

**South Africa**

**Emily Wong, AHRI**  
**Krishna Reddy, Massachusetts General Hospital**  
**Mark Siedner, AHRI**

**AHRI.Vukuzazi Tobacco survey:investigating  
initiation cessation and relapse of tobacco smoking**

**Study Documentation**

June 14, 2024

# Metadata Production

Metadata Producer(s)	Africa Health Research Institute (AHRI)
Identification	DDI.Tobacco.init.cess.relapspaper.datadeposit

# Table of Contents

<a href="#">Overview.....</a>	<a href="#">4</a>
<a href="#">Scope &amp; Coverage.....</a>	<a href="#">4</a>
<a href="#">Producers &amp; Sponsors.....</a>	<a href="#">4</a>
<a href="#">Sampling.....</a>	<a href="#">5</a>
<a href="#">Data Collection.....</a>	<a href="#">5</a>
<a href="#">Data Processing &amp; Appraisal.....</a>	<a href="#">5</a>
<a href="#">Accessibility.....</a>	<a href="#">5</a>
<a href="#">Files Description.....</a>	<a href="#">6</a>
<a href="#">Tobacco init cess relaps paper_datadeposit.....</a>	<a href="#">6</a>
<a href="#">Variables List.....</a>	<a href="#">7</a>
<a href="#">Tobacco init cess relaps paper_datadeposit.....</a>	<a href="#">7</a>
<a href="#">Variables Description.....</a>	<a href="#">10</a>
<a href="#">Tobacco init cess relaps paper_datadeposit.....</a>	<a href="#">11</a>

# AHRI.Vukuzazi Tobacco survey:investigating initiation cessation and relapse of tobacco smoking

Overview	
<b>Identification</b>	AHRI.Tobacco.init.cess.relapsaper.datadeposit
<b>Version</b>	v1.0.0
<b>Abstract</b> Participants enrolled in the Vukuzazi population cohort in rural KwaZulu-Natal, South Africa completed a baseline tobacco behavioural survey during May 2018-March 2020. A follow-up survey was conducted during February 2021-November 2022 among 1448 participants who completed the baseline tobacco behavioural survey. These included all participants aged =15y who reported current and former smoking at baseline (to detect cessation and relapse) and in a random selection of participants aged 15-29y who reported never smoking at baseline (to detect initiation). The dataset contains the records of the 1448 participants who completed the follow-up survey merged to their baseline data records.	
<b>Kind of Data</b>	Survey data. Records at follow-up linked to baseline records
<b>Unit of Analysis</b>	Individual participant responses.

Scope & Coverage	
<b>Keywords</b>	Tobacco smoking; smoking cessation; smoking relapse; smoking initiation
<b>Topics</b>	<a href="https://meshb.nlm.nih.gov/record/ui?name=tobacco%20smoking">https://meshb.nlm.nih.gov/record/ui?name=tobacco%20smoking</a> , <a href="https://meshb.nlm.nih.gov/record/ui?name=smoking%20cessation">https://meshb.nlm.nih.gov/record/ui?name=smoking%20cessation</a> , <a href="https://meshb.nlm.nih.gov/record/ui?name=smoking%20prevention">https://meshb.nlm.nih.gov/record/ui?name=smoking%20prevention</a> , <a href="https://meshb.nlm.nih.gov/record/ui?name=nicotine%20dependence">https://meshb.nlm.nih.gov/record/ui?name=nicotine%20dependence</a> , <a href="https://meshb.nlm.nih.gov/record/ui?name=south%20africa">https://meshb.nlm.nih.gov/record/ui?name=south%20africa</a>
<b>Time Period(s)</b>	2018-2022
<b>Countries</b>	South Africa
<b>Geographic Coverage</b> The Vukuzazi programme is situated in the uMkhanyakude district in rural KwaZulu-Natal province which is approximately 200km north of Durban. The area is approximately 845km <sup>2</sup> in size.  For more refer: Gunda R, Koole O, Gareta D, Olivier S, Surujdeen A, Smit T, et al. Cohort Profile: The Vukuzazi ('Wake Up and Know Yourself' in isiZulu) population science programme. Int J Epidemiol. 2022;51(3):e131-e42. Epub 2021/12/02. doi: 10.1093/ije/dyab229. PubMed PMID: 34849923; PubMed Central PMCID: PMCPMC9189966.	
<b>Universe</b> All participants from AHRI's Vukuzazi study aged 15 years and older who reported current and former smoking at baseline and participants aged 15-29 years who reported never smoking at baseline.	

Producers & Sponsors	
<b>Primary Investigator(s)</b>	Emily Wong, AHRI Krishna Reddy, Massachusetts General Hospital Mark Siedner, AHRI
<b>Other Producer(s)</b>	Africa Health Research Institute (AHRI)
<b>Funding Agency/ies</b>	Wellcome Trust (WT) , Funder National Institute on Drug Abuse, NIH (NIDA) , Funder Fogarty International Center, NIH , Funder

<b>Other Acknowledgment(s)</b>	Mareca Sithole , Research Data management , AHRI Stephen Olivier , Data quality, data analytics , AHRI Hloniphile Ngubane , Data collection , AHRI Thando Zulu , Data collection, data quality , AHRI Ronel Sewpaul , Data analytics , AHRI Gina Kruse , Data interpretation , AHRI Nancy A. Rigotti , Data interpretation , AHRI Dickman Gareta , Research Data management , AHRI
--------------------------------	---

## Sampling

### Sampling Procedure

All baseline current and former smokers aged 15 years and older were sampled. A random sample of never-smokers aged 15-29 years were selected.

## Data Collection

<b>Data Collection Dates</b>	Baseline: start 2018-05-01 Baseline: end 2020-03-30 Follow-Up Survey: start 2021-02-01 Follow-Up Survey: end 2022-11-30
------------------------------	--

## Data Processing & Appraisal

### Data Editing

The index data is completely anonymised.

## Accessibility

### Access Conditions

Access to the data requires accurate completion of the online data access application form accessible on the AHRI Data repository(<<https://data.ahri.org/>>). Data users are required to abide by the data use conditions stipulated on the application for access to the data. Failure to do so may result in their data access privileges being revoked by the Data Custodian. In order to recognise the effort and intellectual contributions of AHRI investigators in producing and curating the data, users of AHRI data must acknowledge the source of the data and abide by the terms and conditions under which the data is accessed and must cite the dataset in publication using the citation provided as part of this documentation. All analytical datasets published on the AHRI Data Repository are assigned digital object identifier (DOIs) and the DOIs can be found on the Data Repository under Study Description tab - Access policy. AHRI data users are required to always cite the dataset using the relevant DOI.

### Citation Requirements

Wong, E., Reddy, K., & Siedner, M. (2024). AHRI.Vukuzazi Tobacco survey:investigating initiation cessation and relapse of tobacco smoking [Data set]. Africa Health Research Institute. DOI:<https://doi.org/10.23664/AHRI.TOBACCO.init.cess.relapspaper.datadeposit>

# Files Description

Dataset contains 1 file(s)

Tobacco init cess relaps paper_datadeposit	
# Cases	1448
# Variable(s)	50

# Variables List

Dataset contains 50 variable(s)

File Tobacco init cess relaps paper_datadeposit							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	<a href="#">record_id</a>	Record ID	discrete	character-7	1448	0	-
2	<a href="#">Individu ..</a>	Individual ID	continuous	numeric-12.0	1448	0	-
3	<a href="#">introduc ..</a>	Follow-up: introduction complete	discrete	numeric-10.0	1448	0	-
4	<a href="#">ts_has_s ..</a>	Followup: Have you smoked 100 cigarettes or other tobacco products, suchas cig	discrete	numeric-8.0	755	693	-
5	<a href="#">ts_how_o ..</a>	Followup: How often do you CURRENTLY smoke tobacco product?	discrete	numeric-31.0	485	963	-
6	<a href="#">ohu_mobi ..</a>	Followup: Problems with mobility	discrete	numeric-28.0	751	697	-
7	<a href="#">ohu_self ..</a>	Followup: Problems with self-care	discrete	numeric-47.0	751	697	-
8	<a href="#">ohu_usua ..</a>	Followup: Problems with usual activities (e.g., work, study, housework, family,	discrete	numeric-56.0	751	697	-
9	<a href="#">ohu_pain</a>	Followup: Pain/Discomfort	discrete	numeric-34.0	751	697	-
10	<a href="#">ohu_anxi ..</a>	Followup: Anxiety/Depression	discrete	numeric-36.0	751	697	-
11	<a href="#">ohu_th</a>	Followup: Since the Vukuzazi health camp,have you experienced or been told by a	discrete	numeric-8.0	751	697	-
12	<a href="#">FU_sless ..</a>	Followup: Smokeless tobacco status	discrete	numeric-22.0	753	695	-
13	<a href="#">phoneout ..</a>	Followup: Follow-up contact attempt outcome	discrete	numeric-24.0	1448	0	-
14	<a href="#">fu_date1</a>	Followup: Follow-up interview date	discrete	character-11	1448	-	-
15	<a href="#">MobileCl ..</a>	Baseline: Date of visit at the Vukuzazi mobile clinic	discrete	character-11	1448	-	-
16	<a href="#">AgeAtEnr ..</a>	Baseline: Age at enrolment	continuous	numeric-12.0	1448	0	-
17	<a href="#">CurrSmoker</a>	Baseline: Do you currently smoke any tobacco products, such as cigarettes, cigar	discrete	numeric-12.0	1448	0	-
18	<a href="#">EverSmoked</a>	Baseline: In the past, did you ever smoke any tobacco products?	discrete	numeric-12.0	710	738	-
19	<a href="#">BL_Smoke ..</a>	Baseline: Smoking group	discrete	numeric-12.0	1448	0	-
20	<a href="#">AgeStart ..</a>	Baseline: How old were you when you first started smoking?	continuous	numeric-12.0	624	824	-
21	<a href="#">HowLongS ..</a>	Baseline: Do you remember how long ago it was?	continuous	numeric-12.0	135	1313	-

File Tobacco init cess relaps paper_datadeposit							
#	Name	Label	Type	Format	Valid	Invalid	Question
22	<a href="#">StopSmok ..</a>	Baseline: During the past 12 months, have you tried to stop smoking?	discrete	numeric-12.0	1448	0	-
23	<a href="#">StopSmok ..</a>	Baseline: During any visit to a doctor or other health worker in the past 12 mon	discrete	numeric-12.0	948	500	-
24	<a href="#">SmokeAmo ..</a>	Baseline: Number of products smoked/used per day	continuous	numeric-10.0	820	628	-
25	<a href="#">BL_Smoke ..</a>	Baseline: Number of products smoked/used per day (category)	discrete	numeric-12.0	820	628	-
26	<a href="#">EverCons ..</a>	Baseline: Have you ever consumed any alcohol such as beer, wine, spirits or [add	discrete	numeric-12.0	1448	0	-
27	<a href="#">Consumed ..</a>	Baseline: Have you consumed any alcohol within the past 12 months?	discrete	numeric-12.0	701	747	-
28	<a href="#">Consumed ..</a>	Baseline: Have you consumed any alcohol within the past 30 days?	discrete	numeric-12.0	701	747	-
29	<a href="#">HIVCascade</a>	Baseline: HIV care cascade status	discrete	numeric-12.0	1448	0	-
30	<a href="#">HTNDiag</a>	Baseline: Hypertension diagnosis (primary defn)	discrete	numeric-12.0	1444	4	-
31	<a href="#">DMDiag</a>	Baseline: Diabetes diagnosis (primary definition)	discrete	numeric-12.0	1447	1	-
32	<a href="#">Employment</a>	Baseline: Employment	discrete	numeric-12.0	1284	164	-
33	<a href="#">BL_smokd ..</a>	Baseline: Number of years since started smoking (in years among current smokers)	continuous	numeric-9.0	758	690	-
34	<a href="#">FU_Smoke ..</a>	Follow_up: Smoking status (based on ever/never 100 cigs)	discrete	numeric-21.0	754	694	-
35	<a href="#">BL_ageca ..</a>	Baseline age at enrolment (category)	discrete	numeric-9.0	1448	0	-
36	<a href="#">BL_ageca ..</a>	Baseline age at enrolment (2 category)	discrete	numeric-9.0	1448	0	-
37	<a href="#">HIVcasca ..</a>	Baseline HIV cascade status (3 category)	discrete	numeric-67.0	1448	0	-
38	<a href="#">alcohol_..</a>	Baseline alcohol consumption (past 30 days) (category)	discrete	numeric-9.0	1448	0	-
39	<a href="#">alcohol_..</a>	Baseline alcohol consumption (past 12 months) (category)	discrete	numeric-9.0	1448	0	-
40	<a href="#">employ</a>	Baseline Employment (4 category)	discrete	numeric-19.0	1448	0	-
41	<a href="#">ses3c</a>	Baseline Socioeconomic status	discrete	numeric-9.0	1405	43	-



File Tobacco init cess relaps paper_datadeposit							
#	Name	Label	Type	Format	Valid	Invalid	Question
42	<a href="#">BL_Smoke_..</a>	Baseline smoking intensity (recoded categories)	discrete	numeric-14.0	948	500	-
43	<a href="#">quit_adv_..</a>	Received advice to quit smoking (reported at baseline)	discrete	numeric-9.0	948	500	-
44	<a href="#">FU_daily_..</a>	Has problems with mobility, self-care, usual activities or pain at follow-up	discrete	numeric-9.0	751	697	-
45	<a href="#">FU_daily_..</a>	Has problems with mobility, self-care, usual activities or pain (category) at fol	discrete	numeric-9.0	751	697	-
46	<a href="#">smok_start</a>	Smoking initiation between Baseline and Follow_up among baseline never smokers	discrete	numeric-9.0	234	1214	-
47	<a href="#">smok_rel_..</a>	Smoking relapse between Baseline and Follow_up among baseline former smokers	discrete	numeric-9.0	46	1402	-
48	<a href="#">smok_quit</a>	Smoking cessation between Baseline and Follow_up among baseline current smokers	discrete	numeric-9.0	411	1037	-
49	<a href="#">responder</a>	Did participant respond to follow-up survey?	discrete	numeric-23.0	1448	0	-
50	<a href="#">BL_smokd_..</a>	Smoking duration	discrete	numeric-9.0	1436	12	-

# Variables Description

**Dataset contains 50 variable(s)**

## File : Tobacco init cess relaps paper\_datadeposit

### # record\_id: Record ID

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=1448 /-] [Invalid=0 /-]

### # IndividualID: Individual ID

Information	[Type= continuous] [Format=numeric] [Range= 24-38747] [Missing=*]
Statistics [NW/ W]	[Valid=1448 /-] [Invalid=0 /-] [Mean=17368.106 /-] [StdDev=10804.541 /-]

### # introduction\_complete: Follow-up: introduction complete

Information	[Type= discrete] [Format=numeric] [Range= 0-2] [Missing=*]
Statistics [NW/ W]	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Incomplete	682	47.1%
1	Unverified	1	0.1%
2	Complete	765	52.8%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # ts\_has\_smoked: Followup\_up: Have you smoked 100 cigarettes or other tobacco products, suchas cig

Information	[Type= discrete] [Format=numeric] [Range= 1-2] [Missing=*]
Statistics [NW/ W]	[Valid=755 /-] [Invalid=693 /-]

Value	Label	Cases	Percentage
1	Yes	486	64.4%
2	No	269	35.6%
Sysmiss		693	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # ts\_how\_often: Follow\_up: How often do you CURRENTLY smoke tobacco product?

Information	[Type= discrete] [Format=numeric] [Range= 1-4] [Missing=*]
Statistics [NW/ W]	[Valid=485 /-] [Invalid=963 /-]

Value	Label	Cases	Percentage
1	Daily	301	62.1%
2	Less than daily	71	14.6%
3	Seldom (less than once a month)	10	2.1%
4	None, I have quit smoking	103	21.2%
Sysmiss		963	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # ohu\_mobility: Follow\_up: Problems with mobility

Information	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
Statistics [NW/ W]	[Valid=751 /-] [Invalid=697 /-]

Value	Label	Cases	Percentage
1	I have no problems walking	722	96.1%
2	I have some problems walking	29	3.9%
3	I am confined to bed	0	
Sysmiss		697	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

## File : Tobacco init cess relaps paper\_datadeposit

### # ohu\_selfcare: Follow\_up: Problems with self-care

**Information** [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

**Statistics [NW/ W]** [Valid=751 /-] [Invalid=697 /-]

Value	Label	Cases	Percentage
1	I have no problems with self-care	738	98.3%
2	I have some problems washing or dressing myself	13	1.7%
3	I am unable to wash or dress myself	0	
Sysmiss		697	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # ohu\_usual\_activities: Follow\_up: Problems with usual activities (e.g., work, study, housework, family,

**Information** [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

**Statistics [NW/ W]** [Valid=751 /-] [Invalid=697 /-]

Value	Label	Cases	Percentage
1	I have no problems with performing my usual activities	727	96.8%
2	I have some problems with performing my usual activities	24	3.2%
3	I am unable to perform my usual activities	0	
Sysmiss		697	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # ohu\_pain: Follow\_up: Pain/Discomfort

**Information** [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

**Statistics [NW/ W]** [Valid=751 /-] [Invalid=697 /-]

Value	Label	Cases	Percentage
1	I have no pain or discomfort	694	92.4%
2	I have moderate pain or discomfort	52	6.9%
3	I have extreme pain or discomfort	5	0.7%
Sysmiss		697	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # ohu\_anxiety: Follow\_up: Anxiety/Depression

**Information** [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

**Statistics [NW/ W]** [Valid=751 /-] [Invalid=697 /-]

Value	Label	Cases	Percentage
1	I am not anxious or depressed	737	98.1%
2	I am moderately anxious or depressed	12	1.6%
3	I am extremely anxious or depressed	2	0.3%
Sysmiss		697	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # ohu\_tb: Follow\_up: Since the Vukuzazi health camp,have you experienced or been told by a

**Information** [Type= discrete] [Format=numeric] [Range= 1-2] [Missing=\*]

**Statistics [NW/ W]** [Valid=751 /-] [Invalid=697 /-]

Value	Label	Cases	Percentage
1	Yes	38	5.1%
2	No	713	94.9%

## File : Tobacco init cess relaps paper\_datadeposit

### # ohu\_tb: Follow\_up: Since the Vukuzazi health camp,have you experienced or been told by a

Value	Label	Cases	Percentage
Sysmiss		697	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # FU\_sless3Cat: Follow\_up: Smokeless tobacco status

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=753 /-] [Invalid=695 /-]

Value	Label	Cases	Percentage
1	Current Smokeless user	46	6.1%
2	Former Smokeless user	11	1.5%
3	Never Smokeless user	696	92.4%
Sysmiss		695	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # phoneoutcome\_final: Follow\_up: Follow-up contact attempt outcome

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-4] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Call was answered	954	65.9%
2	Call gone to Voicemail	383	26.5%
3	Phone remains unanswered	91	6.3%
4	Number does not exist	20	1.4%

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # fu\_date1: Follow\_up: Follow-up interview date

<b>Information</b>	[Type= discrete] [Format=character] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-]

### # MobileClinicVisitDate: Baseline: Date of visit at the Vukuzazi mobile clinic

<b>Information</b>	[Type= discrete] [Format=character] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-]

### # AgeAtEnrolment: Baseline: Age at enrolment

<b>Information</b>	[Type= continuous] [Format=numeric] [Range= 15-101] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-] [Mean=35.114 /-] [StdDev=15.478 /-]

### # CurrSmoker: Baseline: Do you currently smoke any tobacco products, such as cigarettes, cigar

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Yes	948	65.5%
2	No	500	34.5%
3	NA	0	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*


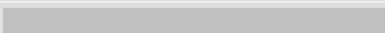
### # EverSmoked: Baseline: In the past, did you ever smoke any tobacco products?

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
--------------------	--

## File : Tobacco init cess relaps paper\_datadeposit

### # EverSmoked: Baseline: In the past, did you ever smoke any tobacco products?

Statistics [NW/ W] [Valid=710 /-] [Invalid=738 /-]



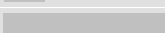
Value	Label	Cases	Percentage
1	Yes	135	 19.0%
2	No	575	 81.0%
3	NA	0	
Sysmiss		738	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # BL\_Smoke3Cat: Baseline: Smoking group

Information [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

Statistics [NW/ W] [Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Current smoker	948	 65.5%
2	Ex-smoker	100	 6.9%
3	Never smoker	400	 27.6%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # AgeStartedSmoking: Baseline: How old were you when you first started smoking?

Information [Type= continuous] [Format=numeric] [Range= 0-52] [Missing=\*]

Statistics [NW/ W] [Valid=624 /-] [Invalid=824 /-] [Mean=20.155 /-] [StdDev=6.808 /-]

### # HowLongStartedSmoking: Baseline: Do you remember how long ago it was?


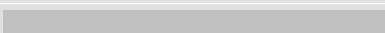
Information [Type= continuous] [Format=numeric] [Range= 0-50] [Missing=\*]

Statistics [NW/ W] [Valid=135 /-] [Invalid=1313 /-] [Mean=18.63 /-] [StdDev=11.691 /-]

### # StopSmoking12mo: Baseline: During the past 12 months, have you tried to stop smoking?

Information [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

Statistics [NW/ W] [Valid=1448 /-] [Invalid=0 /-]


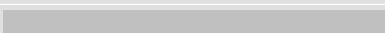
Value	Label	Cases	Percentage
1	Yes	141	 9.7%
2	No	1307	 90.3%
3	NA	0	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # StopSmokingAdvice12mo: Baseline: During any visit to a doctor or other health worker in the past 12 mon

Information [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

Statistics [NW/ W] [Valid=948 /-] [Invalid=500 /-]

Value	Label	Cases	Percentage
1	Yes	55	 5.8%
2	No	716	 75.5%
3	NA	177	 18.7%
Sysmiss		500	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # SmokeAmount: Baseline: Number of products smoked/used per day

Information [Type= continuous] [Format=numeric] [Range= 0-77] [Missing=\*/101]

## File : Tobacco init cess relaps paper\_datadeposit

### # SmokeAmount: Baseline: Number of products smoked/used per day

Statistics [NW/ W] [Valid=820 /-] [Invalid=628 /-] [Mean=6.076 /-] [StdDev=5.569 /-]

Value	Label	Cases	Percentage
101	..		

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # BL\_SmokeAmountCat: Baseline: Number of products smoked/used per day (category)

Information [Type= discrete] [Format=numeric] [Range= 0-3] [Missing=\*]

Statistics [NW/ W] [Valid=820 /-] [Invalid=628 /-]

Value	Label	Cases	Percentage
0	Never	40	4.9%
1	Light	608	74.1%
2	Moderate	143	17.4%
3	Heavy	29	3.5%
Sysmiss		628	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # EverConsumedAlcohol: Baseline: Have you ever consumed any alcohol such as beer, wine, spirits or [add

Information [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

Statistics [NW/ W] [Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Yes	701	48.4%
2	No	747	51.6%
3	NA	0	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # ConsumedAlcohol12mo: Baseline: Have you consumed any alcohol within the past 12 months?

Information [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

Statistics [NW/ W] [Valid=701 /-] [Invalid=747 /-]

Value	Label	Cases	Percentage
1	Yes	610	87.0%
2	No	91	13.0%
3	NA	0	
Sysmiss		747	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # ConsumedAlcohol30d: Baseline: Have you consumed any alcohol within the past 30 days?

Information [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

Statistics [NW/ W] [Valid=701 /-] [Invalid=747 /-]

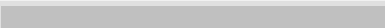



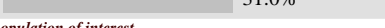
Value	Label	Cases	Percentage
1	Yes	291	41.5%
2	No	410	58.5%
3	NA	0	
Sysmiss		747	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

## File : Tobacco init cess relaps paper\_datadeposit

### # HIVCascade: Baseline: HIV care cascade status


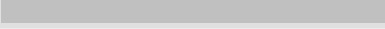
<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 0-4] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Negative	750	 51.8%
1	Positive and undiagnosed	151	 10.4%
2	Diagnosed not in care	39	 2.7%
3	In care, not controlled	59	 4.1%
4	Controlled	449	 31.0%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # HTNDiag: Baseline: Hypertension diagnosis (primary defn)


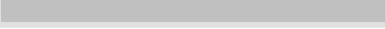
<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1444 /-] [Invalid=4 /-]

Value	Label	Cases	Percentage
1	Yes	204	 14.1%
2	No	1240	 85.9%
3	NA	0	
Sysmiss		4	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # DMDDiag: Baseline: Diabetes diagnosis (primary definition)

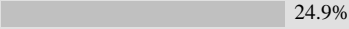
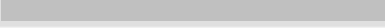
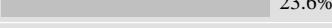


<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1447 /-] [Invalid=1 /-]

Value	Label	Cases	Percentage
1	Yes	38	 2.6%
2	No	1409	 97.4%
3	NA	0	
Sysmiss		1	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # Employment: Baseline: Employment

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-5] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1284 /-] [Invalid=164 /-]

Value	Label	Cases	Percentage
1	Employed	320	 24.9%
2	Unemployed_strict	434	 33.8%
3	Unemployed_expanded	303	 23.6%
4	Not labour force	164	 12.8%
5	Status Unknown	63	 4.9%
Sysmiss		164	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # BL\_smokduration: Baseline: Number of years since started smoking (in years among current smokers)

<b>Information</b>	[Type= continuous] [Format=numeric] [Range= 0-66] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=758 /-] [Invalid=690 /-] [Mean=17.673 /-] [StdDev=12.651 /-]



## File : Tobacco init cess relaps paper\_datadeposit

### # FU\_Smoke3Cat: Follow\_up: Smoking status (based on ever/never 100 cigs)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=754 /-] [Invalid=694 /-]

Value	Label	Cases	Percentage
1	Current Smoker	382	50.7%
2	Former Smoker	103	13.7%
3	Never Smoked 100 cigs	269	35.7%
Sysmiss		694	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # BL\_agecat\_enrol: Baseline age at enrolment (category)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-5] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	<20	196	13.5%
2	20-29	506	34.9%
3	30-39	269	18.6%
4	40-49	194	13.4%
5	>=50	283	19.5%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # BL\_agecat2\_enrol: Baseline age at enrolment (2 category)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	<50	1165	80.5%
1	>=50	283	19.5%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # HIVcascade\_rec: Baseline HIV cascade status (3 category)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 0-2] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Negative	750	51.8%
1	Positive & undiagnosed, Diagnosed not in care, In care,uncontrolled	249	17.2%
2	Controlled	449	31.0%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # alcohol\_30d: Baseline alcohol consumption (past 30 days) (category)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	No	1157	79.9%
1	Yes	291	20.1%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

## File : Tobacco init cess relaps paper\_datadeposit

### # alcohol\_12mo: Baseline alcohol consumption (past 12 months) (category)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	No	838	57.9%
1	Yes	610	42.1%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # employ: Baseline Employment (4 category)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-4] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Employed	320	22.1%
2	Unemployed	737	50.9%
3	Not in labour force	164	11.3%
4	Unknown	227	15.7%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # ses3c: Baseline Socioeconomic status

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=1405 /-] [Invalid=43 /-]

Value	Label	Cases	Percentage
1	Low	607	43.2%
2	Middle	318	22.6%
3	High	480	34.2%
Sysmiss		43	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # BL\_SmokeAmount2Cat: Baseline smoking intensity (recoded categories)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=948 /-] [Invalid=500 /-]

Value	Label	Cases	Percentage
1	Light	648	68.4%
2	Moderate-heavy	172	18.1%
3	Unknown	128	13.5%
Sysmiss		500	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

### # quit\_advice2cat: Received advice to quit smoking (reported at baseline)

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
<b>Statistics [NW/ W]</b>	[Valid=948 /-] [Invalid=500 /-]

Value	Label	Cases	Percentage
0	No	893	94.2%
1	Yes	55	5.8%
Sysmiss		500	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

## File : Tobacco init cess relaps paper\_datadeposit

### # FU\_dailydiff\_score: Has problems with mobility, self-care, usual activities or pain at follow-up

**Information** [Type= discrete] [Format=numeric] [Range= 4-8] [Missing=\*]

**Statistics [NW/ W]** [Valid=751 /-] [Invalid=697 /-]

### # FU\_dailydiff: Has problems with mobility, self-care, usual activities or pain (category) at fol

**Information** [Type= discrete] [Format=numeric] [Range= 0-1] [Missing=\*]

**Statistics [NW/ W]** [Valid=751 /-] [Invalid=697 /-]

Value	Label	Cases	Percentage
0	None	677	90.1%
1	Some	74	9.9%
Sysmiss		697	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # smok\_start: Smoking initiation between Baseline and Follow\_up among baseline never smokers

**Information** [Type= discrete] [Format=numeric] [Range= 0-1] [Missing=\*]

**Statistics [NW/ W]** [Valid=234 /-] [Invalid=1214 /-]

Value	Label	Cases	Percentage
0	No	206	88.0%
1	Yes	28	12.0%
Sysmiss		1214	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # smok\_relapse: Smoking relapse between Baseline and Follow\_up among baseline former smokers

**Information** [Type= discrete] [Format=numeric] [Range= 0-1] [Missing=\*]

**Statistics [NW/ W]** [Valid=46 /-] [Invalid=1402 /-]

Value	Label	Cases	Percentage
0	No	41	89.1%
1	Yes	5	10.9%
Sysmiss		1402	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # smok\_quit: Smoking cessation between Baseline and Follow\_up among baseline current smokers

**Information** [Type= discrete] [Format=numeric] [Range= 0-1] [Missing=\*]

**Statistics [NW/ W]** [Valid=411 /-] [Invalid=1037 /-]

Value	Label	Cases	Percentage
0	No	358	87.1%
1	Yes	53	12.9%
Sysmiss		1037	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

### # responder: Did participant respond to follow-up survey?

**Information** [Type= discrete] [Format=numeric] [Range= 0-1] [Missing=\*]

**Statistics [NW/ W]** [Valid=1448 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Didnt Participate in FU	694	47.9%
1	Participated in FU	754	52.1%

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

## File : Tobacco init cess relaps paper\_datadeposit

### # BL\_smokduration\_3cat: Smoking duration

**Information** [Type= discrete] [Format=numeric] [Range= 1-3] [Missing=\*]

**Statistics [NW/ W]** [Valid=1436 /-] [Invalid=12 /-]

Value	Label	Cases	Percentage
1	1-5 years	132	9.2%
2	>5 years	614	42.8%
3	Unknown	690	48.1%
Sysmiss		12	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*