

Codebook

Harmonised ExPoSE South Africa Dataset

V. 1.2

Dataset name:	DATA_SA
Dataset size:	133.8 Mb
Column count:	210
Row count:	155,626
Updated date:	2025-10-09
Changes from version 1.1	-

Column Attributes:

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
ZAF	155,626	155,626	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
South Africa	155,626	155,626	100.00

3	Column name:	source
---	--------------	--------

Column description: Data source

Data type: Factor

Unique non-missing value count: 11

Missing value count: 0

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
SAGE 2014	3153	NIDS 2008	16765
SAGE 2007-8	3874	NIDS 2010-11	18089
DHS 2003	8107	NIDS 2012	19097
DHS 2016	10333	NIDS 2014-15	23230
DHS 1998	13794	NIDS 2017	25052

4	Column name:	year		
	Column description:	Year of data collection - Survey median		
	Data type:	Numeric		
	Unique non-missing value count:	10		
	Missing value count:	0		
Min	Mean	Median	Max	SD
1998.00	2011.11	2012.00	2017.00	5.41

5	Column name:	hhid		
	Column description:	Household identifier		
	Source information:	Generated in the harmonised dataset to be unique for each household.		
	Data type:	Numeric		
	Unique non-missing value count:	68,456		
	Missing value count:	0		
Min	Mean	Median	Max	SD
100001.00	399913.85	425332.50	611083.00	120231.01

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
Unique non-missing value count:	86,831
Missing value count:	0

Min	Mean	Median	Max	SD
100001.00	468419.05	418486.00	898981.00	200495.11

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
Unique non-missing value count:	3,840
Missing value count:	0

Min	Mean	Median	Max	SD
100001.00	383525.19	404033.00	600750.00	117463.82

8	Column name:	stratum
----------	---------------------	----------------

Column description:	Sampling stratum
Source information:	Generated in the harmonised dataset to be unique.
Data type:	Factor
Unique non-missing value count:	243
Missing value count:	0

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
300922	5	400023	2454
200014	6	400171	2554
300512	6	400021	2703

300017	7	400774	2986
300012	9	400572	3656

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
Unique non-missing value count:	53,418
Missing value count:	0

Min	Mean	Median	Max	SD
0.00	1.00	0.62	125.06	1.45

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
Unique non-missing value count:	49,068
Missing value count:	7,873

Min	Mean	Median	Max	SD
0.00	1.00	0.62	125.06	1.49

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
Unique non-missing value count:	48,496
Missing value count:	9,476

Min	Mean	Median	Max	SD
-----	------	--------	-----	----

0.00	1.00	0.62	125.06	1.50
------	------	------	--------	------

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
Unique non-missing value count:	118,235
Missing value count:	0

Min	Mean	Median	Max	SD
-4.29	1.00	0.60	197.95	1.59

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
Unique non-missing value count:	114,669
Missing value count:	7,873

Min	Mean	Median	Max	SD
-4.29	1.00	0.59	197.95	1.63

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
Unique non-missing value count:	113,739
Missing value count:	9,476

Min	Mean	Median	Max	SD
-----	------	--------	-----	----

-4.29	1.00	0.59	197.95	1.64
-------	------	------	--------	------

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

Unique non-missing value count: 9

Missing value count: 0

Categories	Frequency	Cumulative Frequency	Percent
Eastern Cape	20,693	20,693	13.30
Free State	10,440	31,133	6.71
Gauteng	18,083	49,216	11.62
KwaZulu Natal	37,089	86,305	23.83
Limpopo	15,076	101,381	9.69
Mpumalanga	11,848	113,229	7.61
North West	14,324	127,553	9.20
Northern Cape	10,764	138,317	6.92
Western Cape	17,309	155,626	11.12

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

Unique non-missing value count: 9

Missing value count: 0

Categories	Frequency	Cumulative Frequency	Percent
EC	20,693	20,693	13.30
FS	10,440	31,133	6.71
GT	18,083	49,216	11.62
KZN	37,089	86,305	23.83
LI	15,076	101,381	9.69
MP	11,848	113,229	7.61
NC	10,764	123,993	6.92
NW	14,324	138,317	9.20
WC	17,309	155,626	11.12

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

Unique non-missing value count: 53

Missing value count: 0

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
Metsweding District Municipality	926	Ugu District Municipality	4218
Xhariep District Municipality	1275	Nelson Mandela Bay Metropolitan Municipality	4239

Northern Free State District Municipality	1405	City of Johannesburg Metropolitan Municipality	5675
Umkhanyakude District Municipality	1484	City of Cape Town Metropolitan Municipality	5908
Bohlabela District Municipality	1618	Ethekwini Metropolitan Municipality	8087

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

Unique non-missing value count: 53

Missing value count: 0

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
CBDC2	926	DC21	4218
DC16	1275	NMA	4239
DC20	1405	JHB	5675
DC27	1484	CPT	5908
CBDC4	1618	ETH	8087

19	Column name:	geotype2
----	--------------	----------

Column description: Urban/rural

Data type: Factor

Unique non-missing value count: 2

Missing value count: 0

Categories	Frequency	Cumulative Frequency	Percent
------------	-----------	----------------------	---------

Non-urban	70,830	70,830	45.51
Urban	84,796	155,626	54.49

20	Column name:	intm
-----------	---------------------	-------------

Column description: Interview - Month

Data type: Factor

Unique non-missing value count: 12

Missing value count: 3,493

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
December	3052	April	15028
Missing	3493	June	15666
January	4077	July	16539
November	8633	May	17420
February	12053	March	17696

21	Column name:	inty
-----------	---------------------	-------------

Column description: Interview - Year

Data type: Numeric

Unique non-missing value count: 12

Missing value count: 18

Min	Mean	Median	Max	SD
1998.00	2011.07	2012.00	2017.00	5.41

22	Column name:	vism
-----------	---------------------	-------------

Column description: Anthropometry - Month

Source information: Month of anthropometric measurements.

Data type: Factor

Unique non-missing value count: 12

Missing value count: 3,493

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
December	3052	April	15028
Missing	3493	June	15666
January	4077	July	16539
November	8633	May	17420
February	12053	March	17696

23	Column name:	visq
-----------	---------------------	-------------

Column description:	Anthropometry - Quarter
Source information:	Quarter of anthropometric measurements.
Data type:	Factor
Unique non-missing value count:	4
Missing value count:	3,493

Categories	Frequency	Cumulative Frequency	Percent
I	33,826	33,826	21.74
II	48,114	81,940	30.92
III	45,094	127,034	28.98
IV	25,099	152,133	16.13
Missing	3,493	155,626	2.24

24	Column name:	hh_size
-----------	---------------------	----------------

Column description:	Household size
Data type:	Numeric
Unique non-missing value count:	32
Missing value count:	3,159

Min	Mean	Median	Max	SD
1.00	5.26	5.00	41.00	3.27

25	Column name:	hh_size_cat
-----------	---------------------	--------------------

	Column description:	Household size, categorical		
	Data type:	Factor		
	Unique non-missing value count:	6		
	Missing value count:	3,159		
	Categories	Frequency	Cumulative Frequency	Percent
	1	11,322	11,322	7.28
	2	17,262	28,584	11.09
	3	20,376	48,960	13.09
	4	23,137	72,097	14.87
	5	21,188	93,285	13.61
	6+	59,182	152,467	38.03
	Missing	3,159	155,626	2.03
26	Column name:	hh_ownhome		
	Column description:	Dwelling - Ownership		
	Data type:	Factor		
	Unique non-missing value count:	2		
	Missing value count:	49,774		
	Categories	Frequency	Cumulative Frequency	Percent
	No	23,622	23,622	15.18
	Yes	82,230	105,852	52.84
	Missing	49,774	155,626	31.98
27	Column name:	hh_totrooms		
	Column description:	Number of rooms in dwelling		
	Data type:	Numeric		
	Unique non-missing value count:	27		
	Missing value count:	42,708		
	Min	Mean	Median	Max
				SD

1.00	4.44	4.00	44.00	2.45
------	------	------	-------	------

28	Column name:	hh_sleeprooms
-----------	---------------------	----------------------

Column description:	Number of rooms for sleeping in dwelling
Data type:	Numeric
Unique non-missing value count:	17
Missing value count:	117,980

Min	Mean	Median	Max	SD
0.00	2.56	2.00	18.00	1.28

29	Column name:	hh_ptotrooms
-----------	---------------------	---------------------

Column description:	Number of rooms in dwelling, per household member
Data type:	Numeric
Unique non-missing value count:	210
Missing value count:	42,708

Min	Mean	Median	Max	SD
0.03	1.46	1.14	24.00	1.20

30	Column name:	hh_psleeprooms
-----------	---------------------	-----------------------

Column description:	Number of rooms used for sleeping, per household member
Data type:	Numeric
Unique non-missing value count:	102
Missing value count:	117,980

Min	Mean	Median	Max	SD
0.00	0.65	0.50	12.00	0.52

31	Column name:	hh_dwellingtype
-----------	---------------------	------------------------

Column description:	Dwelling - Type
Data type:	Factor

Unique non-missing value count: 10
Missing value count: 21,802

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
Caravan/tent	71	Backyard dwelling	6073
Other	810	Informal dwelling not in backyard	7125
Room/flatlet	1139	Traditional structure made of traditional materials	16565
Town/cluster/semi-detached house in complex	3137	Missing	21802
Flat/apartment in block	3408	Structure (brick) on separate stand/yard	91017

32	Column name:	hh_wallmaterial	
	Column description:	Dwelling - Wall material	
	Data type:	Factor	
	Unique non-missing value count:	6	
	Missing value count:	21,763	
	Categories	Frequency	Cumulative Frequency
	Brick/cement/prefab/plaster	98,860	98,860
	Corrugated iron/zinc	10,951	109,811
	Mud and cement mix	10,499	120,310
	Mud/tatching/wattle and daub	9,578	129,888
	Other	3,024	132,912
	Plastic/cardboard	951	133,863
	Missing	21,763	155,626
		Percent	
	Brick/cement/prefab/plaster	63.52	
	Corrugated iron/zinc	7.04	
	Mud and cement mix	6.75	
	Mud/tatching/wattle and daub	6.15	
	Other	1.94	
	Plastic/cardboard	0.61	
	Missing	13.98	

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
Unique non-missing value count:	7
Missing value count:	38,224

Categories	Frequency	Cumulative Frequency	Percent
Carpet	14,487	14,487	9.31
Concrete/cement	45,438	59,925	29.20
Linoleum/vinyl	10,342	70,267	6.65
Mud/earth/dung	13,664	83,931	8.78
Other	63	83,994	0.04
Tiles	30,704	114,698	19.73
Wood	2,704	117,402	1.74
Missing	38,224	155,626	24.56

34

Column name:

hh_roofmaterial

Column description:	Dwelling - Wall material
Data type:	Factor
Unique non-missing value count:	9
Missing value count:	43,373

Categories	Frequency	Cumulative Frequency	Percent
Asbestos	10,436	10,436	6.71
Brick/cement	4,750	15,186	3.05
Corrugated iron/zinc	76,554	91,740	49.19
Mud and cement mix	586	92,326	0.38
Mud/thatching/wattle and daub	4,011	96,337	2.58
Other/No roof	87	96,424	0.06
Plastic/cardboard	349	96,773	0.22
Tile	14,699	111,472	9.45

Wood	781	112,253	0.50
Missing	43,373	155,626	27.87

35	Column name:	hh_roof_wall_1
-----------	---------------------	-----------------------

Column description:	Dwelling - roof/wall material: mud/thatching/wattle and daub
Data type:	Factor
Unique non-missing value count:	2
Missing value count:	141,494

Categories	Frequency	Cumulative Frequency	Percent
No	13,219	13,219	8.49
Yes	913	14,132	0.59
Missing	141,494	155,626	90.92

36	Column name:	hh_roof_wall_2
-----------	---------------------	-----------------------

Column description:	Dwelling - roof/wall material: mud and cement mix
Data type:	Factor
Unique non-missing value count:	2
Missing value count:	141,494

Categories	Frequency	Cumulative Frequency	Percent
No	13,393	13,393	8.61
Yes	739	14,132	0.47
Missing	141,494	155,626	90.92

37	Column name:	hh_roof_wall_3
-----------	---------------------	-----------------------

Column description:	Dwelling - roof/wall material: corrugated iron/zinc
Data type:	Factor
Unique non-missing value count:	2
Missing value count:	141,494

Categories	Frequency	Cumulative Frequency	Percent
------------	-----------	----------------------	---------

No	8,476	8,476	5.45
Yes	5,656	14,132	3.63
Missing	141,494	155,626	90.92

38	Column name:	hh_roof_wall_4
-----------	---------------------	-----------------------

Column description:	Dwelling - roof/wall material: plastic/cardboard
Data type:	Factor
Unique non-missing value count:	2
Missing value count:	141,494

Categories	Frequency	Cumulative Frequency	Percent
No	13,730	13,730	8.82
Yes	402	14,132	0.26
Missing	141,494	155,626	90.92

39	Column name:	hh_roof_wall_5
-----------	---------------------	-----------------------

Column description:	Dwelling - roof/wall material: brick/cement/prefab/plaster
Data type:	Factor
Unique non-missing value count:	2
Missing value count:	141,494

Categories	Frequency	Cumulative Frequency	Percent
No	2,884	2,884	1.85
Yes	11,248	14,132	7.23
Missing	141,494	155,626	90.92

40	Column name:	hh_roof_wall_9999
-----------	---------------------	--------------------------

Column description:	Dwelling - roof/wall material: other
Data type:	Factor
Unique non-missing value count:	2
Missing value count:	141,494

Categories	Frequency	Cumulative Frequency	Percent
No	11,170	11,170	7.18
Yes	2,962	14,132	1.90
Missing	141,494	155,626	90.92

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

Unique non-missing value count: 8

Missing value count: 3,964

Categories	Frequency	Cumulative Frequency	Percent
Borehole/well	4,292	4,292	2.76
Other	1,750	6,042	1.12
Piped in dwelling	64,988	71,030	41.76
Piped on site or yard	39,343	110,373	25.28
Public tap	27,010	137,383	17.36
Rainwater tank	1,087	138,470	0.70
Surface water	9,989	148,459	6.42
Water carrier/tanker	3,203	151,662	2.06
Missing	3,964	155,626	2.55

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

Unique non-missing value count: 5

Missing value count: 3,996

Categories	Frequency	Cumulative Frequency	Percent
Flush toilet/chemical toilet	82,413	82,413	52.96
No facility/bush	7,460	89,873	4.79
Other	409	90,282	0.26
Traditional pit latrine/bucket	40,979	131,261	26.33
Ventilated improved pit latrine	20,369	151,630	13.09
Missing	3,996	155,626	2.57

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

Unique non-missing value count: 2

Missing value count: 17,899

Categories	Frequency	Cumulative Frequency	Percent
No	110,570	110,570	71.05
Yes	27,157	137,727	17.45
Missing	17,899	155,626	11.50

44	Column name:	hh_refuseremoved
-----------	---------------------	-------------------------

Column description: Refuse removal

Data type: Factor

Unique non-missing value count: 2

Missing value count: 21,877

Categories	Frequency	Cumulative Frequency	Percent
No	66,985	66,985	43.04
Yes	66,764	133,749	42.90
Missing	21,877	155,626	14.06

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

Unique non-missing value count: 7

Missing value count: 17,775

Categories	Frequency	Cumulative Frequency	Percent
Animal dung	205	205	0.13
Coal	1,732	1,937	1.11
Electricity	101,493	103,430	65.22
Gas	3,907	107,337	2.51
Other	540	107,877	0.35
Paraffin	8,116	115,993	5.22
Wood/straw	21,858	137,851	14.05
Missing	17,775	155,626	11.42

46	Column name:	hh_heatingfuel
-----------	---------------------	-----------------------

Column description: Main heating fuel

Data type: Factor

Unique non-missing value count: 7

Missing value count: 23,187

Categories	Frequency	Cumulative Frequency	Percent
Animal dung	252	252	0.16

Coal	3,247	3,499	2.09
Electricity	78,833	82,332	50.66
Gas	1,565	83,897	1.01
Other	12,692	96,589	8.16
Paraffin	8,036	104,625	5.16
Wood/straw	27,814	132,439	17.87
Missing	23,187	155,626	14.90

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

Unique non-missing value count: 2

Missing value count: 51,294

Categories	Frequency	Cumulative Frequency	Percent
No	7,008	7,008	4.50
Yes	97,324	104,332	62.54
Missing	51,294	155,626	32.96

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

Unique non-missing value count: 2

Missing value count: 138,049

Categories	Frequency	Cumulative Frequency	Percent
No	12,871	12,871	8.27
Yes	4,706	17,577	3.02
Missing	138,049	155,626	88.71

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

Unique non-missing value count: 2

Missing value count: 134,171

Categories	Frequency	Cumulative Frequency	Percent
No	8,958	8,958	5.76
Yes	12,497	21,455	8.03
Missing	134,171	155,626	86.21

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

Unique non-missing value count: 2

Missing value count: 121,132

Categories	Frequency	Cumulative Frequency	Percent
No	9,929	9,929	6.38
Yes	24,565	34,494	15.78
Missing	121,132	155,626	77.84

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

Unique non-missing value count: 2

Missing value count: 140,248

Categories	Frequency	Cumulative Frequency	Percent
No	12,530	12,530	8.05
Yes	2,848	15,378	1.83
Missing	140,248	155,626	90.12

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

Unique non-missing value count: 2

Missing value count: 141,533

Categories	Frequency	Cumulative Frequency	Percent
No	13,747	13,747	8.83
Yes	346	14,093	0.22
Missing	141,533	155,626	90.94

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

Unique non-missing value count: 2

Missing value count: 141,567

Categories	Frequency	Cumulative Frequency	Percent
No	13,885	13,885	8.92
Yes	174	14,059	0.11
Missing	141,567	155,626	90.97

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

Unique non-missing value count: 2

Missing value count: 21,562

Categories	Frequency	Cumulative Frequency	Percent
No	53,015	53,015	34.07
Yes	81,049	134,064	52.08
Missing	21,562	155,626	13.86

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

Unique non-missing value count: 2

Missing value count: 151,764

Categories	Frequency	Cumulative Frequency	Percent
No	2,969	2,969	1.91
Yes	893	3,862	0.57
Missing	151,764	155,626	97.52

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

Unique non-missing value count: 2

Missing value count: 116,934

Categories	Frequency	Cumulative Frequency	Percent
No	26,157	26,157	16.81
Yes	12,535	38,692	8.05
Missing	116,934	155,626	75.14

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

Unique non-missing value count: 2

Missing value count: 121,774

Categories	Frequency	Cumulative Frequency	Percent
No	23,883	23,883	15.35
Yes	9,969	33,852	6.41
Missing	121,774	155,626	78.25

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

Unique non-missing value count: 2
Missing value count: 128,562

Categories	Frequency	Cumulative Frequency	Percent
No	21,519	21,519	13.83
Yes	5,545	27,064	3.56
Missing	128,562	155,626	82.61

59	Column name:	hh_ass_elec
-----------	---------------------	--------------------

Column description: Household assets: Electricity
Data type: Factor
Unique non-missing value count: 2
Missing value count: 18,110

Categories	Frequency	Cumulative Frequency	Percent
No	25,142	25,142	16.16
Yes	112,374	137,516	72.21
Missing	18,110	155,626	11.64

60	Column name:	hh_ass_radio
-----------	---------------------	---------------------

Column description: Household assets: Radio
Data type: Factor
Unique non-missing value count: 2
Missing value count: 4,793

Categories	Frequency	Cumulative Frequency	Percent
No	48,879	48,879	31.41
Yes	101,954	150,833	65.51
Missing	4,793	155,626	3.08

61	Column name:	hh_ass_tv
-----------	---------------------	------------------

Column description: Household assets: TV
Data type: Factor

Unique non-missing value count: 2
Missing value count: 4,724

Categories	Frequency	Cumulative Frequency	Percent
No	36,935	36,935	23.73
Yes	113,967	150,902	73.23
Missing	4,724	155,626	3.04

62	Column name:	hh_ass_fridge
-----------	---------------------	----------------------

Column description: Household assets: Fridge
Data type: Factor
Unique non-missing value count: 2
Missing value count: 4,701

Categories	Frequency	Cumulative Frequency	Percent
No	44,475	44,475	28.58
Yes	106,450	150,925	68.40
Missing	4,701	155,626	3.02

63	Column name:	hh_ass_bicycle
-----------	---------------------	-----------------------

Column description: Household assets: Bicycle
Data type: Factor
Unique non-missing value count: 2
Missing value count: 18,435

Categories	Frequency	Cumulative Frequency	Percent
No	123,704	123,704	79.49
Yes	13,487	137,191	8.67
Missing	18,435	155,626	11.85

64	Column name:	hh_ass_motorcycle
-----------	---------------------	--------------------------

Column description: Household assets: Motorcycle
Data type: Factor

Unique non-missing value count: 2

Missing value count: 18,417

Categories	Frequency	Cumulative Frequency	Percent
No	135,050	135,050	86.78
Yes	2,159	137,209	1.39
Missing	18,417	155,626	11.83

65	Column name:	hh_ass_car_truck
-----------	---------------------	-------------------------

Column description: Household assets: Car/Truck

Data type: Factor

Unique non-missing value count: 2

Missing value count: 5,033

Categories	Frequency	Cumulative Frequency	Percent
No	117,384	117,384	75.43
Yes	33,209	150,593	21.34
Missing	5,033	155,626	3.23

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

Unique non-missing value count: 2

Missing value count: 4,513

Categories	Frequency	Cumulative Frequency	Percent
No	132,060	132,060	84.86
Yes	19,053	151,113	12.24
Missing	4,513	155,626	2.90

67	Column name:	hh_ass_computer
-----------	---------------------	------------------------

Column description: Household assets: Computer

Data type: Factor

Unique non-missing value count: 2

Missing value count: 5,061

Categories	Frequency	Cumulative Frequency	Percent
No	128,900	128,900	82.83
Yes	21,665	150,565	13.92
Missing	5,061	155,626	3.25

68	Column name:	hh_ass_wmachine
-----------	---------------------	------------------------

Column description: Household assets: Washing Machine

Data type: Factor

Unique non-missing value count: 2

Missing value count: 12,990

Categories	Frequency	Cumulative Frequency	Percent
No	98,025	98,025	62.99
Yes	44,611	142,636	28.67
Missing	12,990	155,626	8.35

69	Column name:	hh_ass_cellphone
-----------	---------------------	-------------------------

Column description: Household assets: Electricity

Data type: Factor

Unique non-missing value count: 2

Missing value count: 18,491

Categories	Frequency	Cumulative Frequency	Percent
No	21,236	21,236	13.65
Yes	115,899	137,135	74.47
Missing	18,491	155,626	11.88

70	Column name:	hh_ass_watch
-----------	---------------------	---------------------

Column description: Household assets: Phone (cellular)

Data type: Factor

Unique non-missing value count: 2

Missing value count: 145,293

Categories	Frequency	Cumulative Frequency	Percent
No	5,577	5,577	3.58
Yes	4,756	10,333	3.06
Missing	145,293	155,626	93.36

71	Column name:	hh_ass_animalcart
-----------	---------------------	--------------------------

Column description: Household assets: Animal cart

Data type: Factor

Unique non-missing value count: 2

Missing value count: 40,191

Categories	Frequency	Cumulative Frequency	Percent
No	114,260	114,260	73.42
Yes	1,175	115,435	0.76
Missing	40,191	155,626	25.83

72	Column name:	hh_ass_motorboat
-----------	---------------------	-------------------------

Column description: Household assets: Motorboat

Data type: Factor

Unique non-missing value count: 2

Missing value count: 44,047

Categories	Frequency	Cumulative Frequency	Percent
No	111,038	111,038	71.35
Yes	541	111,579	0.35
Missing	44,047	155,626	28.30

73	Column name:	hh_ass_vacuum
-----------	---------------------	----------------------

Column description: Household assets: Vacuum cleaner

Data type: Factor

Unique non-missing value count: 2

Missing value count: 131,954

Categories	Frequency	Cumulative Frequency	Percent
No	19,957	19,957	12.82
Yes	3,715	23,672	2.39
Missing	131,954	155,626	84.79

74	Column name:	hh_ass_microwave
-----------	---------------------	-------------------------

Column description: Household assets: Microwave oven

Data type: Factor

Unique non-missing value count: 2

Missing value count: 44,034

Categories	Frequency	Cumulative Frequency	Percent
No	62,027	62,027	39.86
Yes	49,565	111,592	31.85
Missing	44,034	155,626	28.29

75	Column name:	hh_ass_stove
-----------	---------------------	---------------------

Column description: Household assets: Stove

Data type: Factor

Unique non-missing value count: 2

Missing value count: 30,368

Categories	Frequency	Cumulative Frequency	Percent
No	24,967	24,967	16.04
Yes	100,291	125,258	64.44
Missing	30,368	155,626	19.51

76	Column name:	hh_ass_sat
-----------	---------------------	-------------------

Column description: Household assets: Satellite TV

Data type: Factor

Unique non-missing value count: 2

Missing value count: 40,991

Categories	Frequency	Cumulative Frequency	Percent
No	84,343	84,343	54.20
Yes	30,292	114,635	19.46
Missing	40,991	155,626	26.34

77	Column name:	hh_ass_video
-----------	---------------------	---------------------

Column description: Household assets: Video player

Data type: Factor

Unique non-missing value count: 2

Missing value count: 40,883

Categories	Frequency	Cumulative Frequency	Percent
No	69,259	69,259	44.50
Yes	45,484	114,743	29.23
Missing	40,883	155,626	26.27

78	Column name:	hh_ass_hifi
-----------	---------------------	--------------------

Column description: Household assets: hifi

Data type: Factor

Unique non-missing value count: 2

Missing value count: 50,492

Categories	Frequency	Cumulative Frequency	Percent
No	64,860	64,860	41.68
Yes	40,274	105,134	25.88
Missing	50,492	155,626	32.44

79	Column name:	hh_ass_camera
-----------	---------------------	----------------------

Column description: Household assets: Photo camera

Data type: Factor

Unique non-missing value count: 2

Missing value count: 54,430

Categories	Frequency	Cumulative Frequency	Percent
No	91,601	91,601	58.86
Yes	9,595	101,196	6.17
Missing	54,430	155,626	34.97

80	Column name:	hh_ass_smachine
-----------	---------------------	------------------------

Column description: Household assets: Sewing machine

Data type: Factor

Unique non-missing value count: 2

Missing value count: 50,608

Categories	Frequency	Cumulative Frequency	Percent
No	96,182	96,182	61.80
Yes	8,836	105,018	5.68
Missing	50,608	155,626	32.52

81	Column name:	hh_ass_sofa
-----------	---------------------	--------------------

Column description: Household assets: Sofa

Data type: Factor

Unique non-missing value count: 2

Missing value count: 54,339

Categories	Frequency	Cumulative Frequency	Percent
No	47,994	47,994	30.84
Yes	53,293	101,287	34.24
Missing	54,339	155,626	34.92

82	Column name:	hh_ass_boat
-----------	---------------------	--------------------

Column description: Household assets: Boat

Data type: Factor

Unique non-missing value count: 2

Missing value count: 54,392

Categories	Frequency	Cumulative Frequency	Percent
No	100,834	100,834	64.79
Yes	400	101,234	0.26
Missing	54,392	155,626	34.95

83	Column name:	hh_ass_plough
-----------	---------------------	----------------------

Column description: Household assets: Plough

Data type: Factor

Unique non-missing value count: 2

Missing value count: 54,345

Categories	Frequency	Cumulative Frequency	Percent
No	98,277	98,277	63.15
Yes	3,004	101,281	1.93
Missing	54,345	155,626	34.92

84	Column name:	hh_ass_tractor
-----------	---------------------	-----------------------

Column description: Household assets: Tractor

Data type: Factor

Unique non-missing value count: 2

Missing value count: 54,379

Categories	Frequency	Cumulative Frequency	Percent
No	100,069	100,069	64.30
Yes	1,178	101,247	0.76
Missing	54,379	155,626	34.94

85	Column name:	hh_ass_wheelbarrow		
	Column description:	Household assets: Wheelbarrow		
	Data type:	Factor		
	Unique non-missing value count:	2		
	Missing value count:	54,367		
Categories	Frequency	Cumulative Frequency	Percent	
No	71,311	71,311	45.82	
Yes	29,948	101,259	19.24	
Missing	54,367	155,626	34.93	

86	Column name:	hh_ass_mill		
	Column description:	Household assets: Mill		
	Data type:	Factor		
	Unique non-missing value count:	2		
	Missing value count:	54,467		
Categories	Frequency	Cumulative Frequency	Percent	
No	99,701	99,701	64.06	
Yes	1,458	101,159	0.94	
Missing	54,467	155,626	35.00	

87	Column name:	hh_ass_tab		
	Column description:	Household assets: Table		
	Data type:	Factor		
	Unique non-missing value count:	2		
	Missing value count:	151,758		
Categories	Frequency	Cumulative Frequency	Percent	
No	363	363	0.23	
Yes	3,505	3,868	2.25	
Missing	151,758	155,626	97.51	

88	Column name:	hh_ass_sink
-----------	---------------------	--------------------

Column description: Household assets: Sink

Data type: Factor

Unique non-missing value count: 2

Missing value count: 151,769

Categories	Frequency	Cumulative Frequency	Percent
No	2,230	2,230	1.43
Yes	1,627	3,857	1.05
Missing	151,769	155,626	97.52

89	Column name:	hh_ass_hotw
-----------	---------------------	--------------------

Column description: Household assets: Hot water

Data type: Factor

Unique non-missing value count: 2

Missing value count: 151,767

Categories	Frequency	Cumulative Frequency	Percent
No	2,570	2,570	1.65
Yes	1,289	3,859	0.83
Missing	151,767	155,626	97.52

90	Column name:	hh_ass_dishwasher
-----------	---------------------	--------------------------

Column description: Household assets: Dishwasher

Data type: Factor

Unique non-missing value count: 2

Missing value count: 151,768

Categories	Frequency	Cumulative Frequency	Percent
No	3,500	3,500	2.25
Yes	358	3,858	0.23
Missing	151,768	155,626	97.52

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		
	Unique non-missing value count:	2		
	Missing value count:	4,713		
	Categories	Frequency	Cumulative Frequency	Percent
	No	141,143	141,143	90.69
	Yes	9,770	150,913	6.28
	Missing	4,713	155,626	3.03

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		
	Unique non-missing value count:	2		
	Missing value count:	3,935		
	Categories	Frequency	Cumulative Frequency	Percent
	No	82,206	82,206	52.82
	Yes	69,485	151,691	44.65

Missing	3,935	155,626	2.53
---------	-------	---------	------

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
Unique non-missing value count:	2
Missing value count:	3,964

Categories	Frequency	Cumulative Frequency	Percent
No	136,720	136,720	87.85
Yes	14,942	151,662	9.60
Missing	3,964	155,626	2.55

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
Unique non-missing value count:	2
Missing value count:	3,981

Categories	Frequency	Cumulative Frequency	Percent
No	122,370	122,370	78.63
Yes	29,275	151,645	18.81

Missing	3,981	155,626	2.56
---------	-------	---------	------

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
Unique non-missing value count:	2
Missing value count:	3,153

Categories	Frequency	Cumulative Frequency	Percent
No	69,501	69,501	44.66
Yes	82,972	152,473	53.31
Missing	3,153	155,626	2.03

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
Unique non-missing value count:	2
Missing value count:	3,153

Categories	Frequency	Cumulative Frequency	Percent
No	121,599	121,599	78.14
Yes	30,874	152,473	19.84
Missing	3,153	155,626	2.03

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

Unique non-missing value count: 2

Missing value count: 3,153

Categories	Frequency	Cumulative Frequency	Percent
No	143,865	143,865	92.44
Yes	8,608	152,473	5.53
Missing	3,153	155,626	2.03

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

Unique non-missing value count: 2

Missing value count: 3,153

Categories	Frequency	Cumulative Frequency	Percent
No	151,455	151,455	97.32
Yes	1,018	152,473	0.65
Missing	3,153	155,626	2.03

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

Unique non-missing value count: 21,188

Missing value count: 50,340

Min	Mean	Median	Max	SD
0.00	7278.19	3800.00	2605523.50	17774.66

100	Column name:	hh_income Quint
------------	---------------------	------------------------

Column description:	Household income quintile
Data type:	Factor
Unique non-missing value count:	5
Missing value count:	49,538

Categories	Frequency	Cumulative Frequency	Percent
I (Lowest)	17,316	17,316	11.13
II	23,613	40,929	15.17
III	25,160	66,089	16.17
IV	23,997	90,086	15.42
V (Highest)	16,002	106,088	10.28
Missing	49,538	155,626	31.83

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			
Unique non-missing value count:	63,999			
Missing value count:	3,153			
Min	Mean	Median	Max	SD

-7.49	0.03	0.01	9.61	2.71
-------	------	------	------	------

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
	Unique non-missing value count:	5
	Missing value count:	3,153
Categories	Frequency	Cumulative Frequency
I (Lowest)	38,169	38,169
II	31,668	69,837
III	30,051	99,888
IV	29,685	129,573
V (Highest)	22,900	152,473
Missing	3,153	155,626
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

Unique non-missing value count: 64,002

Missing value count: 2,435

Min	Mean	Median	Max	SD
-7.31	0.45	0.35	10.11	2.76

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

Unique non-missing value count: 2

Missing value count: 40,174

Categories	Frequency	Cumulative Frequency	Percent
No	108,968	108,968	70.02
Yes	6,484	115,452	4.17
Missing	40,174	155,626	25.81

105	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	91,250	91,250	58.63

Male	64,376	155,626	41.37
------	--------	---------	-------

106	Column name:	age			
	Column description:	Age			
	Data type:	Numeric			
	Unique non-missing value count:	92			
	Missing value count:	0			
	Min	Mean	Median	Max	SD
	15.00	38.61	35.00	110.00	18.04

107	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
80+	2820	35-39	12637
75-79	3025	30-34	14460
70-74	4355	25-29	17362
65-69	6011	20-24	20807
60-64	7890	15-19	24033

108	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
	15-19	24,033	24,033	15.44

20-29	38,169	62,202	24.53
30-39	27,097	89,299	17.41
40-49	21,806	111,105	14.01
50-59	20,420	131,525	13.12
60-69	13,901	145,426	8.93
70-79	7,380	152,806	4.74
80+	2,820	155,626	1.81

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		
	Unique non-missing value count:	4		
	Missing value count:	382		
	Categories	Frequency	Cumulative Frequency	Percent
	Asian	4,554	4,554	2.93
	Black African	121,537	126,091	78.10
	Coloured	21,966	148,057	14.11
	White	7,187	155,244	4.62
	Missing	382	155,626	0.25
110	Column name:	race_imp		

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

Unique non-missing value count: 4

Missing value count: 0

Categories	Frequency	Cumulative Frequency	Percent
Asian	4,618	4,618	2.97
Black African	121,748	126,366	78.23
Coloured	22,036	148,402	14.16
White	7,224	155,626	4.64

111	Column name:	marstatus
------------	---------------------	------------------

Column description: Marital status

Data type: Factor

Unique non-missing value count: 3

Missing value count: 39,115

Categories	Frequency	Cumulative Frequency	Percent
Married/living with partner	38,791	38,791	24.93
Never married/single	63,698	102,489	40.93
Widowed/divorced/separated	14,022	116,511	9.01
Missing	39,115	155,626	25.13

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

Unique non-missing value count: 7

Missing value count: 9,064

Categories	Frequency	Cumulative Frequency	Percent
Any higher education	13,691	13,691	8.80
Completed primary	10,253	23,944	6.59
Completed secondary	23,489	47,433	15.09
No education	14,798	62,231	9.51
Other	329	62,560	0.21
Some primary	20,639	83,199	13.26
Some secondary	63,363	146,562	40.71
Missing	9,064	155,626	5.82

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

Unique non-missing value count: 6

Missing value count: 2,097

Categories	Frequency	Cumulative Frequency	Percent
Completed primary	11,842	11,842	7.61
Completed secondary/Any higher education	38,472	50,314	24.72
No education	16,353	66,667	10.51
Other	329	66,996	0.21
Some primary	21,760	88,756	13.98
Some secondary	64,773	153,529	41.62
Missing	2,097	155,626	1.35

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

Unique non-missing value count: 2

Missing value count: 6,462

Categories	Frequency	Cumulative Frequency	Percent
Employed	52,804	52,804	33.93
Unemployed	96,360	149,164	61.92
Missing	6,462	155,626	4.15

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

Unique non-missing value count: 3

Missing value count: 5,541

Categories	Frequency	Cumulative Frequency	Percent
Current smoker	28,926	28,926	18.59

Former smoker	6,420	35,346	4.13
Never smoker	114,739	150,085	73.73
Missing	5,541	155,626	3.56

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	
	Unique non-missing value count:	2	
	Missing value count:	5,541	
	Categories	Frequency	Cumulative Frequency
	No	121,159	121,159
	Yes	28,926	150,085
	Missing	5,541	155,626
			Percent
	No	121,159	77.85
	Yes	28,926	18.59
	Missing	5,541	3.56

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	
	Unique non-missing value count:	3	
	Missing value count:	43,176	

Categories	Frequency	Cumulative Frequency	Percent
Current drinker	28,895	28,895	18.57
Former drinker	10,148	39,043	6.52
Never drinker	73,407	112,450	47.17
Missing	43,176	155,626	27.74

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

Unique non-missing value count: 2

Missing value count: 29,378

Categories	Frequency	Cumulative Frequency	Percent
No	93,932	93,932	60.36
Yes	32,316	126,248	20.77
Missing	29,378	155,626	18.88

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

Unique non-missing value count: 159

Missing value count: 36,629

Min	Mean	Median	Max	SD
0.00	3.36	0.00	360.00	12.89

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

Unique non-missing value count: 1,119

Missing value count: 136,012

Min	Mean	Median	Max	SD
0.00	3705.24	560.00	84000.00	7393.81

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
Unique non-missing value count:	4
Missing value count:	136,012

Categories	Frequency	Cumulative Frequency	Percent
Highly active [8001+]	2,932	2,932	1.88
Inactive [0-600]	9,934	12,866	6.38
Low active [601-4000]	4,954	17,820	3.18
Moderately active [4001-8000]	1,794	19,614	1.15
Missing	136,012	155,626	87.40

122	Column name:	exercisefreq
------------	---------------------	---------------------

Column description:	Exercise frequency
Source information:	Only considers leisure time physical activity.
Data type:	Factor
Unique non-missing value count:	5
Missing value count:	36,207

Categories	Frequency	Cumulative Frequency	Percent
Less than once a week	7,844	7,844	5.04
Never	85,335	93,179	54.83
Once a week	5,178	98,357	3.33
Three or more times a week	14,857	113,214	9.55
Twice a week	6,205	119,419	3.99
Missing	36,207	155,626	23.27

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		
	Unique non-missing value count:	4		
	Missing value count:	18,391		
	Categories	Frequency	Cumulative Frequency	Percent
	Average/fair	21,071	21,071	13.54
	Good	41,407	62,478	26.61
	Poor/bad	7,869	70,347	5.06
	Very good/excellent	66,888	137,235	42.98
	Missing	18,391	155,626	11.82

124	Column name:	diag_hbp		
	Column description:	Diagnosis: Hypertension		
	Data type:	Factor		
	Unique non-missing value count:	2		
	Missing value count:	9,504		
	Categories	Frequency	Cumulative Frequency	Percent
	No	123,290	123,290	79.22
	Yes	22,832	146,122	14.67
	Missing	9,504	155,626	6.11

125	Column name:	diag_isch		
-----	--------------	-----------	--	--

Source information: SANHANES, DHS and SAGE specify heart attack or angina which were included in the variable diag_isch.

Data type: Factor

Unique non-missing value count: 2

Missing value count: 103,176

Categories	Frequency	Cumulative Frequency	Percent
No	50,192	50,192	32.25
Yes	2,258	52,450	1.45
Missing	103,176	155,626	66.30

126	Column name:	diag_stroke
------------	---------------------	--------------------

Column description: Diagnosis: Stroke

Data type: Factor

Unique non-missing value count: 2

Missing value count: 5,459

Categories	Frequency	Cumulative Frequency	Percent
No	148,343	148,343	95.32
Yes	1,824	150,167	1.17
Missing	5,459	155,626	3.51

127	Column name:	diag_chol
------------	---------------------	------------------

Column description: Diagnosis: Hypercholesterolaemia

Data type: Factor

Unique non-missing value count: 2

Missing value count: 110,336

Categories	Frequency	Cumulative Frequency	Percent
No	44,052	44,052	28.31
Yes	1,238	45,290	0.80
Missing	110,336	155,626	70.90

128	Column name:	diag_diab
------------	---------------------	------------------

Column description: Diagnosis: Diabetes/hyperglycaemia

Data type: Factor

Unique non-missing value count: 2

Missing value count: 6,597

Categories	Frequency	Cumulative Frequency	Percent
No	142,656	142,656	91.67
Yes	6,373	149,029	4.10
Missing	6,597	155,626	4.24

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

Unique non-missing value count: 2

Missing value count: 116,756

Categories	Frequency	Cumulative Frequency	Percent
No	37,833	37,833	24.31
Yes	1,037	38,870	0.67
Missing	116,756	155,626	75.02

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

Unique non-missing value count: 2

Missing value count: 19,732

Categories	Frequency	Cumulative Frequency	Percent
No	131,622	131,622	84.58
Yes	4,272	135,894	2.75
Missing	19,732	155,626	12.68

131	Column name:	diag_tb
------------	---------------------	----------------

Column description: Diagnosis: Tuberculosis

Data type: Factor

Unique non-missing value count: 2

Missing value count: 13,322

Categories	Frequency	Cumulative Frequency	Percent
No	136,203	136,203	87.52
Yes	6,101	142,304	3.92
Missing	13,322	155,626	8.56

132	Column name:	diag_cancer
------------	---------------------	--------------------

Column description: Diagnosis: Cancer

Data type: Factor

Unique non-missing value count: 2

Missing value count: 25,878

Categories	Frequency	Cumulative Frequency	Percent
No	128,821	128,821	82.78
Yes	927	129,748	0.60
Missing	25,878	155,626	16.63

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
Unique non-missing value count:	2
Missing value count:	58,385

Categories	Frequency	Cumulative Frequency	Percent
No	95,124	95,124	61.12
Yes	2,117	97,241	1.36
Missing	58,385	155,626	37.52

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
Unique non-missing value count:	2
Missing value count:	6,574

Categories	Frequency	Cumulative Frequency	Percent
No	131,270	131,270	84.35
Yes	17,782	149,052	11.43
Missing	6,574	155,626	4.22

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

Unique non-missing value count: 2

Missing value count: 15,939

Categories	Frequency	Cumulative Frequency	Percent
No	134,702	134,702	86.55
Yes	4,985	139,687	3.20
Missing	15,939	155,626	10.24

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

Unique non-missing value count: 2

Missing value count: 119,765

Categories	Frequency	Cumulative Frequency	Percent
No	35,673	35,673	22.92
Yes	188	35,861	0.12
Missing	119,765	155,626	76.96

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

Unique non-missing value count: 2

Missing value count: 127,449

Categories	Frequency	Cumulative Frequency	Percent
No	27,772	27,772	17.85
Yes	405	28,177	0.26
Missing	127,449	155,626	81.89

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

Unique non-missing value count: 2

Missing value count: 29,598

Categories	Frequency	Cumulative Frequency	Percent
No	123,289	123,289	79.22
Yes	2,739	126,028	1.76
Missing	29,598	155,626	19.02

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

Unique non-missing value count: 2

Missing value count: 36,619

Categories	Frequency	Cumulative Frequency	Percent
------------	-----------	----------------------	---------

No	117,691	117,691	75.62
Yes	1,316	119,007	0.85
Missing	36,619	155,626	23.53

140	Column name:	strokedmed
-----	--------------	------------

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

Unique non-missing value count: 2

Missing value count: 29,758

Categories	Frequency	Cumulative Frequency	Percent
No	125,242	125,242	80.48
Yes	626	125,868	0.40
Missing	29,758	155,626	19.12

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

Unique non-missing value count: 2

Missing value count: 123,392

Categories	Frequency	Cumulative Frequency	Percent
No	31,245	31,245	20.08
Yes	989	32,234	0.64
Missing	123,392	155,626	79.29

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
Unique non-missing value count:	2
Missing value count:	123,392

Categories	Frequency	Cumulative Frequency	Percent
No	31,670	31,670	20.35
Yes	564	32,234	0.36
Missing	123,392	155,626	79.29

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
Unique non-missing value count:	2
Missing value count:	123,392

Categories	Frequency	Cumulative Frequency	Percent
No	32,056	32,056	20.60
Yes	178	32,234	0.11
Missing	123,392	155,626	79.29

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

Unique non-missing value count: 2

Missing value count: 123,392

Categories	Frequency	Cumulative Frequency	Percent
No	32,157	32,157	20.66
Yes	77	32,234	0.05
Missing	123,392	155,626	79.29

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

Unique non-missing value count: 2

Missing value count: 123,392

Categories	Frequency	Cumulative Frequency	Percent
No	31,734	31,734	20.39
Yes	500	32,234	0.32
Missing	123,392	155,626	79.29

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

Unique non-missing value count: 2

Missing value count: 123,392

Categories	Frequency	Cumulative Frequency	Percent
No	32,152	32,152	20.66
Yes	82	32,234	0.05
Missing	123,392	155,626	79.29

147	Column name:	strokedmed_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
Unique non-missing value count:	2
Missing value count:	123,392

Categories	Frequency	Cumulative Frequency	Percent
No	32,233	32,233	20.71
Yes	1	32,234	0.00
Missing	123,392	155,626	79.29

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
Unique non-missing value count:	18
Missing value count:	104,051

Min	Mean	Median	Max	SD
0.00	3.02	2.00	17.00	2.34

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
Unique non-missing value count:	2

Missing value count: 83,654

Categories	Frequency	Cumulative Frequency	Percent
No/don't know	69,739	69,739	44.81
Yes	2,233	71,972	1.43
Missing	83,654	155,626	53.75

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

Unique non-missing value count: 2

Missing value count: 89,586

Categories	Frequency	Cumulative Frequency	Percent
No	14,406	14,406	9.26
Yes	51,634	66,040	33.18
Missing	89,586	155,626	57.56

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

Unique non-missing value count: 1,169

Missing value count: 20,939

Min	Mean	Median	Max	SD
120.00	162.06	161.50	207.40	9.06

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
Unique non-missing value count:	732
Missing value count:	56,563

Min	Mean	Median	Max	SD
120.00	162.28	161.60	207.40	8.87

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
Unique non-missing value count:	264
Missing value count:	154,818

Min	Mean	Median	Max	SD
122.40	161.83	161.00	207.10	9.42

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
Unique non-missing value count:	1,480
Missing value count:	20,882

Min	Mean	Median	Max	SD
120.00	162.07	161.50	207.40	9.06

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			
Unique non-missing value count:	2,065			
Missing value count:	21,112			
Min	Mean	Median	Max	SD
26.90	68.95	65.80	210.00	17.65

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			
	Unique non-missing value count:	1,182			
	Missing value count:	62,874			
	Min	Mean	Median	Max	SD
	28.40	69.12	66.00	150.60	17.49

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			
	Unique non-missing value count:	282			
	Missing value count:	155,208			
	Min	Mean	Median	Max	SD
	33.00	71.03	66.55	136.20	18.92

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			
	Unique non-missing value count:	2,552			
	Missing value count:	21,047			
	Min	Mean	Median	Max	SD
	26.90	68.95	65.75	210.00	17.65
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			
	Unique non-missing value count:	1,928			
	Missing value count:	20,950			
	Min	Mean	Median	Max	SD
	30.00	86.12	84.00	218.00	17.56
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			
	Unique non-missing value count:	1,234			
	Missing value count:	56,941			
	Min	Mean	Median	Max	SD
	30.00	86.71	84.60	218.00	17.85
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
Unique non-missing value count:	361
Missing value count:	154,732

Min	Mean	Median	Max	SD
33.00	89.12	87.00	201.00	17.42

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
Unique non-missing value count:	2,457
Missing value count:	20,891

Min	Mean	Median	Max	SD
30.00	86.14	84.00	218.00	17.56

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
Unique non-missing value count:	547
Missing value count:	127,789

Min	Mean	Median	Max	SD
6.40	29.13	28.50	50.00	5.02

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
Unique non-missing value count:	311
Missing value count:	149,750

Min	Mean	Median	Max	SD
6.40	29.14	28.40	50.00	5.65

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
Unique non-missing value count:	180
Missing value count:	155,026

Min	Mean	Median	Max	SD
6.00	30.73	30.25	50.00	6.64

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			
Unique non-missing value count:	584			
Missing value count:	127,774			
Min	Mean	Median	Max	SD
6.40	29.13	28.50	50.00	5.02

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			
Unique non-missing value count:	1,495			
Missing value count:	122,214			
Min	Mean	Median	Max	SD
40.00	99.61	98.00	230.00	16.38

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			
Unique non-missing value count:	677			
Missing value count:	149,725			
Min	Mean	Median	Max	SD
40.00	100.15	98.00	192.30	15.32

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			
Unique non-missing value count:	269			
Missing value count:	155,003			
Min	Mean	Median	Max	SD
59.40	101.72	99.50	187.00	16.23

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			
	Unique non-missing value count:	1,396			
	Missing value count:	122,202			
	Min	Mean	Median	Max	SD
	40.00	99.61	98.00	230.00	16.38

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			
	Unique non-missing value count:	192			
	Missing value count:	21,337			
	Min	Mean	Median	Max	SD
	60.00	126.22	122.00	267.00	23.32
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
Unique non-missing value count:	190
Missing value count:	21,882

Min	Mean	Median	Max	SD
60.00	123.97	120.00	263.00	22.71

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
Unique non-missing value count:	187
Missing value count:	115,466

Min	Mean	Median	Max	SD
60.00	127.10	123.00	256.00	23.63

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
Unique non-missing value count:	676
Missing value count:	20,734

Min	Mean	Median	Max	SD
63.00	124.75	121.00	260.33	22.32

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			
	Unique non-missing value count:	358			
	Missing value count:	21,542			
	Min	Mean	Median	Max	SD
	60.00	123.69	120.00	257.00	22.46

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			
	Unique non-missing value count:	121			
	Missing value count:	21,527			
	Min	Mean	Median	Max	SD
	30.00	80.81	79.00	150.00	14.36

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
	Unique non-missing value count:	121

Missing value count:		22,012		
Min	Mean	Median	Max	SD
30.00	79.47	78.00	150.00	14.11

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		
Unique non-missing value count:		121		
Missing value count:		115,510		
Min	Mean	Median	Max	SD
30.00	79.30	78.00	150.00	15.07

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		
Unique non-missing value count:		445		
Missing value count:		20,784		
Min	Mean	Median	Max	SD
30.00	79.99	78.50	150.00	13.68

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			
Unique non-missing value count:	231			
Missing value count:	21,609			
Min	Mean	Median	Max	SD
30.00	79.33	78.00	150.00	13.97

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			
Unique non-missing value count:	149			
Missing value count:	20,317			
Min	Mean	Median	Max	SD
21.00	76.40	76.00	189.00	13.54

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			
Unique non-missing value count:	153			
Missing value count:	20,729			
Min	Mean	Median	Max	SD
22.00	76.10	75.00	186.00	13.38

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			
Unique non-missing value count:	175			
Missing value count:	118,050			
Min	Mean	Median	Max	SD
21.00	87.06	78.00	250.00	32.25

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			
	Unique non-missing value count:	484			
	Missing value count:	20,161			
	Min	Mean	Median	Max	SD
	24.00	77.19	76.00	177.00	14.31

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			
	Unique non-missing value count:	305			
	Missing value count:	20,649			
	Min	Mean	Median	Max	SD
	25.00	77.53	76.00	188.50	15.81

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			
Unique non-missing value count:	96,113			
Missing value count:	22,307			
Min	Mean	Median	Max	SD
10.35	26.34	24.88	103.14	6.95

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/m ² < BMI < 40 kg/m ² . Obesity III = 40 kg/m ² < BMI	
	Data type:	Factor	
	Unique non-missing value count:	7	
	Missing value count:	22,307	
Categories	Frequency	Cumulative Frequency	Percent
Healthy weight	57,218	57,218	36.77
Obesity I	18,940	76,158	12.17
Obesity II	9,464	85,622	6.08
Obesity III	3,796	89,418	2.44
Obesity IV	2,191	91,609	1.41
Overweight	31,357	122,966	20.15
Underweight	10,353	133,319	6.65
Missing	22,307	155,626	14.33

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			
	Unique non-missing value count:	238			
	Missing value count:	144,134			
	Min	Mean	Median	Max	SD
	5.00	13.48	13.50	19.90	1.98

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 : $\text{mmol/mol} = 10.93 * \% - 23.5$ mmol/mol.			
	Data type:	Numeric			
	Unique non-missing value count:	225			
	Missing value count:	145,015			
	Min	Mean	Median	Max	SD
	14.75	43.05	40.98	189.99	13.19

190	Column name:	chol_tot			
	Column description:	Total cholesterol [mmol/l]			
	Source information:	Implausible values were considered to be those with total cholesterol <1.75 mmol/L, >20 mmol/L, or where total cholesterol was lower than HDL cholesterol, 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:	78			
Missing value count:	151,175			
Min	Mean	Median	Max	SD
1.90	4.47	4.30	11.80	1.13

191	Column name:	chol_hdl		
	Column description:	High-density lipoprotein (HDL) cholesterol [mmol/l]		
	Source information:	Implausible values were considered to be those with HDL cholesterol <0.40 mmol/L, >5.00 mmol/L, or where total cholesterol was lower than HDL cholesterol 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:	239		
	Missing value count:	151,201		
Min	Mean	Median	Max	SD
0.40	1.28	1.20	4.00	0.41

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
	Unique non-missing value count:	56
	Missing value count:	153,893

Min	Mean	Median	Max	SD
0.10	2.54	2.40	6.10	0.89

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

Unique non-missing value count: 435

Missing value count: 151,191

Min	Mean	Median	Max	SD
0.15	1.35	1.09	14.22	0.99

194	Column name:	medaid
-----	--------------	--------

Column description: Covered by medical insurance

Data type: Factor

Unique non-missing value count: 2

Missing value count: 12,633

Categories	Frequency	Cumulative Frequency	Percent
No	126,128	126,128	81.05
Yes	16,865	142,993	10.84
Missing	12,633	155,626	8.12

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

Unique non-missing value count: 2

Missing value count: 55,424

Categories	Frequency	Cumulative Frequency	Percent
No	53,613	53,613	34.45
Yes	46,589	100,202	29.94
Missing	55,424	155,626	35.61

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

Unique non-missing value count: 2

Missing value count: 33,305

Categories	Frequency	Cumulative Frequency	Percent
No	89,542	89,542	57.54
Yes	32,779	122,321	21.06
Missing	33,305	155,626	21.40

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

Unique non-missing value count: 2

Missing value count: 26,214

Categories	Frequency	Cumulative Frequency	Percent
No	107,079	107,079	68.81
Yes	22,333	129,412	14.35
Missing	26,214	155,626	16.84

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

Unique non-missing value count: 2

Missing value count: 37,727

Categories	Frequency	Cumulative Frequency	Percent
No	107,999	107,999	69.40
Yes	9,900	117,899	6.36
Missing	37,727	155,626	24.24

199	Column name:	ohcare1mo
-----	--------------	-----------

Column description: Outpatient consultations last month

Data type: Factor

Unique non-missing value count: 2

Missing value count: 145,293

Categories	Frequency	Cumulative Frequency	Percent
No	9,053	9,053	5.82
Yes	1,280	10,333	0.82

Missing	145,293	155,626	93.36
---------	---------	---------	-------

200	Column name:	ohcare1mo_public
-----	--------------	------------------

Column description:	Outpatient healthcare last month: public hospital/clinic
---------------------	--

Data type:	Factor
------------	--------

Unique non-missing value count:	2
---------------------------------	---

Missing value count:	145,293
----------------------	---------

Categories	Frequency	Cumulative Frequency	Percent
No	9,398	9,398	6.04
Yes	935	10,333	0.60
Missing	145,293	155,626	93.36

201	Column name:	ohcare1mo_private
-----	--------------	-------------------

Column description:	Outpatient healthcare last month: private hospital/clinic/doctor
---------------------	--

Data type:	Factor
------------	--------

Unique non-missing value count:	2
---------------------------------	---

Missing value count:	145,293
----------------------	---------

Categories	Frequency	Cumulative Frequency	Percent
No	10,041	10,041	6.45
Yes	292	10,333	0.19
Missing	145,293	155,626	93.36

202	Column name:	hcare1mo_chem_nurse
-----	--------------	---------------------

Column description:	Healthcare last month: chemist/pharmacist/nurse
---------------------	---

Data type:	Factor
------------	--------

Unique non-missing value count:	2
---------------------------------	---

Missing value count:	47,953
----------------------	--------

Categories	Frequency	Cumulative Frequency	Percent
No	106,188	106,188	68.23

Yes	1,485	107,673	0.95
Missing	47,953	155,626	30.81

203	Column name:	hcare1mo_trad
------------	---------------------	----------------------

Column description: Healthcare last month: traditional/faith healer

Data type: Factor

Unique non-missing value count: 2

Missing value count: 48,861

Categories	Frequency	Cumulative Frequency	Percent
No	105,661	105,661	67.89
Yes	1,104	106,765	0.71
Missing	48,861	155,626	31.40

204	Column name:	hcare1mo_other
------------	---------------------	-----------------------

Column description: Healthcare last month: other

Data type: Factor

Unique non-missing value count: 2

Missing value count: 104,974

Categories	Frequency	Cumulative Frequency	Percent
No	48,861	48,861	31.40
Yes	1,791	50,652	1.15
Missing	104,974	155,626	67.45

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

Unique non-missing value count: 50,204

Missing value count: 105,397

Min	Mean	Median	Max	SD
0.29	12.23	8.45	87.34	11.34

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

Unique non-missing value count: 1,797

Missing value count: 153,823

Min	Mean	Median	Max	SD
0.46	14.66	9.87	91.23	13.86

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

Unique non-missing value count: 1,797

Missing value count: 153,823

Min	Mean	Median	Max	SD
0.09	8.18	3.62	87.78	11.25

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](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: 52,590

Missing value count: 102,972

Min	Mean	Median	Max	SD
0.40	7.10	5.21	53.71	5.84

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](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:	1,879			
Missing value count:	153,741			
Min	Mean	Median	Max	SD
0.82	8.09	5.98	47.50	6.66

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		
Unique non-missing value count:		2,899		
Missing value count:		85,225		
Min	Mean	Median	Max	SD
1.00	10.21	6.55	30.00	9.30