

**South Africa**

**Pillay, Deenan, Africa Health Research Institute**

**AHRI.PANGEA2 Vukuzazi Clinical Data.Release 2021-02**

**Study Documentation**

February 16, 2021

# Metadata Production

<b>Metadata Producer(s)</b>	Africa Health Research Institute (AHRI)
<b>Identification</b>	DDI.PANGEA2.Vukuzazi.ClinicalData.202102

# Table of Contents

<a href="#">Overview</a> .....	4
<a href="#">Scope &amp; Coverage</a> .....	4
<a href="#">Producers &amp; Sponsors</a> .....	4
<a href="#">Sampling</a> .....	4
<a href="#">Data Collection</a> .....	5
<a href="#">Accessibility</a> .....	5
<a href="#">Files Description</a> .....	6
<a href="#">AHRI.PANGEA2.VukuPANGEA2ClinicalData.2021</a> .....	6
<a href="#">Variables List</a> .....	7
<a href="#">AHRI.PANGEA2.VukuPANGEA2ClinicalData.2021</a> .....	7
<a href="#">Variables Description</a> .....	8
<a href="#">AHRI.PANGEA2.VukuPANGEA2ClinicalData.2021</a> .....	9

## AHRI.PANGEA2 Vukuzazi Clinical Data.Release 2021-02

Overview	
<b>Identification</b>	AHRI.PANGEA2.Vukuzazi.ClinicalData.202102
<b>Version</b>	V3.0.0
<b>Abstract</b> PANGEA is an international consortium of researchers based in Africa, the US and the UK studying transmission dynamics in HIV epidemics in sub-Saharan Africa. The goal of the consortium is to analyse HIV phylogenetic and demographic data to identify individual and population-level factors that drive the epidemic, analyse the dynamics of the epidemic and translate these findings into information that can be used to more effectively target interventions. PANGEA encompasses four analysis themes: (1) molecular epidemiology and mathematical modelling, (2) phylodynamics, (3) mobility and migration, and (4) clinical science, drug resistance and ethics. Social, Demographic and Clinical data will be used for these analysis.	
<b>Kind of Data</b>	Clinical and Demographic Data
<b>Unit of Analysis</b>	NA Samples are not sequenced at AHRI, but shipped to the UK for sequencing.

Scope & Coverage	
<b>Keywords</b>	Viremic Vukuzazi participants; date of sampling; phylodynamics, phylogeny, HIV-1, ART history
<b>Topics</b>	HIV-1; Incidence; Phylogeny; Epidemics; Population Surveillance; Rural Population; HIV Infections; Africa
<b>Time Period(s)</b>	2018-2020
<b>Countries</b>	South Africa
<b>Geographic Coverage</b> South Africa	
<b>Universe</b> All Vukuzazi participants in the intervention area, in Umkhanyakude district.	

Producers & Sponsors	
<b>Primary Investigator(s)</b>	Pillay, Deenan, Africa Health Research Institute
<b>Other Producer(s)</b>	Africa Health Research Institute (AHRI)
<b>Funding Agency/ies</b>	Bill & Melinda Gates Foundation (BMGF) , Genotyping funding source
<b>Other Acknowledgment(s)</b>	Anne Derache , Africa Health Research Institute Sweetness Dube , Africa Health Research Institute Tshwaraganang Modise , Africa Health Research Institute Dickman Gareta , Africa Health Research Institute

Sampling
<b>Sampling Procedure</b> HIV positive individuals within the Vukuzazi study, enrolled from June 1st 2018 to 31st May 2019, with a VL >100 copies/ml.

## Data Collection

<b>Data Collection Dates</b>	start 2018-05-25 end 2020-03-13
------------------------------	------------------------------------

## Accessibility

### Access Conditions

The representative of the Receiving Organization agrees to comply with the following conditions:

1. Access to the restricted data will be limited to the Lead Researcher and other members of the research team listed in this request.
2. Copies of the restricted data or any data created on the basis of the original data will not be copied or made available to anyone other than those mentioned in this Data Access Agreement, unless formally authorized by the Data Archive.
3. The data will only be processed for the stated statistical and research purpose. They will be used for solely for reporting of aggregated information, and not for investigation of specific individuals or organizations. Data will not in any way be used for any administrative, proprietary or law enforcement purposes.
4. The Lead Researcher must state if it is their intention to match the restricted microdata with any other micro-dataset. If any matching is to take place, details must be provided of the datasets to be matched and of the reasons for the matching. Any datasets created as a result of matching will be considered to be restricted and must comply with the terms of this Data Access Agreement.
5. The Lead Researcher undertakes that no attempt will be made to identify any individual person, family, business, enterprise or organization. If such a unique disclosure is made inadvertently, no use will be made of the identity of any person or establishment discovered and full details will be reported to the Data Archive. The identification will not be revealed to any other person not included in the Data Access Agreement.
6. The Lead Researcher will implement security measures to prevent unauthorized access to licensed microdata acquired from the Data Archive. The microdata must be destroyed upon the completion of this research, unless the Data Archive obtains satisfactory guarantee that the data can be secured and provides written authorization to the Receiving Organization to retain them. Destruction of the microdata will be confirmed in writing by the Lead Researcher to the Data Archive.
7. Any books, articles, conference papers, theses, dissertations, reports, or other publications that employ data obtained from the Data Archive will cite the source of data in accordance with the citation requirement provided with the dataset.
8. An electronic copy of all reports and publications based on the requested data will be sent to the Data Archive.
9. The original collector of the data, the Data Archive, and the relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.
10. This agreement will come into force on the date that approval is given for access to the restricted dataset and remain in force until the completion date of the project or an earlier date if the project is completed ahead of time.
11. If there are any changes to the project specification, security arrangements, personnel or organization detailed in this application form, it is the responsibility of the Lead Researcher to seek the agreement of the Data Archive to these changes. Where there is a change to the employer organization of the Lead Researcher this will involve a new application being made and termination of the original project.
12. Breaches of the agreement will be taken seriously and the Data Archive will take action against those responsible for the lapse if willful or accidental. Failure to comply with the directions of the Data Archive will be deemed to be a major breach of the agreement and may involve recourse to legal proceedings. The Data Archive will maintain and share with partner data archives a register of those individuals and organizations which are responsible for breaching the terms of the Data Access Agreement and will impose sanctions on release of future data to these parties.

### Citation Requirements

Gareta, D., Dube, S. and Modise, T. (2021) "AHRI.PANGEA2 Vukuzazi Clinical Data.Release 2021-02." Africa Health Research Institute (AHRI). doi: 10.23664/AHRI.PANGEA2.VUKUZAZI.CLINICALDATA.202102

# Files Description

Dataset contains 1 file(s)

<b>AHRI.PANGEA2.VukuPANGEA2ClinicalData.2021</b>	
<b># Cases</b>	562
<b># Variable(s)</b>	24

# Variables List

Dataset contains 24 variable(s)

File AHRI.PANGEA2.VukuPANGEA2ClinicalData.2021							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	<a href="#">study_id</a>	study_id	continuous	numeric.0	562	0	-
2	<a href="#">sex</a>	sex	discrete	character-1	562	0	-
3	<a href="#">dob</a>	dob	discrete	character-10	562	0	-
4	<a href="#">study</a>	study	discrete	character-6	562	0	-
5	<a href="#">sampled_..</a>	sampled_date	discrete	character-10	562	0	-
6	<a href="#">parent_t..</a>	parent_tube	discrete	character-14	562	0	-
7	<a href="#">aliquot_id</a>	aliquot_id	discrete	character-13	562	0	-
8	<a href="#">sample_t..</a>	sample_type	discrete	character-10	562	0	-
9	<a href="#">sample_v..</a>	sample_volume	continuous	numeric.0	562	0	-
10	<a href="#">sample_u..</a>	sample_units	discrete	character-2	562	0	-
11	<a href="#">knows_st..</a>	knows_status	discrete	character-3	562	0	-
12	<a href="#">hiv_status</a>	hiv_status	discrete	character-8	562	0	-
13	<a href="#">last_neg..</a>	last_neg_dt	discrete	character-10	195	0	-
14	<a href="#">had_hiv_..</a>	had_hiv_pos_result	discrete	character-3	489	0	-
15	<a href="#">first_po..</a>	first_pos_dt	discrete	character-10	445	0	-
16	<a href="#">on_art</a>	on_art	discrete	character-7	562	0	-
17	<a href="#">art_star..</a>	art_start_dt	discrete	character-10	250	0	-
18	<a href="#">latest_a..</a>	latest_art_drug_regimen	discrete	character-20	223	0	-
19	<a href="#">latest_a..</a>	latest_art_visit_date	discrete	character-10	250	0	-
20	<a href="#">latest_a..</a>	latest_art_visit_type	discrete	character-18	250	0	-
21	<a href="#">cd4_count</a>	cd4_count	continuous	numeric.0	558	4	-
22	<a href="#">vl_result</a>	vl_result	continuous	numeric.0	562	0	-
23	<a href="#">is_pregn..</a>	is_pregnant	discrete	character-16	326	0	-
24	<a href="#">tb_ever</a>	tb_ever	discrete	character-3	537	0	-

# Variables Description

Dataset contains 24 variable(s)

## File : AHRI.PANGEA2.VukuPANGEA2ClinicalData.2021

### # study\_id: study\_id

**Information** [Type= continuous] [Format=numeric] [Range= 6796-38745] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-] [Mean=28830.1491683328 /-] [StdDev=6126.71179194368 /-]

### # sex: sex

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # dob: dob

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # study: study

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # sampled\_date: sampled\_date

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # parent\_tube: parent\_tube

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # aliquot\_id: aliquot\_id

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # sample\_type: sample\_type

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # sample\_volume: sample\_volume

**Information** [Type= continuous] [Format=numeric] [Range= 500-1000] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-] [Mean=694.840010473472 /-] [StdDev=244.054000402432 /-]

### # sample\_units: sample\_units

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # knows\_status: knows\_status

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # hiv\_status: hiv\_status

**Information** [Type= discrete] [Format=character] [Missing=\*]

**Statistics [NW/ W]** [Valid=562 /-] [Invalid=0 /-]

### # last\_neg\_dt: last\_neg\_dt

**Information** [Type= discrete] [Format=character] [Missing=\*]

**File : AHRI.PANGEA2.VukuPANGEA2ClinicalData.2021****# last\_neg\_dt: last\_neg\_dt**

Statistics [NW/ W]	[Valid=195 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# had\_hiv\_pos\_result: had\_hiv\_pos\_result**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=489 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# first\_pos\_dt: first\_pos\_dt**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=445 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# on\_art: on\_art**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=562 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# art\_start\_dt: art\_start\_dt**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=250 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# latest\_art\_drug\_regimen: latest\_art\_drug\_regimen**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=223 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# latest\_art\_visit\_date: latest\_art\_visit\_date**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=250 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# latest\_art\_visit\_type: latest\_art\_visit\_type**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=250 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# cd4\_count: cd4\_count**

Information	[Type= continuous] [Format=numeric] [Range= 4-1749] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=558 /-] [Invalid=4 /-] [Mean=439.218992054272 /-] [StdDev=281.992016953344 /-]
--------------------	---

**# vl\_result: vl\_result**

Information	[Type= continuous] [Format=numeric] [Range= 104-6715552] [Missing=*]
-------------	--

Statistics [NW/ W]	[Valid=562 /-] [Invalid=0 /-] [Mean=96459.9859904512 /-] [StdDev=480347.162673152 /-]
--------------------	---

**# is\_pregnant: is\_pregnant**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=326 /-] [Invalid=0 /-]
--------------------	-------------------------------

**# tb\_ever: tb\_ever**

Information	[Type= discrete] [Format=character] [Missing=*]
-------------	---

Statistics [NW/ W]	[Valid=537 /-] [Invalid=0 /-]
--------------------	-------------------------------