Supplementary appendix

Table of Contents

Supplementary Materials Section 1 – Methods.

Population and setting

The surveys employed a multi-stage, stratified cluster sampling design stratified by province, locality/geographic type, and race, covering the whole of South Africa. The surveys are representative of the entire country as they are based on randomly selecting 1000 primary sampling units (PSU) from a national population sampling frame developed by Statistics South Africa and then randomly selecting secondary sampling units (SSU) comprising 15 visiting points (households) per PSU. After obtaining informed consent/assent, data was collected through face-to-face interviews by trained fieldwork staff using structured questionnaires. HIV status is based on laboratory testing of blood samples in all the survey rounds, and so is ARV use (2012 and 2017 only) among HIV-positive samples.

Data measurements

Instructions were provided to the respondents during the survey interviews that drugs prescribed for medication by a medical doctor, such as amphetamines, sedatives, or pain medications, were not recorded unless they had used "such medications for reasons other than prescription or had taken them more frequently or at higher doses than prescribed". We note, however, that this instruction was not provided in the 2017 questionnaire. We undertook a sensitivity analysis to assess if analysis results differ if we exclude individuals from the 2017 survey who were only using drugs that could potentially be prescribed and not using any other illicit drug.

Even though the survey included a question on injection drug use ("Besides drugs prescribed by a health professional, have you ever used a drug by injection?"), there were validity concerns related to injection drug use responses, as only 6.4% of those reporting recent injection drug use also reported any recent drug use in the pooled dataset. Possibly, the respondents may not have distinguished between drugs injected for medical treatment and illicit drug injection, and therefore, this question was not used.

Statistical analyses

The education attainment variable had 11% and 12% missing values in the 2012 and 2017 survey waves, respectively, due to differences in how this question was administered in those years. Consequently, we imputed missing data for the education attainment variable separately for the 2012 and 2017 datasets. Conditional multiple imputation (assuming data was missing completely at random) was used with a random forest regression model used to predict education level (assuming ordered variable with five categories) with age (treated as a continuous variable), sex, race, and the province as independent variables in the model. Details of this method have been described elsewhere (Ramosaj & Pauly, 2019) (see **Supplementary Material Section 4** – missing data imputation, for additional information).

Supplementary Materials Section 2 – Survey questions used to construct some of the analysis variables.

Supplementary Table S1. Recent drug use

Instruction	A doctor may prescribe some of the substances listed below (e.g., amphetamines, sedatives, pain medications). For this interview, we will not record medications that are prescribed by your doctor. However, if you have taken such medications for reasons other than prescription, or have taken them more frequently or at higher doses than prescribed, please let me know. While we are also interested in knowing about your use of various recreational drugs, please be assured that information on such use will be treated as strictly confidential."				
In the past three months, how often have you used any of the following substances?	**Never**	**Once or twice**	**Monthly**	**Weekly**	**Almost daily**
Cannabis (dagga, marijuana, pot, grass, hash, etc.)					
Cocaine (coke, rocks, crack, etc.)					
Amphetamine-type stimulants (speed, ecstasy, tik, etc.)					
Inhalants (nitrates, glue, petrol, paint thinners, etc.)					
Sedatives or sleeping pills (Valium, Mandrax, Serepax, Rohypnol, etc.)					
Hallucinogens (LSD, acid, mushrooms, PCP, Special K, etc.)					
Opiates (heroin, morphine, methadone, codeine, etc.)					
Whoonga (mixture of heroin, dagga and ARVs), Nyaope					
Other					

"*Whoonga*" was not available as an option to the question in the 2002, 2005, and 2008 surveys, while "*Inhalants*"

was not available as an option in the 2002 survey.

The recent drug use variable was constructed based on those who responded to having used any of the drugs in the

past three months, i.e., "*once or twice*", "*monthly*", "*weekly*", or "*almost daily*".

Supplementary Table S2. Hazardous alcohol use and alcohol dependence

Hazardous alcohol use and alcohol dependence were measured using the 'Alcohol Disorder Identification Test

(AUDIT)' screening tool(World Health Organization, 2001). The following questions were used:

The Alcohol Use Disorders Identification Test	
1. How often did you have a drink containing alcohol in the past 12 months? (0) Not in the past 12 months (1) Once a month or less (2) 2 to 4 times a month (3) 2 to 3 times a week (4) 4 or more times a week	6. How often during the past 12 months did you need a first drink in the morning to get yourself going after a heavy drinking session? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily
2. How many drinks containing alcohol do you have on a typical day when you are drinking? (0) 1 or 2	7. How often during the past 12 months did you feel guilt or remorse after drinking? (0) Never

(1) 3 or 4 (2) 5 or 6 (3) 7 to 9 (4) 10 or more	(1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily
3. How often do you have (*for men*) five or more and (*for women*) four or more drinks on one occasion? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily	8. How often during the past 12 months were you unable to remember what happened the night before because of your drinking? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily
4. How often during the past 12 months were you not able to stop drinking once you had started? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily	9. Have you or someone else been injured as a result of your drinking? (0) No (2) Yes, but not in the past 12 months (4) Yes, during the past 12 months
5. How often during the past 12 months did you fail to do what was normally expected from you because of drinking? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily	10. Has a concerned relative, friend or a doctor or health worker ever suggested that you should cut down on your drinking? (0) No (2) Yes, but not in the past 12 months (4) Yes, during the past 12 months

The AUDIT score, the sum of the above 10-item questions, ranges from 0-40. Hazardous alcohol use was based on an AUDIT score ≥ 8, while alcohol dependence was based on an AUDIT score ≥ 15.

Supplementary Table S3. Psychological distress

Psychological distress was measured on a 10-item scale based on the Kessler Psychological Distress Scale [K-10] (Kessler et al., 2003). The following questions were asked on psychological distress.

Kessler Psychological Distress Test	
1. During the last 30 days, about how often did you feel tired out for no good reason? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time	6. About how often did you feel so restless you could not sit still? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time
2. During the last 30 days, about how often did you feel nervous? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time	7. About how often did you feel depressed? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time

3. About how often did you feel so nervous that nothing could calm you down? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time	8. During the last 30 days, about how often did you feel that everything was an effort? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time
4. About how often did you feel hopeless? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time	9. About how often did you feel so sad that nothing could cheer you up? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time
5. During the last 30 days, about how often did you feel restless or fidgety? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time	10. About how often did you feel worthless? (1) None of the time (2) A little of the time (3) Some of the time (4) Most of the time (5) All of the time

The Kessler Psychological Distress score, sum of the above 10-item questions, ranges from 0-50. Psychological Distress was based on score ≥ 22.

Supplementary Table S4. Intimate partner violence

Intimate partner violence (IPV) was created based on a "Yes" response to any of the eleven questions related to physical, sexual, and emotional abuse. The questions were as follows:

Intimate partner violence	
1. Did your partner ever - Push you, shake you, or throw something at you (1) Yes (2) No	7. Did your partner ever - Threaten or attack you with a knife, gun, or other weapon (1) Yes (2) No
2. Did your partner ever - Slap you (1) Yes (2) No	8. Did your partner ever - Physically force you to have sexual intercourse with him/her (1) Yes (2) No
3. Did your partner ever - Twist your arm or pull your hair (1) Yes (2) No	9. Did your partner ever - Physically force you to perform any other sexual acts you did not want to (1) Yes (2) No
4. Did your partner ever - Punch you with his fist or with something (1) Yes (2) No	10. Did your partner ever - Force you with threats or in any other way (1) Yes (2) No
5. Did your partner ever - Kick you, drag you, or beat you up (1) Yes (2) No	11. Did your partner ever - Perform sexual acts you did not want to (1) Yes (2) No
6. Did your partner ever - Try to choke you or burn you on purpose	

(1) Yes (2) No	

Supplementary Materials Section 3 – Survey response rates

The individual interview response rates were generally high ranging from 73.7% in 2002 to 93.6% in 2017, while the household and individual interview rates combined ranged from 52.4% in 2002 to 76.9% in 2017. It has been noted that surveys with a response rate of 50% are considered adequate, 60% are good, while 75% are considered very good (Babbie, 2020; Shisana O & Simbayi L, Nelson Mandela/HSRC study of HIV/AIDS:, 2004). From the response rates in Supplementary Table S5, most surveys had very good response rates.

Supplementary Table S5. Household (visiting points) and individual interview response rates

| Survey wave	Household (visiting point) response rate	Individual interview response rate	Household and individual interview response rate*
2002	71.1%	73.7%	52.4%
2005	84.1%	96.0%	80.7%
2008	80.8%	89.1%	72.0%
2012	84.7%	89.5%	75.8%
2017	82.2%	93.6%	76.9%

*The response rate is a product of the household response rate and individual interview response rate.

Supplementary Materials Section 4 – Missing data imputation.

We summarise the method used to impute missing education values for the 2012 and 2017 survey waves. There were 11% and 12% missing values for the highest education attained in 2012 and 2017, respectively, due to differences in how this question was administered in those years. We imputed missing data using only data from the 2012 and 2017 survey waves, undertaking imputation separately for each dataset. For example, we only used the 2012 dataset to impute missing data for the education variable in the 2012 survey wave; similarly, so for the 2017 survey. We assumed data were missing completely at random and used a random forest regression model to determine education level with age, sex, race, and the province as factors associated with education level. These variables were chosen because South African census data has shown differences in educational attainment by population group, province, and slightly by sex (Statistics South Africa, 2011). Data from the 2012 and 2017 survey waves showed that age correlated with education attainment; hence, it was also considered a factor in the imputation model. We, therefore, included age, sex, race, and province in our imputation model. Random forest regression was chosen since it is robust to the non-linearity of data, accepts mixed continuous and/or categorical data, can handle outliers, and has high predictive accuracy (Hong & Lynn, 2020). Some studies have shown random forest performs better than standard imputation methods (Ramosaj & Pauly, 2019; Tang & Ishwaran, 2017). In brief, using this approach, we pre-impute the data, grow a forest, in turn, using the variable with the missing values, determine the missing value using the grown forest, and iterate for improved results (Tang & Ishwaran, 2017). We tried both a

polynomial regression and random forest to impute the missing education variable. We demonstrate the imputation process using the 2017 data in the chart and table below. Supplementary Figure S1 shows no major differences in polynomial and random forest regression models in fitting the observed missing data pattern (left panel vs. right panel, respectively). The weighted prevalence estimates of recent drug use were the same for either complete case or imputed datasets (Supplementary Table S6). However, we emphasise that imputing the missing education attainment variable was ideal for purposes of analysis, such as for Aim 2, where a multivariable regression model was run, thus ensuring sufficient sample size in the model.

A. Imputation: Using polynomial regression
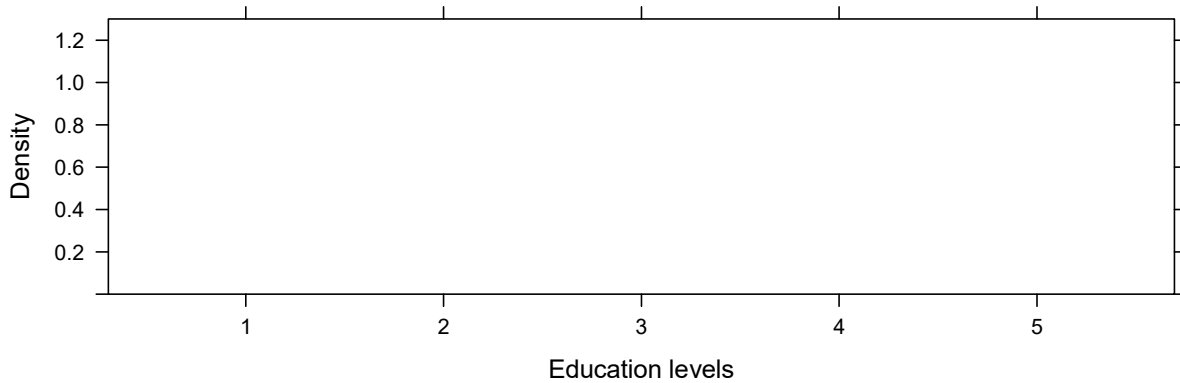
B. Imputation: Using random forest regression
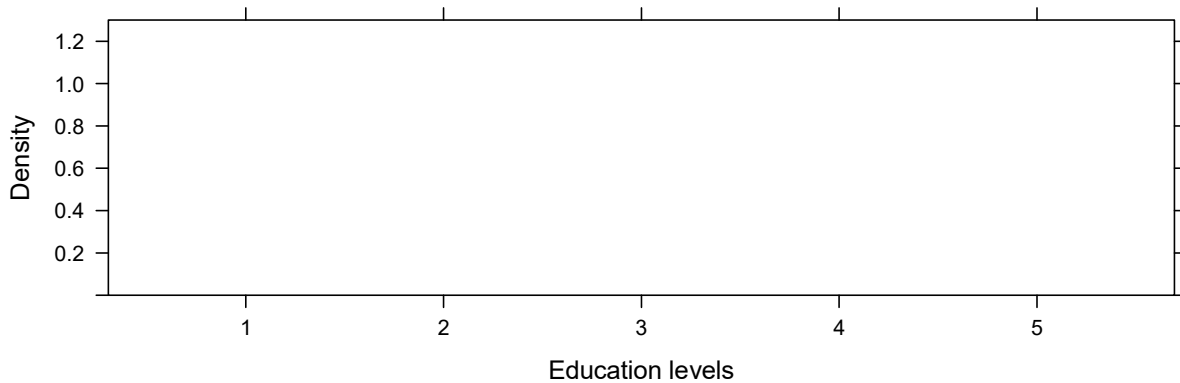
Supplementary Figure S1: Multiple imputation of missing education attainment variable, 2017
Figure S3A is based on a standard regression model (predicting missing values using polynomial regression), while Figure S3B is based on a random forest regression model. The blue line is the density for the complete cases, while the red lines are the densities based on the imputed data. The random forest regression tends to fit much closer to the observed data.

Supplementary Table S6: Comparison of prevalence of any recent drug use by educational status in 2017 (weighted prevalence for complete case vs. imputed data analysis).

	Complete case analysis	Imputed analysis (standard method) [b]	Imputed analysis (random-forest method) [c]
Education status [a]			
Grade 0-7	8.4 (7.4-9.6)	8.4 (7.4-9.6)	8.4 (7.4-9.6)
Grade 8-11	11.9 (10.8-13.0)	11.9 (10.8-13.0)	11.9 (10.8-13.0)
Grade 12 or more	9.3 (8.3-10.6)	9.3 (8.3-10.6)	9.3 (8.3-10.6)

[a] Highest educational level obtained.
[b] Missing education attainment variable modelled using a polynomial regression model
[c] Missing education attainment variable modelled using a random forest regression model.

Supplementary Materials Section 5 – Additional Results.

Supplementary Figure S2. Multicollinearity assessment (variance inflation factors) for the model assessing factors associated with any recent drug use.
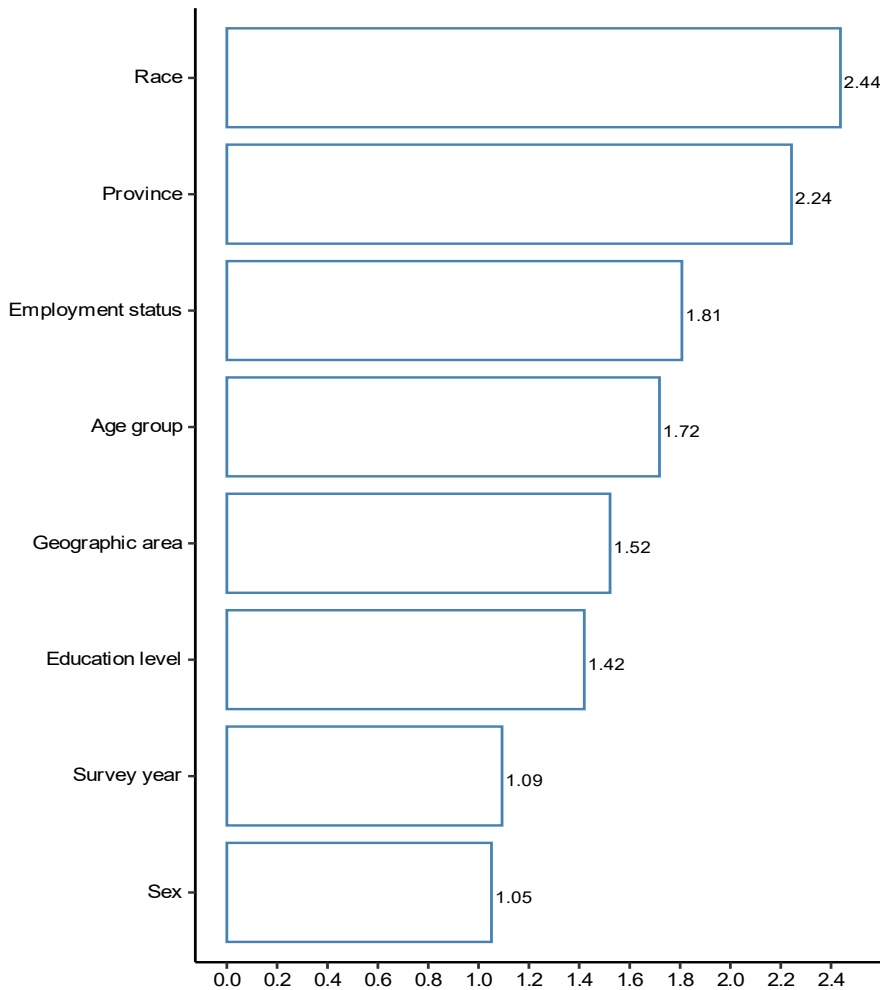
Variance inflation factors for the model assessing factors associated with any recent drug use for the multivariable model in Table 3.

Supplementary Table S7. Behavioural and outcome characteristics of the survey participants (individuals aged 15 years and older) across the years (2002, 2005, 2008, 2012 and 2017 SABSSM surveys).

Characteristics	2002		2005		2008		2012		2017		2002-2017	
	n	Weighted percent (%)	n	Weighted percent (%)	n	Weighted percent (%)	n	Weighted percent (%)	n	Weighted percent (%)	n	Weighted percent (%)
All (N)	7084	100.0	16398	100.0	13828	100.0	26807	100.0	24996	100.0.	89113	100.0
Behavioral and outcome characteristics **Hazardous or harmful alcohol use (AUDIT score ≥8)** [a]												
No	NA	NA	13932	92.3	10771	90.4	21036	88.2	19513	89.0	65252	89.9
Yes	NA	NA	1082	7.7	1188	9.6	2497	11.8	2118	11.0	6885	10.1
Alcohol dependence (AUDIT score ≥15) [a]												
No	NA	NA	14701	98.0	11656	97.4	22790	96.5	20980	96.6	70127	97.1
Yes	NA	NA	313	2.0	303	2.6	743	3.5	651	3.4	2010	2.9
Multiple sexual partners												
None or 1 sexual partner	6653	94.1	14857	94.1	12202	93.9	23920	91.8	20702	93.5	78334	93.4
2 or more sexual partners	379	5.9	680	5.9	652	6.1	1482	8.2	1073	6.5	4266	6.6
Ever experienced psychological distress (≥22) [b] (only available in 2012 and 2017 surveys)												
No	NA	NA	NA	NA	NA	Na	21787	82.4	20113	85.8	41900	84.1
Yes	NA	NA	NA	NA	NA	NA	4088	17.6	3414	14.2	7502	15.9
Sexual debut before age 15 years (among youth aged 15-24 years old) [γ]												
<15 years	122	8.4	248	8.4	194	8	393	10.6	390	13.6	1347	9.9
≥15 years	1199	91.6	2662	91.6	2141	92	3543	89.4	2803	86.4	12348	90.1
Age-disparate relationships (among women aged 15-24 years old) [c, γ]												
<5 years	351	59.2	870	63.3	700	68.4	1118	64.3	761	60.0	3800	63.2
5+ years	212	40.8	525	36.7	364	31.6	649	35.7	534	40.0	2284	36.8

Condom use at last sex with most recent sexual partner												
No	3302	73.1	6348	64.8	5049	59.0	11118	64.6	7976	61.8	33793	64.5
Yes	1241	26.9	2965	35.2	2907	41.0	4725	35.4	4186	38.2	16024	35.5
Ever been tested for HIV												
Yes	1658	21.2	4937	30.5	6352	50.7	16619	65.4	16939	74.1	46505	50.5
No	5360	78.2	11175	69.3	6732	49.2	9762	34.4	6251	24.4	39280	49.0
HIV status												
HIV positive	719	13.6	1351	14.0	1302	14.3	2632	16.4	2819	18.8	8823	15.7
HIV negative	5361	86.4	10681	86.0	9506	85.7	18075	83.6	14420	81.2	58043	84.3
Antiretroviral therapy status (only available in 2012 and 2017 surveys)												
Not on ART	NA	NA	NA	NA	NA	NA	1810	69.7	899	37.3	2709	52.5
On ART	NA	NA	NA	NA	NA	NA	820	30.3	1602	62.7	2422	47.5
Ever experienced intimate partner violence [d] (only available in 2017 survey)												
No	NA	NA	NA	NA	NA	NA	NA	NA	4624	83.2	4624	83.2
Yes	NA	NA	NA	NA	NA	NA	NA	NA	901	16.8	901	16.8

[a] Based on a 10-item 'Alcohol Disorder Identification Test (AUDIT)' score. A score of 8 or more indicated hazardous or harmful drinking. A score of 15 or more indicates dependent alcohol drinking (moderate-severe alcohol use disorder).

[b] Psychological distress is measured based on 10 variables where each question has a scale of 1-5 (individuals with a score $\geq$22 are considered psychologically distressed).

[c] Age-disparate relationships involving a sexual partner more than five years older among women aged 15-24 years old.

[d] Experience of intimate partner violence among male and female respondents.

[γ] Analysis was done only among youth aged 15-24 years old.

Supplementary Table S8: Prevalence of any recent drug use for various drugs across the years (2002, 2005, 2008, 2012 and 2017 SABSSM surveys).

Type of drug used:	2002		2005		2008		2012		2017	
	n/N	Weighted percent (95% CI)	n/N	Weighted percent (95% CI)	n/N	Weighted percent (95% CI)	n/N	Weighted percent (95% CI)	n/N	Weighted percent (95% CI)
Cannabis	117 / 7055	1.48 (1.10-1.99)	265 / 16145	2.07 (1.70-2.51)	424 / 13122	3.32 (2.84-3.89)	941 / 26404	4.04 (3.60-4.53)	1728 / 23575	7.78 (7.16-8.46)
Cocaine	3 / 7055	0.02 (0.00-0.06)	35 / 16124	0.3 (0.18-0.49)	63 / 13113	0.61 (0.41-0.92)	101 / 26401	0.33 (0.23-0.47)	440 / 23526	1.76 (1.47-2.11)
Amphetamine	11 / 7055	0.10 (0.04-0.24)	37 / 16126	0.18 (0.10-0.31)	88 / 13113	0.73 (0.51-1.04)	102 / 26405	0.3 (0.22-0.40)	380 / 23529	1.45 (1.2-1.77)
Inhalants	NA / 0	0	9 / 16117	0.10 (0.04-0.25)	57 / 13111	0.5 (0.32-0.80)	44 / 26400	0.18 (0.11-0.29)	340 / 23530	1.32 (1.07-1.62)
Sedatives	13 / 7055	0.07 (0.03-0.17)	63 / 16126	0.27 (0.18-0.41)	110 / 13112	0.84 (0.60-1.16)	144 / 26402	0.40 (0.29-0.55)	459 / 23522	1.71 (1.43-2.03)
Hallucinogens	NA / 0	0	12 / 16120	0.12 (0.06-0.26)	54 / 13112	0.52 (0.33-0.82)	39 / 26402	0.14 (0.07-0.28)	325 / 23514	1.24 (0.99-1.54)
Opioids†	1 / 7055	0.01 (0.00-0.05)	14 / 16109	0.08 (0.03-0.21)	56 / 13107	0.52 (0.33-0.81)	82 / 26807	0.28 (0.19-0.41)	416 / 24996	1.61 (1.35-1.93)
Other drugs	4 / 7055	0.01 (0.00-0.05)	31 / 10719	0.43 (0.26-0.70)	122 / 11540	1.21 (0.90-1.62)	456 / 25012	1.62 (1.31-2.01)	555 / 22906	2.73 (2.23-3.32)
Any recent drug use	**127 / 7055**	**1.50 (1.10-2.00)**	**363 / 16164**	**2.60 (2.20-3.00)**	**560 / 13128**	**4.3 (3.7-4.9)**	**1508 / 26425**	**5.9 (5.4-6.5)**	**2184 / 23590**	**10 (9.3-10.9)**

† Includes Opiates and Whoonga (local name for a mixture of opiates and cannabis).
CI – confidence interval.
NA – data on drug type was not collected in the survey round.

Supplementary Table S9: Distribution of number of drugs used among those who were using any recent drugs in the pooled 2002-2017 SABSSM surveys.

Number of drugs used	2002		2005		2008		2012		2017		2002-2017 (Combined data)	
	n	Weighted percent (95% CI)	n	Weighted percent (95% CI)	n	Weighted percent (95% CI)	n	Weighted percent (95% CI)	n	Weighted percent (95% CI)	n	Weighted percent (95% CI)
Used 1 drug	110	89.8 (80.8-94.9)	304	84.5 (78.1-89.3)	458	79.7 (73.6-84.8)	1306	88.8 (85.8-91.2)	1635	76.7 (73.5-79.6)	3813	81.6 (79.6-83.4)
Used 2 drugs	13	9 (4.2-18.2)	42	10.9 (7.1-16.5)	40	5.9 (3.8-9.1)	121	6.3 (4.6-8.6)	176	8.9 (7.4-10.7)	392	8 (6.9-9.2)
Used 3 drugs	3	0.7 (0.2-3.1)	8	1.4 (0.5-4)	9	2.6 (1.1-6.1)	39	2.5 (1.5-4)	43	2.2 (1.4-3.2)	102	2.2 (1.6-2.9)
Used 4 drugs	1	0.4 (0.1-3.1)	3	0.4 (0.1-1.5)	2	0.2 (0-0.7)	11	0.8 (0.2-3)	15	0.5 (0.3-1)	32	0.5 (0.3-0.9)
Used 5 drugs	None	None	2	0.6 (0.1-3.3)	1	0.5 (0.1-3.7)	11	0.5 (0.2-1.2)	4	0.2 (0.1-0.5)	18	0.3 (0.2-0.7)
Used 6 drugs	None	None	None	None	None	None	3	0.3 (0.1-1.1)	5	0.1 (0-0.5)	8	0.1 (0.1-0.4)
Used 7 drugs	None	None	None	None	4	0.7 (0.2-2.2)	1	0.1 (0-0.5)	23	1 (0.5-2)	28	0.6 (0.3-1)
Used 8 drugs	None	None	4	2.2 (0.6-7.9)	46	10.3 (6.2-16.5)	15	0.7 (0.3-1.3)	282	10.3 (8.1-13.1)	347	6.6 (5.3-8.2)

Supplementary Table S10: The proportion of the population using drugs with higher frequency (i.e., weekly, or almost daily)

Type of drug	2008	2012	2017
Cannabis	1.60%	2.17%	3.45%
Cocaine	0.01%	0.04%	0.16%
Amphetamines	0.06%	0.07%	0.11%
Inhalants	<0.01%	0.02%	0.05%
Sedatives	0.09%	0.11%	0.01%
Hallucinogens	<0.01%	0.008%	0.002%
Opioids	<0.01%	0.07%	0.23%
Any drug use	2.10%	3.46%	4.88%

References

Babbie, E. R. (2020). *The practice of social research*. Cengage learning.

Hong, S., & Lynn, H. S. (2020). Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. *BMC Medical Research Methodology*, *20*(1), 199. https://doi.org/10.1186/s12874-020-01080-1

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S.-L. T., Manderscheid, R. W., & Walters, E. E. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, *60*(2), Article 2.

Ramosaj, B., & Pauly, M. (2019). Predicting missing values: A comparative study on non-parametric approaches for imputation. *Computational Statistics*, *34*(4), 1741–1764. https://doi.org/10.1007/s00180-019-00900-3

Shisana O & Simbayi L, Nelson Mandela/HSRC study of HIV/AIDS: (2004). *South African national HIV prevalence, behavioural risks and mass media household survey 2002. Cape Town: HSRC Press*.

Statistics South Africa. (2011). *Census 2011 Monogram: A profile of education enrolment, attainment and progression in South Africa*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www.statssa.gov.za/publications/Report-03-01-81/Report-03-01-812011.pdf

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *10*(6), 363–377.

World Health Organization. (2001). *AUDIT: The alcohol use disorders identification test: Guidelines for use in primary health care*. World Health Organization.