yes’ (1) or missing since respondents were only allowed to select the main cooking fuel and not all cooking fuels used. For these surveys, the indicator was only coded ‘no’ (0) when it was reported that the household did not use a cooking fuel. Caution therefore needs to be taken when using these indicators since many of those with missing values are likely to be ‘no’ and not included in the numerator or denominator.
	Data type:	Factor

54	Column name:	hh_recgrant
	Column description:	Household member receives government grant
	Source information:	Government grants include old age pensions, child maintenance/child support grants, disability grants, foster child grants, care dependency grants, war veteran, social relief of distress or any other kind of social grant. In the NIDS this refers to whether any household member received a grant in the last month. In the DHS, a time frame is not specified. Unavailable in SAGE.
	Data type:	Factor

55	Column name:	hh_govsupport
	Column description:	Household receives government support
	Source information:	Available only in SAGE 2007. Refers to financial or in-kind support from the government (excludes pensions which are considered elsewhere).
	Data type:	Factor

56	Column name:	hh_foodinsec
	Column description:	Food insecurity
	Source information:	NIDS: If the response was ‘never’ for both children and adults in the household for how often a household member ran out of food/went hungry in the past 30 days, it was coded as ‘no’. If responses were missing or

		not applicable for both questions, it was coded as missing.
	Data type:	Factor
57	**Column name:**	**hh_foodinsec_adult**
	Column description:	Food insecurity: adult
	Source information:	NIDS: If the response was 'never' for adults in the household, it was coded as 'no' (0). It was coded as 'yes' (1) if it was any other response (rarely, sometimes, always, etc.) and missing if there was no response. SAGE: Food insecurity questions were phrased as whether the individual ever ate less than they felt they should because there was not enough food? This was in reference to the last 30 days. Because there was no household level variable, this was collapsed across household members to be 'yes' (1) if any adults interviewed in the household responded yes, 'no' if none of them did, and missing if some responses were missing but none were yes. This is likely to be an underestimate since not all adults in the household were interviewed.
	Data type:	Factor
58	**Column name:**	**hh_foodinsec_child**
	Column description:	Food insecurity: child
	Source information:	NIDS: If the response was 'never' for children in the household, it was coded as 'no' (0). It was coded as 'yes' (1) if it was any other response (rarely, sometimes, always, etc.) and missing if there was no response.
	Data type:	Factor
59	**Column name:**	**hh_ass_elec**
	Column description:	Household assets: Electricity
	Data type:	Factor
60	**Column name:**	**hh_ass_radio**
	Column description:	Household assets: Radio
	Data type:	Factor

61	**Column name:**	**hh_ass_tv**
	Column description:	Household assets: TV
	Data type:	Factor

62	**Column name:**	**hh_ass_fridge**
	Column description:	Household assets: Fridge
	Data type:	Factor

63	**Column name:**	**hh_ass_bicycle**
	Column description:	Household assets: Bicycle
	Data type:	Factor

64	**Column name:**	**hh_ass_motorcycle**
	Column description:	Household assets: Motorcycle
	Data type:	Factor

65	**Column name:**	**hh_ass_car_truck**
	Column description:	Household assets: Car/Truck
	Data type:	Factor

66	**Column name:**	**hh_ass_phone**
	Column description:	Household assets: Phone (landline)
	Source information:	For landline telephone, this was limited to those with a working telephone where available (e.g., NIDS, DHS) but some surveys (e.g., SAGE) did not specify whether the landline was functioning or not.
	Data type:	Factor

67	**Column name:**	**hh_ass_computer**
	Column description:	Household assets: Computer
	Data type:	Factor

68	**Column name:**	**hh_ass_wmachine**
	Column description:	Household assets: Washing Machine
	Data type:	Factor
69	**Column name:**	**hh_ass_cellphone**
	Column description:	Household assets: Electricity
	Data type:	Factor
70	**Column name:**	**hh_ass_watch**
	Column description:	Household assets: Phone (cellular)
	Data type:	Factor
71	**Column name:**	**hh_ass_animalcart**
	Column description:	Household assets: Animal cart
	Data type:	Factor
72	**Column name:**	**hh_ass_motorboat**
	Column description:	Household assets: Motorboat
	Data type:	Factor
73	**Column name:**	**hh_ass_vacuum**
	Column description:	Household assets: Vacuum cleaner
	Data type:	Factor
74	**Column name:**	**hh_ass_microwave**
	Column description:	Household assets: Microwave oven
	Data type:	Factor
75	**Column name:**	**hh_ass_stove**
	Column description:	Household assets: Stove

	Data type:	Factor

76	**Column name:**	**hh_ass_sat**
	Column description:	Household assets: Satellite TV
	Data type:	Factor

77	**Column name:**	**hh_ass_video**
	Column description:	Household assets: Video player
	Data type:	Factor

78	**Column name:**	**hh_ass_hifi**
	Column description:	Household assets: hifi
	Data type:	Factor

79	**Column name:**	**hh_ass_camera**
	Column description:	Household assets: Photo camera
	Data type:	Factor

80	**Column name:**	**hh_ass_smachine**
	Column description:	Household assets: Sewing machine
	Data type:	Factor

81	**Column name:**	**hh_ass_sofa**
	Column description:	Household assets: Sofa
	Data type:	Factor

82	**Column name:**	**hh_ass_boat**
	Column description:	Household assets: Boat
	Data type:	Factor

83	**Column name:**	**hh_ass_plough**
	Column description:	Household assets: Plough
	Data type:	Factor
84	**Column name:**	**hh_ass_tractor**
	Column description:	Household assets: Tractor
	Data type:	Factor
85	**Column name:**	**hh_ass_wheelbarrow**
	Column description:	Household assets: Wheelbarrow
	Data type:	Factor
86	**Column name:**	**hh_ass_mill**
	Column description:	Household assets: Mill
	Data type:	Factor
87	**Column name:**	**hh_ass_tab**
	Column description:	Household assets: Table
	Data type:	Factor
88	**Column name:**	**hh_ass_sink**
	Column description:	Household assets: Sink
	Data type:	Factor
89	**Column name:**	**hh_ass_hotw**
	Column description:	Household assets: Hot water
	Data type:	Factor
90	**Column name:**	**hh_ass_dishwasher**
	Column description:	Household assets: Dishwasher

	Data type:	Factor

91	Column name:	hh_edu_deprived
	Column description:	Deprivation: education (Anchor variable 1 for the Comparative Wealth Index)
	Source information:	Indicator for deprivation in education. ‘Yes’ indicates that no household member has completed primary school.
	Data type:	Factor

92	Column name:	hh_unimp_toilet
	Column description:	Deprivation: sanitation (Anchor variable 2 for the Comparative Wealth Index)
	Source information:	Indicator for deprivation in sanitation. ‘Yes’ indicates unimproved sanitation facilities. Flush toilet/chemical toilet and ventilated improved pit latrine were considered improved while traditional pit latrine/bucket, no facility/bush and other toilet facilities were considered unimproved. Household toilet facilities that were shared with other households were considered unimproved. [Note: For DHS 1998, information on sharing of facilities was only available for those with flush toilets. There is a possibility that some of those with ventilated improved pit latrines were classified as having improved facilities where they would have been classified as unimproved if it was shared]
	Data type:	Factor

93	Column name:	hh_unimp_water
	Column description:	Deprivation: water (Anchor variable 3 for the Comparative Wealth Index)
	Source information:	Indicator for deprivation in water source. ‘Yes’ indicates unimproved water source. Piped/tap water, boreholes/wells and rainwater tanks were considered to be improved. Water carrier/tanker, surface water (stream, dam, pool or spring) or other drinking water source were considered to be unimproved. All surface water sources were considered to be unimproved because some surveys did not separate types of surface water sources.
	Data type:	Factor

94	Column name:	hh_unimp_cooking
	Column description:	Deprivation: cooking fuel (Anchor variable 4 for the Comparative Wealth Index)
	Source information:	Indicator for deprivation in cooking fuel. 'Yes' indicates unimproved/solid cooking fuels. Unimproved/solid cooking fuels included coal, wood/straw, dung or other/none. Improved cooking fuels included electricity, gas and paraffin.
	Data type:	Factor

95	Column name:	hh_dep1plus
	Column description:	Deprivation: 1+ indicator
	Source information:	Deprived in one or more of the Comparative Wealth Index anchor variables.
	Data type:	Factor

96	Column name:	hh_dep2plus
	Column description:	Deprivation: 2+ indicators
	Source information:	Deprived in two or more of the Comparative Wealth Index anchor variables.
	Data type:	Factor

97	Column name:	hh_dep3plus
	Column description:	Deprivation: 3+ indicators
	Source information:	Deprived in three or more of the Comparative Wealth Index anchor variables.
	Data type:	Factor

98	Column name:	hh_dep4plus
	Column description:	Deprivation: 4+ indicators
	Source information:	Deprived in four or more of the Comparative Wealth Index anchor variables.
	Data type:	Factor

99	Column name:	hh_income
	Column description:	Household income
	Source information:	In Rands. Only available in NIDS. Values with full imputations (see NIDS documentation). Values are not adjusted for inflation.
	Data type:	Numeric

100	Column name:	hh_income_quint
	Column description:	Household income quintile
	Data type:	Factor

101	Column name:	hh_windex
	Column description:	Household wealth index
	Source information:	The household wealth index score was derived by principal component analysis (keeping only the first component), separately in each dataset. The score was constructed from the indicators available in each survey, using the DHS 2016 wealth index creation as a guide. The original classifications in each survey were used to create indicator variables for the principal components analysis. Indicators included: Dwelling ownership & dwelling type, Durable asset ownership (TV, washing machine, stove, car, etc.), Toilet facilities, Source of drinking water, Main material for roof/walls/floors of the dwelling, Number of household members per number of rooms (sleeping rooms where available. NIDS only has total number of rooms and not sleeping rooms), Type of refuse disposal, Type of cooking fuel. This wealth index is a relative measure that is specific to each survey and not comparable across surveys or time. Within each source dataset, higher values of the wealth index correspond to higher socioeconomic status.
	Data type:	Numeric

102	Column name:	hh_windex_quint
	Column description:	Household wealth index quintile
	Source information:	The wealth quintiles are a relative measure (e.g. ranking) that are specific to each survey.
	Data type:	Factor

103	Column name:	hh_cwi
	Column description:	Comparative wealth index
	Source information:	The comparative wealth index was constructed using the method described in: Rutstein SO, Staveteig S. Making the Demographic and Health Surveys Wealth Index Comparable. DHS Methodological Reports 9. 2014. Anchoring points to distinguish those at lower levels of socioeconomic status were modified from the Multidimensional Poverty Index. These included the four following indicators available in all surveys: (1) No household member has completed primary school, (2) The household uses unimproved sanitation facilities (traditional pit latrine/bucket, no facility/bush or shared facilities), (3) The household uses unimproved drinking water (water carrier/tanker, surface water, other), (4) The household cooks with solid cooking fuel. Anchoring points used to distinguish those at higher levels of socioeconomic status included possession of the following four assets: Television, Refrigerator, Car/truck, Computer. Wealth indices were calibrated using the 2008 NIDS survey as a reference. Within and across source datasets, higher values of the comparative wealth index correspond to higher socioeconomic status.
	Data type:	Numeric

104	Column name:	hh_deaths12mo
	Column description:	Death in the household last 12 months
	Source information:	NIDS asks for household deaths in past 24 months. Those with a missing month of death or interview but where the calendar year of death was the year before the interview year are coded as having a death in the past year (only one of such observation in Wave 5). Those who reported a household death in the past 24 months but who did not provide the year of death were coded as missing. SAGE data not publicly available for date of death, so not available in SAGE.
	Data type:	Factor

105	Column name:	sex
	Column description:	Sex
	Data type:	Factor

106	Column name:	age
	Column description:	Age
	Data type:	Numeric

107	Column name:	agecat1
	Column description:	Age category (5 years)
	Data type:	Factor

108	Column name:	agecat2
	Column description:	Age category (10 years)
	Data type:	Factor

109	Column name:	race
	Column description:	Population group
	Source information:	The four categories are those currently used in the official statistics and match the socially defined 'racial' groups used under the apartheid regime. See reference in: Statistics South Africa, Standards Division. Concepts and Definitions for Statistics South Africa 2010. v 3. Pretoria: Statistics South Africa. 2010:72.
	Data type:	Factor

110	Column name:	race_imp
	Column description:	Population group, imputed
	Source information:	Where race was missing and another household member had a valid race, this race was used to impute the household member with missing race. This mostly affected SAGE 2007 data where approximately 300 individuals were imputed using this method. The imputed race variable was used in the recalibration of weights to the population totals.
	Data type:	Factor

111	Column name:	marstatus
	Column description:	Marital status

	Data type:	Factor

112	Column name:	edu1
	Column description:	Education: categorisation 1
	Source information:	Highest level of education. Education categories were defined as in the 2011 South Africa Census: Stats SA. 2014. Census 2011 10% Sample Metadata. Pretoria: Stats SA, pp.60-67. SANHANES: NTC Level 2-4 were grouped. NTC Level 2 & 3 (NTC I-II) is considered some secondary, while NTC 3 (Level 4) is completed secondary. Because they could not be separated in SANHANES, we classified NTC 2-4 as 'some secondary'.
	Data type:	Factor

113	Column name:	edu2
	Column description:	Education: categorisation 2
	Source information:	SAGE 2007-8 and 2014 classify those with some higher education but without a degree as 'Completed secondary'. For this reason, an education variable with 5 categories was constructed, combining codes 4 and 5 from the variable edu1.
	Data type:	Factor

114	Column name:	emp
	Column description:	Employment
	Source information:	Unemployed includes those who are not economically active where the survey has a separate category for not economically active.
	Data type:	Factor

115	Column name:	smokstatus
	Column description:	Smoking status
	Source information:	Where survey asks about cigarettes only (e.g. NIDS), based on cigarette smoking. SAGE does not distinguish between smoking and smokeless tobacco, so it includes all tobacco users. Current smokers include those who smoke, regardless of whether they smoke daily, where the survey makes a distinction between daily and non-

		daily smoking. DHS 1998 and 2003: Former smokers include those who have ever smoked daily but currently do not smoke at all. In DHS 2016, former smokers include those who ever smoked every day or some days but do not currently smoke at all. NIDS: Former smokers are those who reported not currently smoking but have ever smoked 'regularly'. SAGE: Former smokers include those who have ever smoked (or used smokeless tobacco) but do not currently use tobacco products at all. SANHANES: former smokers include those who have ever smoked daily or less than daily but do not currently smoke at all.
	Data type:	Factor
116	**Column name:**	**currsmok**
	Column description:	Current smoker
	Source information:	Non-smokers include former smokers. See notes in variable smokingstatus for details on definitions.
	Data type:	Factor
117	**Column name:**	**alcstatus**
	Column description:	Alcohol status
	Source information:	Never drinkers are those reporting having never drunk alcohol/a drink containing alcohol. Current drinkers are those reporting: drinking alcohol now (DHS 1998); having consumed a drink containing alcohol in the last 12 months (DHS 2003 & 2016); having ever consumed an alcoholic beverage with a frequency of having at least one alcoholic drink at greater than 0 days in the past 12 months (SAGE); a frequency of drinking alcohol that's greater than 'never' or 'no longer' (NIDS); Former drinkers are those who reported: ever consuming an alcoholic drink and a frequency of having at least one alcoholic drink in the past 12 months of 'No days' (SAGE); no longer drinking alcohol (NIDS); ever drinking alcohol but not drinking now (DHS 1998); ever drinking a beverage containing alcohol but not within the past 12 months (DHS 2003 & 2016).
	Data type:	Factor
118	**Column name:**	**Curralc**
	Column description:	Current drinker

	Source information:	Non-drinkers include former drinkers. See note above for further details. For SANHANES, current drinkers are those reporting a frequency of having a drink in the past 12 months greater than ‘never’.
	Data type:	Factor
119	**Column name:**	**alcavg**
	Column description:	Average alcohol consumption [g/day]
	Source information:	Estimated by multiplying the midpoint of each category for frequency of drinking (e.g. days per week, month, etc) by the midpoint of the number of drinks per occasion of drinking and dividing by the number of days in the interval. For the highest category of drinks per occasion (e.g., n or more), the value was estimated as 1.5 times the endpoint of the previous category following Probst, Shuper & Rehm (2016) in Probst, C., Shuper, P. A., & Rehm, J. (2017). Coverage of alcohol consumption by national surveys in South Africa. Addiction, 112(4), 705–710. https://doi.org/10.1111/add.13692. A standard drink was assumed to be 12 grams (Probst, Shuper & Rehm, 2016). NIDS: Frequency was reported in days of drinking per week. Those reporting drinking ‘very rarely’ were estimated as drinking 0.25 days a week, and ‘less than once a week’ was estimated as 0.5 days a week to generate drinks per week and convert to drinks and grams per day. All other categories used the midpoint. SAGE and SANHANES: Drinking frequency over the past 12 months was reported. Those reporting drinking ‘less than once per month’ or ‘monthly or less’ were estimated as drinking 6 days per year. All other categories used the midpoint for frequency to estimate drinking days per year, multiplying by 12 if it was a monthly frequency or 52 if it was a weekly frequency, then dividing by 365 to estimate daily drinking quantity. (For the last category of n days per week, the midpoint between n and 7 was used.). DHS 1998: The average number of drinks per day were estimated using the midpoint of each category or 1.5 times the endpoint of the highest category for weekdays and weekend days and summed across days to generate weekly consumption. DHS 2003 & 2016: The number of standard drinks reported each day of the week during the past 7 days were summed to estimate weekly consumption and converted to daily consumption.
	Data type:	Numeric
120	**Column name:**	**Gpaq**

	Column description:	GPAQ
	Source information:	Calculated using the Global Physical Activity Questionnaire (GPAQ): World Health Organisation. Global physical activity questionnaire (GPAQ). Available at: https://www.who.int/publications/m/item/global-physical-activity-questionnaire
	Data type:	Numeric
121	**Column name:**	**gpaqcat**
	Column description:	GPAQ category
	Source information:	Calculated using the Global Physical Activity Questionnaire (GPAQ): World Health Organisation. Global physical activity questionnaire (GPAQ). Available at: https://www.who.int/publications/m/item/global-physical-activity-questionnaire
	Data type:	Factor
122	**Column name:**	**exercisefreq**
	Column description:	Exercise frequency
	Source information:	Only considers leisure time physical activity.
	Data type:	Factor
123	**Column name:**	**self_health**
	Column description:	Self-rated health
	Source information:	Different Likert scales were used in different surveys: In NIDS, this variable was originally coded as: poor, fair, good, very good, excellent; In SAGE, this variable was originally coded as: very bad, bad, average, good, very good; In DHS, this variable was originally coded as: poor, average, good, very good; In the SANHANES, the variable was originally coded as very bad, bad, moderate, good, very good. Care should be taken in interpretation of this variable.
	Data type:	Factor
124	**Column name:**	**diag_hbp**

	Column description:	Diagnosis: Hypertension
	Data type:	Factor
125	**Column name:**	**diag_isch**
	Column description:	Diagnosis: Heart attack/angina
	Source information:	SANHANES, DHS and SAGE specify heart attack or angina which were included in the variable diag_isch.
	Data type:	Factor
126	**Column name:**	**diag_stroke**
	Column description:	Diagnosis: Stroke
	Data type:	Factor
127	**Column name:**	**diag_chol**
	Column description:	Diagnosis: Hypercholesterolaemia
	Data type:	Factor
]		
128	**Column name:**	**diag_diab**
	Column description:	Diagnosis: Diabetes/hyperglycaemia
	Data type:	Factor
129	**Column name:**	**diag_emph**
	Column description:	Diagnosis: Emphysema/Chronic bronchitis
	Source information:	Where the survey specifies chronic lung disease diagnosis separately from asthma (e.g., SAGE), or emphysema/bronchitis diagnosis (e.g., DHS), these were included in the variable diag_emph. Where the survey only specifies asthma diagnosis (e.g., NIDS), this was included in the variable diag_asth.
	Data type:	Factor
130	**Column name:**	**diag_asth**
	Column description:	Diagnosis: Asthma

	Source information:	Where the survey specifies chronic lung disease diagnosis separately from asthma (e.g., SAGE), or emphysema/bronchitis diagnosis (e.g., DHS), these were included in the variable diag_emph. Where the survey only specifies asthma diagnosis (e.g., NIDS), this was included in the variable diag_asth.
	Data type:	Factor
131	**Column name:**	**diag_tb**
	Column description:	Diagnosis: Tuberculosis
	Data type:	Factor
132	**Column name:**	**diag_cancer**
	Column description:	Diagnosis: Cancer
	Data type:	Factor
133	**Column name:**	**diag_heart**
	Column description:	Diagnosis: Heart problems
	Source information:	NIDS does not specify specific heart conditions such as heart attack or angina but has a generic heart problems variable which was recoded as diag_heart.
	Data type:	Factor
134	**Column name:**	**bpmed**
	Column description:	Medication: Hypertension
	Source information:	NIDS asks ‘Are you currently taking medication for this condition? SAGE is based on responses to, ‘Have you been taking any medication for it in the last two weeks? DHS 1998 & 2003 ask about current use. DHS 2016 only asks ‘Did you receive treatment at the time of diagnosis?’ for the condition, so this has been excluded.
	Data type:	Factor
135	**Column name:**	**diabmed**
	Column description:	Medication: Diabetes/hyperglycaemia

	Source information:	NIDS asks ‘Are you currently taking medication for this condition? SAGE is based on responses to, ‘Have you been taking any medication for it in the last two weeks? DHS 1998 & 2003 ask about current use. DHS 2016 only asks ‘Did you receive treatment at the time of diagnosis?’ for the condition, so this has been excluded.
	Data type:	Factor
136	**Column name:**	**cholmed**
	Column description:	Medication: Hypercholesterolaemia
	Source information:	NIDS asks ‘Are you currently taking medication for this condition? SAGE is based on responses to, ‘Have you been taking any medication for it in the last two weeks? DHS 1998 & 2003 ask about current use. DHS 2016 only asks ‘Did you receive treatment at the time of diagnosis?’ for the condition, so this has been excluded.
	Data type:	Factor
137	**Column name:**	**ischmed**
	Column description:	Medication: Heart attack/angina
	Source information:	NIDS asks ‘Are you currently taking medication for this condition? SAGE is based on responses to, ‘Have you been taking any medication for it in the last two weeks? DHS 1998 & 2003 ask about current use. DHS 2016 only asks ‘Did you receive treatment at the time of diagnosis?’ for the condition, so this has been excluded.
	Data type:	Factor
138	**Column name:**	**lungmed**
	Column description:	Medication: Respiratory problems
	Source information:	NIDS asks ‘Are you currently taking medication for this condition? SAGE is based on responses to, ‘Have you been taking any medication for it in the last two weeks? DHS 1998 & 2003 ask about current use. DHS 2016 only asks ‘Did you receive treatment at the time of diagnosis?’ for the condition, so this has been excluded.
	Data type:	Factor
139	**Column name:**	**tbmed**

	Column description:	Medication: Tuberculosis
	Source information:	NIDS asks ‘Are you currently taking medication for this condition? SAGE is based on responses to, ‘Have you been taking any medication for it in the last two weeks? DHS 1998 & 2003 ask about current use. DHS 2016 only asks ‘Did you receive treatment at the time of diagnosis?’ for the condition, so this has been excluded.
	Data type:	Factor
140	**Column name:**	**strokemed**
	Column description:	Medication: Stroke
	Source information:	NIDS asks ‘Are you currently taking medication for this condition? SAGE is based on responses to, ‘Have you been taking any medication for it in the last two weeks? DHS 1998 & 2003 ask about current use. DHS 2016 only asks ‘Did you receive treatment at the time of diagnosis?’ for the condition, so this has been excluded.
	Data type:	Factor
141	**Column name:**	**bpmed_coded**
	Column description:	Medication: Hypertension, coded
	Source information:	Directly coded from the medication names and details listed on medication containers presented by respondents. Only available for the DHS surveys.
	Data type:	Factor
142	**Column name:**	**diabmed_coded**
	Column description:	Medication: Diabetes/hyperglycaemia, coded
	Source information:	Directly coded from the medication names and details listed on medication containers presented by respondents. Only available for the DHS surveys.
	Data type:	Factor
143	**Column name:**	**cholmed_coded**
	Column description:	Medication: Hypercholesterolaemia, coded

	Source information:	Directly coded from the medication names and details listed on medication containers presented by respondents. Only available for the DHS surveys.
	Data type:	Factor

144	**Column name:**	**ischmed_coded**
	Column description:	Medication: Heart attack/angina, coded
	Source information:	Directly coded from the medication names and details listed on medication containers presented by respondents. Only available for the DHS surveys.
	Data type:	Factor

145	**Column name:**	**lungmed_coded**
	Column description:	Medication: Respiratory problems, coded
	Source information:	Directly coded from the medication names and details listed on medication containers presented by respondents. Only available for the DHS surveys.
	Data type:	Factor

146	**Column name:**	**tbmed_coded**
	Column description:	Medication: Tuberculosis, coded
	Source information:	Directly coded from the medication names and details listed on medication containers presented by respondents. Only available for the DHS surveys.
	Data type:	Factor

147	**Column name:**	**strokemed_coded**
	Column description:	Medication: Stroke, coded
	Source information:	Directly coded from the medication names and details listed on medication containers presented by respondents. Only available for the DHS surveys.
	Data type:	Factor

148	**Column name:**	**parity**
	Column description:	Parity

	Source information:	A few respondents listed as male in NIDS reported a number of children they had given birth to. These were coded as missing.
	Data type:	Numeric
149	**Column name:**	**currpreg**
	Column description:	Currently pregnant
	Source information:	Pregnancy status was set to 0 (No/don’t know) for females over the age of 49 to be consistent with the DHS where this question was not asked of women over childbearing age (50 and over).
	Data type:	Factor
150	**Column name:**	**everpreg**
	Column description:	Ever pregnant
	Source information:	In NIDS, the question is phrased as 'Have you ever given birth?' In SANHANES, the question is phrased as ‘Ever been pregnant?’. The question is not available in the DHS adult health file or in the SAGE. (A few respondents who were listed as male in NIDS reported having ever given birth (yes or no)).
	Data type:	Factor
151	**Column name:**	**height1**
	Column description:	Height [cm] - First reading
	Source information:	Implausible values removed from the dataset: Height<120 cm or height > 220 cm.
	Data type:	Numeric
152	**Column name:**	**height2**
	Column description:	Height [cm] - Second reading
	Source information:	Implausible values removed from the dataset: Height<120 cm or height > 220 cm.
	Data type:	Numeric
153	**Column name:**	**height3**

	Column description:	Height [cm] - Third reading
	Source information:	Implausible values removed from the dataset: Height<120 cm or height > 220 cm.
	Data type:	Numeric
154	**Column name:**	**height**
	Column description:	Height [cm] - Average of available readings
	Source information:	Arithmetic mean of the available plausible readings where more than one reading was taken.
	Data type:	Numeric
155	**Column name:**	**weight1**
	Column description:	Weight [Kg] - First reading
	Source information:	Implausible values removed from the dataset: Females: Weight<25 Kg; Weight > 250 Kg; Males: Weight <35 Kg; Weight > 250 Kg. Ref: 2nd South African Comparative Risk Assessment (SACRA) study.
	Data type:	Numeric
156	**Column name:**	**weight2**
	Column description:	Weight [Kg] - Second reading
	Source information:	Implausible values removed from the dataset: Females: Weight<25 Kg; Weight > 250 Kg; Males: Weight <35 Kg; Weight > 250 Kg. Ref: 2nd South African Comparative Risk Assessment (SACRA) study.
	Data type:	Numeric
157	**Column name:**	**weight3**
	Column description:	Weight [Kg] - Third reading
	Source information:	Implausible values removed from the dataset: Females: Weight<25 Kg; Weight > 250 Kg; Males: Weight <35 Kg; Weight > 250 Kg. Ref: 2nd South African Comparative Risk Assessment (SACRA) study.
	Data type:	Numeric

158	Column name:	weight
	Column description:	Weight [Kg] - Average of available readings
	Source information:	Arithmetic mean of the available plausible readings where more than one reading was taken.
	Data type:	Numeric

159	Column name:	waist1
	Column description:	Waist circumference [cm] - First reading
	Source information:	Implausible values removed from the dataset: waist < 30 cm or waist > 220 cm.
	Data type:	Numeric

160	Column name:	waist2
	Column description:	Waist circumference [cm] - Second reading
	Source information:	Implausible values removed from the dataset: waist < 30 cm or waist > 220 cm.
	Data type:	Numeric

161	Column name:	waist3
	Column description:	Waist circumference [cm] - Third reading
	Source information:	Implausible values removed from the dataset: waist < 30 cm or waist > 220 cm.
	Data type:	Numeric

162	Column name:	Waist
	Column description:	Waist circumference [cm] - Average of available readings
	Source information:	Arithmetic mean of the available plausible readings where more than one reading was taken.
	Data type:	Numeric

163	Column name:	arm1
	Column description:	Arm circumference [cm] - First reading

	Source information:	Implausible values removed from the dataset (Based on the measurement range of the standard measuring tape according to the UNICEF product Specification Sheet V.4, 2020. Available from: https://www.unicef.org/supply/media/3996/file/MUAC-tape-adult-specifications-May2020.pdf): Arm < 6 cm, Arm > 50 cm
	Data type:	Numeric
164	**Column name:**	**arm2**
	Column description:	Arm circumference [cm] - Second reading
	Source information:	Implausible values removed from the dataset (Based on the measurement range of the standard measuring tape according to the UNICEF product Specification Sheet V.4, 2020. Available from: https://www.unicef.org/supply/media/3996/file/MUAC-tape-adult-specifications-May2020.pdf): Arm < 6 cm, Arm > 50 cm
	Data type:	Numeric
165	**Column name:**	**arm3**
	Column description:	Arm circumference [cm] - Third reading
	Source information:	Implausible values removed from the dataset (Based on the measurement range of the standard measuring tape according to the UNICEF product Specification Sheet V.4, 2020. Available from: https://www.unicef.org/supply/media/3996/file/MUAC-tape-adult-specifications-May2020.pdf): Arm < 6 cm, Arm > 50 cm
	Data type:	Numeric
166	**Column name:**	**Arm**
	Column description:	Arm circumference [cm] - Average of available readings
	Source information:	Arithmetic mean of the available plausible readings where more than one reading was taken.
	Data type:	Numeric
167	**Column name:**	**hip1**

	Column description:	Hip circumference [cm] - First reading
	Source information:	Implausible values removed from the dataset (Based on the distribution in the sample): Hip circumference < 40 cm or hip circumference > 230.
	Data type:	Numeric
168	**Column name:**	**hip2**
	Column description:	Hip circumference [cm] - Second reading
	Source information:	Implausible values removed from the dataset (Based on the distribution in the sample): Hip circumference < 40 cm or hip circumference > 230.
	Data type:	Numeric
169	**Column name:**	**hip3**
	Column description:	Hip circumference [cm] - Third reading
	Source information:	Implausible values removed from the dataset (Based on the distribution in the sample): Hip circumference < 40 cm or hip circumference > 230.
	Data type:	Numeric
170	**Column name:**	**hip**
	Column description:	Hip circumference [cm] - Average of available readings
	Source information:	Arithmetic mean of the available plausible readings where more than one reading was taken.
	Data type:	Numeric
171	**Column name:**	**sbp1**
	Column description:	Systolic Blood Pressure [mmHg] - First reading
	Source information:	Implausible values removed from the dataset: SBP < 60 mmHg or SBP > 270 mmHg. SBP readings were also set to missing if they were less than 15 mmHg greater than the corresponding DBP reading.
	Data type:	Numeric
172	**Column name:**	**sbp2**

	Column description:	Systolic Blood Pressure [mmHg] - Second reading
	Source information:	Implausible values removed from the dataset: SBP < 60 mmHg or SBP > 270 mmHg. SBP readings were also set to missing if they were less than 15 mmHg greater than the corresponding DBP reading.
	Data type:	Numeric
173	**Column name:**	**sbp3**
	Column description:	Systolic Blood Pressure [mmHg] - Third reading
	Source information:	Implausible values removed from the dataset: SBP < 60 mmHg or SBP > 270 mmHg. SBP readings were also set to missing if they were less than 15 mmHg greater than the corresponding DBP reading.
	Data type:	Numeric
174	**Column name:**	**sbp_mean1**
	Column description:	Systolic Blood Pressure [mmHg] - Average of available readings
	Source information:	Arithmetic mean of all plausible available readings.
	Data type:	Numeric
175	**Column name:**	**sbp_mean2**
	Column description:	Systolic Blood Pressure [mmHg] - Average of available readings excluding the first
	Source information:	Arithmetic mean of all plausible available readings excluding the first.
	Data type:	Numeric
176	**Column name:**	**dbp1**
	Column description:	Diastolic Blood Pressure [mmHg] - First reading
	Source information:	Implausible values removed from the dataset (From the 2nd South African Comparative Risk Assessment (SACRA) study): DBP < 30 mmHg; DBP > 150 mmHg. DBP readings were also set to missing if they were less than 15 mmHg lower than the corresponding SBP reading.

	Data type:	Numeric

177	**Column name:**	**dbp2**
	Column description:	Diastolic Blood Pressure [mmHg] - Second reading
	Source information:	Implausible values removed from the dataset (From the 2nd South African Comparative Risk Assessment (SACRA) study): DBP < 30 mmHg; DBP > 150 mmHg. DBP readings were also set to missing if they were less than 15 mmHg lower than the corresponding SBP reading.
	Data type:	Numeric

178	**Column name:**	**dbp3**
	Column description:	Diastolic Blood Pressure [mmHg] - Third reading
	Source information:	Implausible values removed from the dataset (From the 2nd South African Comparative Risk Assessment (SACRA) study): DBP < 30 mmHg; DBP > 150 mmHg. DBP readings were also set to missing if they were less than 15 mmHg lower than the corresponding SBP reading.
	Data type:	Numeric

179	**Column name:**	**dbp_mean1**
	Column description:	Diastolic Blood Pressure [mmHg] - Average of available readings
	Source information:	Arithmetic mean of all plausible available readings.
	Data type:	Numeric

180	**Column name:**	**dbp_mean2**
	Column description:	Diastolic Blood Pressure [mmHg] - Average of available readings excluding the first
	Source information:	Arithmetic mean of all plausible available readings excluding the first.
	Data type:	Numeric

181	**Column name:**	**rhr1**

	Column description:	Resting Heart Rate [ppm] - First reading
	Source information:	Biologically implausible values removed from the dataset: RHR < 20 bpm; RHR > 250 bpm. (Informed by clinical opinion).
	Data type:	Numeric
182	**Column name:**	**rhr2**
	Column description:	Resting Heart Rate [ppm] - Second reading
	Source information:	Biologically implausible values removed from the dataset: RHR < 20 bpm; RHR > 250 bpm. (Informed by clinical opinion).
	Data type:	Numeric
183	**Column name:**	**rhr3**
	Column description:	Resting Heart Rate [ppm] - Third reading
	Source information:	Biologically implausible values removed from the dataset: RHR < 20 bpm; RHR > 250 bpm. (Informed by clinical opinion).
	Data type:	Numeric
184	**Column name:**	**rhr_mean1**
	Column description:	Resting Heart Rate [ppm] - Average of available readings
	Source information:	Arithmetic mean of all plausible available readings.
	Data type:	Numeric
185	**Column name:**	**rhr_mean2**
	Column description:	Resting Heart Rate [ppm] - Average of available readings excluding the first
	Source information:	Arithmetic mean of all plausible available readings excluding the first.
	Data type:	Numeric
186	**Column name:**	**bmi**

	Column description:	BMI [kg/m^2]
	Source information:	Calculated using the mean 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
187	**Column name:**	**Bmicat**
	Column description:	BMI category
	Source information:	Underweight = BMI <18.5 kg/m^2. Healthy weight = 18.5 kg/m^2 < 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/m2 < BMI < 40 kg/m^2. Obesity III = 40 kg/m2 < BMI
	Data type:	Factor
188	**Column name:**	**hb**
	Column description:	Haemoglobin [g/dl]
	Source information:	In DHS 2016, haemoglobin levels are adjusted for altitude in primary sampling units above 1,000 metres and for cigarette smoking among women and men. (National Department of Health SSA South African Medical Research Council, ICF. South Africa Demographic and Health Survey 2016: Report. National Department of Health; 2019. Page 272.)
	Data type:	Numeric
189	**Column name:**	**HbA1c**
	Column description:	Glycated Haemoglobin (HbA1c) [mmol/mol]
	Source information:	DHS 2016 values were converted to mmol/mol using the formula from https://diabetessociety.com.au/documents/HbA1cConversionTable.pdf: mmol/mol = 10.93 * % – 23.5 mmol/mol.
	Data type:	Numeric

190	Column name:	chol_tot
	Column description:	Total cholesterol [mmol/l]
	Source information:	Conservative cutoffs for implausible values adapted from: Castelli WP, Cooper GR, Doyle JT, et al. Distribution of triglyceride and total, LDL and HDL cholesterol in several populations: A cooperative lipoprotein phenotyping study. Journal of Chronic Diseases. 1977;30(3):147-169.
	Data type:	Numeric

191	Column name:	chol_hdl
	Column description:	High-density lipoprotein (HDL) cholesterol [mmol/l]
	Source information:	Conservative cutoffs for implausible values adapted from: Castelli WP, Cooper GR, Doyle JT, et al. Distribution of triglyceride and total, LDL and HDL cholesterol in several populations: A cooperative lipoprotein phenotyping study. Journal of Chronic Diseases. 1977;30(3):147-169.
	Data type:	Numeric

192	Column name:	chol_ldl
	Column description:	Low-density lipoprotein (LDL) cholesterol [mmol/l]
	Source information:	Conservative cutoffs for implausible values adapted from: Castelli WP, Cooper GR, Doyle JT, et al. Distribution of triglyceride and total, LDL and HDL cholesterol in several populations: A cooperative lipoprotein phenotyping study. Journal of Chronic Diseases. 1977;30(3):147-169.
	Data type:	Numeric

193	Column name:	trig
	Column description:	Triglycerides [mmol/l]
	Source information:	Conservative cutoffs for implausible values adapted from: Castelli WP, Cooper GR, Doyle JT, et al. Distribution of triglyceride and total, LDL and HDL cholesterol in several populations: A cooperative lipoprotein phenotyping study. Journal of Chronic Diseases. 1977;30(3):147-169.

	Data type:	Numeric

194	Column name:	medaid
	Column description:	Covered by medical insurance
	Data type:	Factor

195	Column name:	hcare12mo
	Column description:	Health consultation last 12 months
	Source information:	In SAGE, those who reported the last time they needed healthcare as being 0 or 1 year ago and who reported receiving healthcare the last time they needed it were coded as having a health consultation in the last year.
	Data type:	Factor

196	Column name:	hcare1mo
	Column description:	Healthcare consultations last month
	Source information:	Where the survey categorised healthcare visits into last 30 days, one month to five months ago, etc. (e.g. NIDS), those responding last 30 days were coded as yes, while those who had a healthcare visit one month to five months ago were coded as 'no'. In SAGE, those who reported the last time they needed healthcare as being 0 or 1 month ago and who reported receiving healthcare the last time they needed it were coded as having a health consultation in the last month.
	Data type:	Factor

197	Column name:	hcare1mo_public
	Column description:	Healthcare last month: public hospital/clinic
	Source information:	For surveys other than the DHS (e.g., NIDS), they were coded as either 'yes' (1) or missing since respondents were only allowed to select the last point of medical care in the past 30 days. [The DHS allowed multiple instances of care to be recorded so respondents could select multiple care types used in the past 30 days]. For the other surveys where only the last location of care could be specified, the indicator was only coded 'no' (0) when it was reported that the individual did not receive care in the past 30 days. Caution therefore needs to be taken when using these indicators since many of those

		with missing values are likely to be ‘no’ and not included in the totals.
	Data type:	Factor

198	**Column name:**	**hcare1mo_private**
	Column description:	Healthcare last month: private hospital/clinic/doctor
	Source information:	SANHANES only specified public or private.
	Data type:	Factor

199	**Column name:**	**ohcare1mo**
	Column description:	Outpatient consultations last month
	Data type:	Factor

200	**Column name:**	**ohcare1mo_public**
	Column description:	Outpatient healthcare last month: public hospital/clinic
	Data type:	Factor

201	**Column name:**	**ohcare1mo_private**
	Column description:	Outpatient healthcare last month: private hospital/clinic/doctor
	Data type:	Factor

202	**Column name:**	**hcare1mo_chem_nurse**
	Column description:	Healthcare last month: chemist/pharmacist/nurse
	Data type:	Factor

203	**Column name:**	**hcare1mo_trad**
	Column description:	Healthcare last month: traditional/faith healer
	Data type:	Factor

204	**Column name:**	**hcare1mo_other**

	Column description:	Healthcare last month: other
	Data type:	Factor
205	**Column name:**	**globorisk_nonlab**
	Column description:	Globorisk 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 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
206	**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
207	**Column name:**	**globorisk_lab_fatal**
	Column description:	Globorisk CVD fatal risk score
	Source information:	10-year 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

208	**Column name:**	**who_nonlab**
	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 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

209	**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

210	**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