

South Africa

Herbst, Kobus, SAPRIN/AHRI

Siedner, Mark, AHRI/MGH

Harling, Guy, AHRI/UCL

**AHRI.COVID-19 vaccine uptake, confidence and
hesitancy in rural KwaZulu-Natal, South Africa between
April 2021 and April 2022:a subset of the individual
population-wide surveillance system annual interview**

Study Documentation

June 1, 2023

Metadata Production

Metadata Producer(s)	Africa Health Research Institute (AHRI)
Identification	DDI.AHRI.SAPRIN.COVID-19.Vaccine.Hesitancy.Dataset

Table of Contents

Overview.....	4
Scope & Coverage.....	4
Producers & Sponsors.....	4
Sampling.....	5
Data Collection.....	5
Data Processing & Appraisal.....	5
Accessibility.....	5
Files Description.....	6
covid_vax_plos_minimum_dataset.....	6
Variables List.....	7
covid_vax_plos_minimum_dataset.....	7
Variables Description.....	9
covid_vax_plos_minimum_dataset.....	10

AHRI.COVID-19 vaccine uptake, confidence and hesitancy in rural KwaZulu-Natal, South Africa between April 2021 and April 2022: a subset of the individual population-wide surveillance system annual interview

Overview	
Identification	AHRI.SAPRIN.COVID-19.Vaccine.Hesitancy.Dataset
Version	v1.0.0
<p>Abstract</p> <p>This dataset is a subset of the wider population-based Covid-19 surveillance run at the Africa Health Research Institute from 2020 onwards. The dataset covers one complete year of data collection, such that all residents had the opportunity to participate. The dataset specifically provides all observations and variables needed to replicate the analyses described in the journal article “COVID-19 vaccine uptake, confidence and hesitancy in rural KwaZulu-Natal, South Africa between April 2021 and April 2022: a continuous cross-sectional surveillance study” published in PLOS Global Public Health in 2023. The dataset includes variable on Covid vaccine uptake and willingness to take a hypothetical vaccine offer on the day of interview, as well as variables measuring four groups of potential predictors of these vaccine outcomes: demographics (age, sex), pre-existing conditions (information sources, government trust, education, urbanicity), contextual factors (impact of Covid on household economics and community wellbeing, Covid-related stigma, household age composition) and cues to action (recent case counts in KwaZulu-Natal, concern about impact if infected with Covid, knowledge of others with past Covid infection, household vaccination status, depression/anxiety) and interview date.</p>	
Kind of Data	Survey data
Unit of Analysis	Individual

Scope & Coverage	
Keywords	Covid-19; Surveillance; South Africa; Health and Demographic Surveillance System; Behaviour; Attitudes
Topics	COVID-19; Vaccination; Public Health Surveillance; Epidemiology; South Africa
Time Period(s)	2021-2022
Countries	South Africa
<p>Geographic Coverage</p> <p>AHRI demographic surveillance area, uMkhanyakude district in northern KwaZulu-Natal</p>	
<p>Universe</p> <p>All individuals aged 18 and over resident within the areas of the Africa Health Research Institute Population Intervention Programme</p>	

Producers & Sponsors	
Primary Investigator(s)	Herbst, Kobus, SAPRIN/AHRI Siedner, Mark, AHRI/MGH Harling, Guy, AHRI/UCL
Other Producer(s)	Africa Health Research Institute (AHRI)
Funding Agency/ies	Wellcome Trust (WT) , Core funding SAPRIN (SAPRIN) , Surveillance funding
Other Acknowledgment(s)	Tinofa Mutevedzi , SAPRIN Thobeka Mngomezulu , PIP Coordinator , AHRI

Nomathamsanqa Majozi , Head of Public Engagement , AHRI
 Siyabonga Nxumalo , SAPRIN Research Data Manager , AHRI
 Njabulo Dayi , Research Data Manager , AHRI
 Dickman Gareta , Head of Research Data Management , AHRI
 Jaco Dreyer , Senior Research Data Manager , AHRI
 Eugene Ehlers , Senior Software Developer , AHRI

Sampling

Sampling Procedure

All adult residents in the geographic area were eligible via an individual face-to-face interview. Multiple attempts were made to reach each individual if necessary. The final sample reflects all those who consented to and completed an interview.

Data Collection

Data Collection Dates	start 2021-04-20 end 2022-04-20
------------------------------	------------------------------------

Data Processing & Appraisal

Data Editing

Pentaho Data Integration was used to extract the datasets. NESSTAR Publisher was used to document the datasets.

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

Herbst, K., Siedner, M., & Harling, G. (2023). COVID-19 vaccine uptake, confidence and hesitancy in rural KwaZulu-Natal, South Africa between April 2021 and April 2022: a subset of the individual population-wide surveillance system annual interview [Data set]. Africa Health Research Institute. <https://doi.org/10.23664/AHRI.SAPRIN.COVID-19.VACCINE.HESITANCY.DATASET>

Files Description

Dataset contains 1 file(s)

covid_vax_plos_minimum_dataset	
# Cases	15709
# Variable(s)	27

Variables List

Dataset contains 27 variable(s)

File covid_vax_plos_minimum_dataset							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	IntId	Unique or Internal ID of Individual	continuous	numeric-12.0	15709	0	-
2	HHIntId	Household ID	continuous	numeric-12.0	15369	340	-
3	Interview_..	Interview Date	discrete	character-11	15709	-	-
4	Sex	Sex	discrete	numeric-12.0	15709	0	-
5	agegroup	Age of respondent (4 levels)	discrete	numeric-9.0	15709	0	-
6	Interview_..	Month of interview	discrete	numeric-9.0	15709	0	-
7	covax_wi_..	Covid vaccine uptake or willingness to take (5 levels)	discrete	numeric-20.0	15709	0	-
8	info_tra_..	Information from traditional sources	discrete	numeric-9.0	11075	4634	-
9	info_per_..	Information from personal network sources	discrete	numeric-9.0	11075	4634	-
10	info_he_..	Information from healthcare sources	discrete	numeric-9.0	11075	4634	-
11	info_com_..	Information from community sources	discrete	numeric-9.0	11075	4634	-
12	gov_mist_..	Mistrust in government - per standard deviation	continuous	numeric-9.0	15664	45	-
13	educ_att_..	Highest educational attainment	discrete	numeric-19.0	14455	1254	-
14	HH_older	Has household member aged 60 or older	discrete	numeric-9.0	15369	340	-
15	IsUrbanO_..	Urbanicity of household	discrete	numeric-12.0	15709	0	-
16	subUnits_..	Isigodi of residence	discrete	numeric-22.0	15709	0	-
17	phq4_cats	PHQ-4 categories	discrete	numeric-9.0	15610	99	-
18	pandemic_..	Change in community wellbeing since March 2020	discrete	numeric-15.0	11073	4636	-
19	econ_sta_..	Change in economic stability during Covid	discrete	numeric-19.0	11073	4636	-
20	kzn_7day_..	Past 7-day KZN case count per 1000 population	continuous	numeric-9.0	15709	0	-
21	exposure_..	Future COVID infection concern level	discrete	numeric-12.0	15677	32	-
22	covid_kn_..	Knows someone who has had COVID	discrete	numeric-9.0	15658	51	-
23	stigma_s_..	Covid stereotype stigma score, z-scores	continuous	numeric-9.0	11073	4636	-
24	stigma_a_..	Covid anticipated stigma score, z-scores	continuous	numeric-9.0	11073	4636	-
25	cumul_co_..	Any other household members vaccinated up to today	discrete	numeric-9.0	15709	0	-

File covid_vax_plos_minimum_dataset							
#	Name	Label	Type	Format	Valid	Invalid	Question
26	covax_an_..	Binary outcome, any negative response	discrete	numeric-9.0	15709	0	-
27	covax_ne_..	Binary outcome, definite negative response	discrete	numeric-9.0	15709	0	-

Variables Description

Dataset contains 27 variable(s)

File : covid_vax_plos_minimum_dataset

IntId: Unique or Internal ID of Individual

Information	[Type= continuous] [Format=numeric] [Range= 17-258195] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-] [Invalid=0 /-] [Mean=77973.106 /-] [StdDev=56476.976 /-]

HHIntId: Household ID

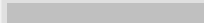
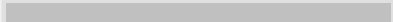
Information	[Type= continuous] [Format=numeric] [Range= 11-41783] [Missing=*]
Statistics [NW/ W]	[Valid=15369 /-] [Invalid=340 /-] [Mean=12566.992 /-] [StdDev=10594.969 /-]

InterviewDate: Interview Date

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-]

Sex: Sex


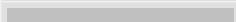
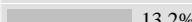
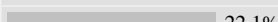
Information	[Type= discrete] [Format=numeric] [Range= 1-2147483642] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Male	5367	 34.2%
2	Female	10342	 65.8%
2147483642	Unknown	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.

agegroup: Age of respondent (4 levels)


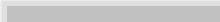
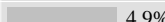

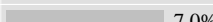
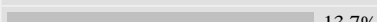

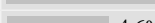

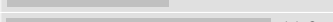
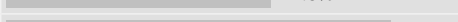
Information	[Type= discrete] [Format=numeric] [Range= 1-4] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	18-34	6394	 40.7%
2	35-49	3769	 24.0%
3	50-59	2078	 13.2%
4	60+	3468	 22.1%

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.

InterviewMonth: Month of interview

Information	[Type= discrete] [Format=numeric] [Range= 4-16] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
4	Apr 2021	449	 2.9%
5	May 2021	1539	 9.8%
6	Jun 2021	773	 4.9%
7	Jul 2021	0	
8	Aug 2021	587	 3.7%
9	Sep 2021	1107	 7.0%
10	Oct 2021	2159	 13.7%
11	Nov 2021	2028	 12.9%
12	Dec 2021	718	 4.6%
13	Jan 2022	1137	 7.2%
14	Feb 2022	1859	 11.8%
15	Mar 2022	2704	 17.2%

File : covid_vax_plos_minimum_dataset

InterviewMonth: Month of interview

Value	Label	Cases	Percentage
16	Apr 2022	649	<div><div></div></div> 4.1%

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.

covax_willing: Covid vaccine uptake or willingness to take (5 levels)

Information	[Type= discrete] [Format=numeric] [Range= 0-4] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Already vaccinated	3854	<div><div></div></div> 24.5%
1	Definitely would	5115	<div><div></div></div> 32.6%
2	Probably would	4130	<div><div></div></div> 26.3%
3	Probably would not	1366	<div><div></div></div> 8.7%
4	Definitely would not	1244	<div><div></div></div> 7.9%

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.

info_tradition: Information from traditional sources

Information	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
Statistics [NW/ W]	[Valid=11075 /-] [Invalid=4634 /-]

Value	Label	Cases	Percentage
0	No	197	<div><div></div></div> 1.8%
1	Yes	10878	<div><div></div></div> 98.2%
Sysmiss		4634	

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.

info_personal: Information from personal network sources

Information	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
Statistics [NW/ W]	[Valid=11075 /-] [Invalid=4634 /-]

Value	Label	Cases	Percentage
0	No	10548	<div><div></div></div> 95.2%
1	Yes	527	<div><div></div></div> 4.8%
Sysmiss		4634	

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.

info_healthcare: Information from healthcare sources

Information	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
Statistics [NW/ W]	[Valid=11075 /-] [Invalid=4634 /-]

Value	Label	Cases	Percentage
0	No	9335	<div><div></div></div> 84.3%
1	Yes	1740	<div><div></div></div> 15.7%
Sysmiss		4634	

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.

info_community: Information from community sources

Information	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
Statistics [NW/ W]	[Valid=11075 /-] [Invalid=4634 /-]

File : covid_vax_plos_minimum_dataset

info_community: Information from community sources

Value	Label	Cases	Percentage
0	No	10682	<div><div></div></div> 96.5%
1	Yes	393	<div><div></div></div> 3.5%
Sysmiss		4634	

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.

gov_mistrust_std: Mistrust in government - per standard deviation

Information	[Type= continuous] [Format=numeric] [Range= -2.32624650001526-2.47611045837402] [Missing=*]
Statistics [NW/ W]	[Valid=15664 /-] [Invalid=45 /-] [Mean=-0.00452 /-] [StdDev=1.001 /-]

educ_attain_cat: Highest educational attainment

Information	[Type= discrete] [Format=numeric] [Range= 1-5] [Missing=*]
Statistics [NW/ W]	[Valid=14455 /-] [Invalid=1254 /-]

Value	Label	Cases	Percentage
1	None	1540	<div><div></div></div> 10.7%
2	Primary	2193	<div><div></div></div> 15.2%
3	Some secondary	6332	<div><div></div></div> 43.8%
4	Completed secondary	3418	<div><div></div></div> 23.6%
5	Any tertiary	972	<div><div></div></div> 6.7%
Sysmiss		1254	

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.

HH_older: Has household member aged 60 or older

Information	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
Statistics [NW/ W]	[Valid=15369 /-] [Invalid=340 /-]

Value	Label	Cases	Percentage
0	No	6994	<div><div></div></div> 45.5%
1	Yes	8375	<div><div></div></div> 54.5%
Sysmiss		340	

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.

IsUrbanOrRural: Urbanicity of household

Information	[Type= discrete] [Format=numeric] [Range= 2-2147483625] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
2	Peri-Urban	3921	<div><div></div></div> 25.0%
3	Rural	10596	<div><div></div></div> 67.5%
4	Urban	1192	<div><div></div></div> 7.6%
2147483625	Unknown	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.

subUnits_num: Isigodi of residence

Information	[Type= discrete] [Format=numeric] [Range= 1-29] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Ebaswazini	1082	<div><div></div></div> 6.9%

File : covid_vax_plos_minimum_dataset

subUnits_num: Isigodi of residence

Value	Label	Cases	Percentage
2	Esiyembeni	421	2.7%
3	Gunjaneni	576	3.7%
4	KwaMsane (Isigodi)	1848	11.8%
5	KwaMsane (Township)	667	4.2%
6	Kwahoho	309	2.0%
7	Macambini	366	2.3%
8	Machibini	1193	7.6%
9	Madwaleni	2	0.0%
10	Mahunjini	216	1.4%
11	Makhambane	37	0.2%
12	Mapheleni	814	5.2%
13	Mfekayi	0	
14	Mshaya	343	2.2%
15	Mvutshini	68	0.4%
16	Myeki	1354	8.6%
17	Ndlovu Village	542	3.5%
18	Nkolokotho	1291	8.2%
19	Nkombose	1459	9.3%
20	Nkundusi	0	
21	Nomathiya	618	3.9%
22	Nompondo	0	
23	Nqopheni	314	2.0%
24	Nsolweni	664	4.2%
25	Ogengele	822	5.2%
26	Ophaphasi (Mpukunyoni)	428	2.7%
27	Ophondweni	275	1.8%
28	Qakwini	0	
29	Shikishela	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.

phq4_cats: PHQ-4 categories

Information	[Type= discrete] [Format=numeric] [Range= 0-3] [Missing=*]		
Statistics [NW/ W]	[Valid=15610 /-] [Invalid=99 /-]		
Value	Label	Cases	Percentage
0	Normal	10172	65.2%
1	Mild	2974	19.1%
2	Moderate	1992	12.8%
3	Severe	472	3.0%
Sysmiss		99	

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.

pandemic_change: Change in community wellbeing since March 2020

Information	[Type= discrete] [Format=numeric] [Range= 1-3] [Missing=*]		
Statistics [NW/ W]	[Valid=11073 /-] [Invalid=4636 /-]		

File : covid_vax_plos_minimum_dataset

pandemic_change: Change in community wellbeing since March 2020

Value	Label	Cases	Percentage
1	Got better	185	1.7%
2	Stayed the same	8850	79.9%
3	Got worse	2038	18.4%
Sysmiss		4636	

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.

econ_stability: Change in economic stability during Covid

Information	[Type= discrete] [Format=numeric] [Range= 1-5] [Missing=*]
Statistics [NW/ W]	[Valid=11073 /-] [Invalid=4636 /-]

Value	Label	Cases	Percentage
1	Much better off	67	0.6%
2	A little better off	406	3.7%
3	About the same	8792	79.4%
4	A little worse off	871	7.9%
5	Much worse off	937	8.5%
Sysmiss		4636	

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.

kzn_7day_incid_1000: Past 7-day KZN case count per 1000 population

Information	[Type= continuous] [Format=numeric] [Range= 0.0208416990935802-1.52049267292023] [Missing=*]
Statistics [NW/ W]	[Valid=15709 /-] [Invalid=0 /-] [Mean=0.227 /-] [StdDev=0.282 /-]

exposure_concern: Future COVID infection concern level

Information	[Type= discrete] [Format=numeric] [Range= 1-5] [Missing=*]
Statistics [NW/ W]	[Valid=15677 /-] [Invalid=32 /-]

Value	Label	Cases	Percentage
1	Not at all	3355	21.4%
2	Slightly concerned	6132	39.1%
3	Moderately concerned	2738	17.5%
4	Very concerned	3452	22.0%
5	Not applicable	0	
Sysmiss		32	

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.

covid_knowperson: Knows someone who has had COVID

Information	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
Statistics [NW/ W]	[Valid=15658 /-] [Invalid=51 /-]

Value	Label	Cases	Percentage
0	No	14744	94.2%
1	Yes	914	5.8%
Sysmiss		51	

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.

stigma_stereotype_std: Covid stereotype stigma score, z-scores

Information	[Type= continuous] [Format=numeric] [Range= -2.18814563751221-5.40825080871582] [Missing=*]
-------------	---

File : covid_vax_plos_minimum_dataset

stigma_stereotype_std: Covid stereotype stigma score, z-scores

Statistics [NW/ W] [Valid=11073 /-] [Invalid=4636 /-] [Mean=0.153 /-] [StdDev=0.814 /-]

stigma_anticipate_std: Covid anticipated stigma score, z-scores

Information [Type= continuous] [Format=numeric] [Range= -1.42416882514954-4.94502449035644] [Missing=*]

Statistics [NW/ W] [Valid=11073 /-] [Invalid=4636 /-] [Mean=0.0311 /-] [StdDev=0.937 /-]

cumul_covax_bin: Any other household members vaccinated up to today

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

Statistics [NW/ W] [Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	No	10904	69.4%
1	Yes	4805	30.6%

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.

covax_anyneg: Binary outcome, any negative response

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

Statistics [NW/ W] [Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	No	13099	83.4%
1	Yes	2610	16.6%

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.

covax_negative: Binary outcome, definite negative response

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

Statistics [NW/ W] [Valid=15709 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	No	14465	92.1%
1	Yes	1244	7.9%

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.