

# India - Vadu HDSS INDEPTH Core Dataset 2009-2015 (Release 2017)

**Dr. Sanjay Juvekar (Founding Co Investigator and presently Investigator: 2002 to date) - VADU HDSS, KEMHRC (IN021), Dr. Siddhivinayak Hirve (Founding Investigator: from 2002-2009) - VADU HDSS, KEMH**

Report generated on: June 30, 2017

Visit our data catalog at: <http://www.indepth-ishare.org/index.php>

## Overview

### Identification

---

**ID NUMBER**

INDEPTH.IN021.CMD2015.v1

### Version

---

**VERSION DESCRIPTION**

CMD2015.v1: Edited dataset for public distribution

**PRODUCTION DATE**

2017-05-31

### Overview

---

**ABSTRACT**

Vadu Rural Health Program, KEM Hospital Research Centre Pune has a rich tradition in health care and development being in the forefront of needs-based, issue-driven research over almost 35 years. During the decades of 1980 and 1990 the research at Vadu focused on mother and child with epidemiological and social science research exploring low birth weight, child survival, maternal mortality, safe abortion and domestic violence. The research portfolio has ever since expanded to include adult health and aging, non-communicable and communicable diseases and to clinical trials in recent years. It started with establishment of Health and Demographic Surveillance System at Vadu (HDSS Vadu) in August, 2002 that seeks to establish a quasi-experimental design setting to allow evaluation of impact of health interventions as well as monitor secular trends in diseases, risk factors and health behavior of humans.

Vadu has struck research alliances with Regional, National and International Organizations. The iSHARE ([www.indepth-ishare.org](http://www.indepth-ishare.org)) - is a novel globally applicable solution from Vadu for enabling INDEPTH Network members to share their data on the web. IMVAC is an international collaboration with the Chest Research Foundation, Pune and the Imperial College, London to better understand COPD in developing world context. The Influenza Disease Burden in India (IDBI) study, in partnership with the National Institute of Virology, Pune and CDC Atlanta, is well placed to follow the influenza in rural areas. Vadu's foray into large community based GCP compliant phase II/III vaccine trials for meningitis, measles, Hib, typhoid and rota-virus is another important step in getting safe, effective and yet less expensive/affordable indigenously produced vaccines to licensure and use in India and globally. A recent strategic 5-year award jointly with paediatrics department of KEM, from the Dept. of Biotechnology, GOI is helping us build capacity to develop a Core Vaccine Research Unit at Vadu.

An exciting array of Translational Research (G2P studies) is ongoing at Vadu in collaboration with Institute of Genomics & Integrative Biology, GOI from both Ayurvedic and disease perspective. Vadu has extensively contributed to Project Ayurgenomics: Integration of Ayurveda to Genomics for field based research. 10000 volunteers from Vadu area participated in this study. Vadu has also developed a detailed phenotyping (DP) unit and molecular biology laboratory for primary investigations. The DP unit includes State of the art labs for heart rate variability testing, spirometry, skin testing, anthropometry and gustatory testing. Probably the only molecular biology lab in rural area, our lab is well equipped for RNA and DNA isolation and detection besides a dedicated fully equipped section for the clinical trials laboratory work.

Today's strength of Vadu is its ability to integrate research among and rational medical service delivery to the populations with support from the community and with capacity building of this very community in conducting community or hospital based research. Vadu strives to address its research needs through partnerships and collaborations with national and international institutions guided by principles of interdependence, equality, technology transfer and joint data ownership as also authorship. Above all, we believe in ethics and practice in conducting utmost ethical research.

We independently track vital events including births, deaths, causes of death, pregnancy outcome, migrations and marital status changes and serves to monitor trends in fertility and mortality inter alia over time.

The word "Demographic Surveillance" means to keep close track of the population dynamics. Vadu HDSS deals with keeping track of health issues and demographic changes in Vadu rural health program (VRHP) area. It is one of the most promising projects of National relevance that aims at establishing a quasi-experimental intervention research setting with objectives.

1. To create a longitudinal data base for

- a. Efficient service delivery
  - b. Future research
  - c. Linking all past micro-studies in Vadu area
2. Monitoring trends in public health problems
  3. Keeping track of population dynamics
  4. Evaluation of intervention services

**KIND OF DATA**

Event history data

**UNITS OF ANALYSIS**

Individual

## Scope

---

**NOTES**

The Scope of demographic surveillance includes demographic characteristics, households, living conditions, social security, access to facilities and services, hygiene, education levels, employment status, migration, fertility, mortality, etc

- Household: Household characteristics, household listing, occupation, education and maternal mortality.
- Individual: Individual listing, education, occupation, marital status and movements.
- Women: Pregnancy monitoring

**TOPICS**

Topic	Vocabulary	URI
Demography [N01.224]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>
Emigration and Immigration [N01.224.625.350]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>
Mortality [N01.224.935.698]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>
Cause of Death [N01.224.935.698.100]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>
Birth Rate [N01.224.935.849.500]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>

**KEYWORDS**

Fertility, Mortality, Migration, Population

## Coverage

---

**GEOGRAPHIC COVERAGE**

Vadu HDSS falls in two administrative blocks namely: (1) Shirur and (2) Haweli of Pune district in Maharashtra in western India.

It covers an area of approximately 232 square kilometers

**UNIVERSE**

Vadu HDSS covers as many as 50,000 households having 140,000 population spread across 22 villages. This dataset contains the events of all individuals ever resident during the study period (1 Jan 2009 to 31 Dec 2015)

## Producers and Sponsors

---

**PRIMARY INVESTIGATOR(S)**

Name	Affiliation
Dr. Sanjay Juvekar (Founding Co Investigator and presently Investigator: 2002 to date)	VADU HDSS, KEMHRC (IN021)
Dr. Siddhivinayak Hirve (Founding Investigator: from 2002-2009)	VADU HDSS, KEMHRC (IN021)

**OTHER PRODUCER(S)**

Name	Affiliation	Role
The International Network for the Demographic Evaluation of Populations and Their Health, Accra Ghana		Offered technical support through in establishment of Vadu HDSS and its full membership to Vadu HDSS.
Bharat Chaudhary	VADU HDSS, KEMHRC Pune (IN021)	Manager, Vadu HDSS
Field Research Assistants/ Field Research Assistants/ Field Research Assistants / Supervisors	VADU HDSS, KEMHRC Pune (IN021)	Data Collection

**FUNDING**

Name	Abbreviation	Role
Self Funded		

**OTHER ACKNOWLEDGEMENTS**

Name	Affiliation	Role
Dr Sanjay Juvekar	Vadu HDSS, KEMHRC Pune	Site Leader
Dr. Siddhi Vinayak Hirve	Vadu HDSS, KEMHRC Pune	Conceptualized formation and establishment of Vadu HDSS. Also offered seed money from his Fellowship received from Bill and Milinda Gates Institution of Population and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, USA
Dr Laila Garda	KEMHRC Pune	Advise and Support
Dr. V S Padbidri	KEMHRC, Pune	Advise and Support
Tathagata Bhattacharjee	INDEPTH Network and Vadu HDSS, KEMHRC Pune	Software development and Senior Data Manager wef 2012
Sandeep Bhujbal	INDEPTH Network and Vadu HDSS, KEMHRC Pune	IT Support and Data Manager wef 2013
Bharat Chaudhary	Vadu HDSS	Field Coordination, Vadu HDSS
Nidhi Patharia	INDEPTH Network and Vadu HDSS, KEMHRC Pune	Vadu HDSS website management and Data Manager wef 2012 - 2015
Neeraj Kashyap	Vadu HDSS, KEMHRC Pune	Data management software development (from 2003 to 2012)
Suhas Bhalerao	Vadu HDSS, KEMHRC Pune	IT Hardware and Networking Support (wef 2016)
Anant Shinde	Vadu HDSS, KEMHRC Pune	IT Hardware and Networking Support (Upto 2015)
Pallavi Lele	Vadu HDSS, KEMHRC Pune	Advise
Dr. Sudipto Roy	Vadu HDSS, KEMHRC Pune	Advise
Rajlaxmi Rangan	KEMHRC, Pune	Administration and Finance
Ramesh Pardeshi	Vadu HDSS, KEMHRC Pune	Site Administration
Pandurang Jadhav	Vadu HDSS	Human Resource support, Vadu HDSS
Hanif Shaikh	Vadu HDSS, KEMHRC Pune	
Dr. Girish Dayma	Vadu HDSS, KEMHRC Pune	

Name	Affiliation	Role
Dr. Ankita Srivastava	Vadu HDSS, KEMHRC Pune	
Rutuja Patil	Vadu HDSS, KEMHRC Pune	
Veena Muralidharan	Vadu HDSS, KEMHRC Pune	
Sagar Patil		Electronic data collection supervision (2012-2013)
Padma Sambhudas (Khuha)	I2IT, Pune	Data management software development (from 2003 to 2012)
Somnath Sambhudas	Vadu HDSS, KEMHRC Pune	Data Processing (Upto 2012)
Trupti Varpe	Vadu HDSS, KEMHRC Pune	Data management software development (from 2003 to 2012)

## Metadata Production

---

### METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
iSHARE2 Technical Team	iS2TT	INDEPTH Network	Documentation of the study
INDEPTH Network	int.INDEPTH	INDEPTH Network	Agency
Tathagata Bhattacharjee	TB	INDEPTH Network - IN021	DDI Author

### DATE OF METADATA PRODUCTION

2017-06-30

### DDI DOCUMENT VERSION

CMD2015.V1: 2017-06-30

### DDI DOCUMENT ID

DDI.INDEPTH.IN021.CMD2015.v1

## Sampling

### Sampling Procedure

---

Vadu area including 22 villages in two administrative blocks is the study area. This area was selected as this is primarily coverage area of Vadu Rural Health Program which is in function since more than four decade. Every individual household is included in HDSS. There is no sampling strategy employed as 100% population coverage in the area is expected.

### Response Rate

---

On an average the response rate is 99.99% in all rounds over the years.

### Weighting

---

Not applicable

# Questionnaires

## Overview

---

Language of communication is in Marath or Hindi. The form labels are multilingual - in English and Marathi, but the data entered through the forms are in English only.

- 1) Field Worker Checklist Form - The checklist provides a guideline to ensure that all the households are covered during the round and the events occurred in each household are captured.
- 2) Enumeration Form: To capture the population details at the start of the HDSS or any addition of villages afterwards.
- 3) Pregnancy Form: To capture pregnancy details of women in the age group 15 to 49.
- 4) Birth Form: To capture the details of the birth events.
- 5) Immigration Form: To capture inward population movement from outside the HDSS area and also for movement within the HDSS area.
- 6) Outmigration Form: To capture outward population movement from inside the HDSS area and also for movement within the HDSS area.
- 7) Death Form: To capture death events.

## Data Collection

### Data Collection Dates

Start	End	Cycle
2009-01-01	2015-12-31	Release Coverage

### Time Periods

Start	End	Cycle
2002-10-15		Round 1 - Baseline
2004-01-01		Round 2
2004-07-01		Round 3
2005-01-01		Round 4
2005-07-01		Round 5
2006-01-01		Round 6
2006-07-01		Round 7
2007-01-01		Round 8
2007-07-01		Round 9
2008-01-01		Round 10
2008-07-01		Round 11
2009-01-01		Round 12
2009-07-01		Round 13
2010-01-01		Round 14
2010-07-01		Round 15
2011-01-01		Round 16
2011-07-01		Round 17
2012-01-15		Round 18
2012-07-01		Round 19
2013-01-01		Round 20
2013-07-01		Round 21
2014-01-01		Round 22
2014-07-01		Round 23
2015-01-01		Round 24
2015-07-01		Round 25
2016-01-01		Round 26
2016-07-01		Round 27

### Data Collection Mode

Proxy Respondent [proxy]

#### DATA COLLECTION NOTES

##### 1) Training of Fieldworkers:

Training is planned and executed at the centre ie at Vadu HDSS every six months before the data collection round begins. Five days training session is conducted and every training session is evaluated by both the trainees and trainers. Trainers complete daily evaluation by asking questions in order to identify problems that trainees had experienced during that particular day's training. Areas that needed re- training are revisited the following day. At the end of the training mock interviews are conducted by the trainees.

##### Survey:

The data collection approach revolves around the use of a team of Field Research Assistants and Field Research Supervisors. The team is assigned a fixed number of EAs to enumerate. The team works together in each sampled EA and moves to the next one, once the targeted EA had been completed. The advantage of this method is that the supervisor is in daily in contact with the team, which improves the quality of the data collection during fieldwork.

Supervisors go to the FRA who have completed the interview and immediately check the questionnaire for errors, consistency and completeness. Where errors are found, the FRA is sent back to the household to correct the information that



had been recorded previously. If the supervisors are satisfied, they sign off the questionnaire and store it in a safe place. Supervisors do the same for all the members of their team until the EA had been completed. The team then moves to another selected EA.

The interview is conducted face - to -face which takes approximately 15 - 30 minutes each and this entire process takes around 5 months for every round.

2) Team Size: 20 - 40

3) Language: Marathi & Hindi

## Data Collectors

<b>Name</b>	<b>Abbreviation</b>	<b>Affiliation</b>
Bharat Chaudhary	BC	Vadu HDSS, KEMHRC (IN021)
Vadu HDSS Field Research Assistants	FRA	Vadu HDSS, KEMHRC (IN021)
Vadu HDSS Field Research Supervisors	FRS	Vadu HDSS, KEMHRC (IN021)

### **SUPERVISION**

Field Research Assistants (FRA) collect data from the field on hard copies. There is one Field Research Supervisor (FRS) for every 5 to 6 FRA. Two Field Coordinators supervise over FRS. Besides the field staffs Vadu HDSS has supervisors in each of the departments including Data Quality Assurance and Data Management team. Entire HDSS is managed by a Project Manager under the leadership of the Center Leader.

# Data Processing

## Data Editing

---

Entered data undergo a data cleaning process. During the cleaning process all error data are either corrected in consultation with the data QC team or the respective forms are sent back to the field for re collection of correct data. Data editors have the access to the raw dataset for making necessary editing after corrected data are brought from the field.

For all individuals whose enumeration (ENU), Immigration (IMG) or Birth (BTH) have occurred before the left censoring date (2009-01-01) and have not outmigrated (OMG) or not died (DTH) before the left censoring date (2009-01-01) are included in the dataset as Enumeration (ENU) with EventDate as the left censored date (2009-01-01). But the actual date of observation of the event (ENU, BTH, IMG) is retained in the dataset as observation date for these left censored ENU events. The individual is dropped from the dataset if their end event (OMG or DTH) is prior to the left censoring date (2009-01-01)

## Other Processing

---

1. Preparation of list of households was automated using HDSS software built using PHP and MySQL
2. All the forms are manually checked to ensure all census forms were completed.
3. Data Entry was done manually using HDSS software built using PHP and MySQL. There are 5 Data Entry Operators.
4. Automated field validation was used on all the forms along with pre manual checking.
5. MySQL Database was used.
6. The ETL process is performed using Pentaho Data Integration Community Edition (PDI CE) - Kettle. All necessary validation checks, quality checks and transformations are designed and implemented using PDI and inconsistencies identified are eliminated from the data set and thus the core micro data set is produced.

## Data Appraisal

### Estimates of Sampling Error

Not Applicable

### Other forms of Data Appraisal

Data is cleaned to an acceptable level against the standard data rules using Pentaho Data Integration Community Edition (PDI CE) tool. After the cleaning process, quality metrics is as follows:

CentreId	MetricTable	QMetric	Illegal	Legal	Total	Metric	RunDate
IN021	MicroDataCleaned Starts	1	301112	301113	0		2017-05-31 20:06
IN021	MicroDataCleaned Transitions	0	667010	667010	0		2017-05-31 20:07
IN021	MicroDataCleaned Ends	301113					2017-05-31 20:07
IN021	MicroDataCleaned SexValues	29	666981	667010	0		2017-05-31 20:07
IN021	MicroDataCleaned DoBValues	575	666435	667010	0		2017-05-31 20:07

Note: Except lower under five mortality in 2012 and lower adult mortality among females in 2013, all other estimates are fairly within expected range. Data underwent additional review in terms of electronic data capture, data cleaning and management to look for reasons for lower under five mortality rates in 2013 and lower female adult mortality in 2013. The additional review returned marginally higher rates and this supplements the validity of collected data. Further field related review of 2012 and 2013 data are underway and any revisions to published data/figures will be shared at a later stage.

# File Description

# Variable List

**IN021.CMD2015.v1**

Content	This file contains INDEPTH Core microdataset. This file was generated using ETL through Pentaho Kettle.
Cases	692091
Variable(s)	14
Structure	Type: Keys: ()
Version	CMD2015.v1: Updated dataset for public distribution
Producer	Vadu HDSS
Missing Data	

**Variables**

ID	Name	Label	Type	Format	Question
V1	RecNr	RecNr	contin	numeric	
V2	CountryId	CountryId	discrete	numeric	
V3	CentreId	CentreId	discrete	character	
V4	IndividualId	IndividualId	contin	numeric	
V5	Sex	Sex	discrete	numeric	
V6	DoB	DoB	discrete	character	
V7	EventCount	EventCount	discrete	numeric	
V8	EventNr	EventNr	discrete	numeric	
V9	EventCode	EventCode	discrete	character	
V10	EventDate	EventDate	discrete	character	
V11	ObservationDate	ObservationDate	discrete	character	
V12	LocationId	LocationId	contin	numeric	
V13	MotherId	MotherId	contin	numeric	
V14	DeliveryId	DeliveryId	contin	numeric	



**RecNr (RecNr)**

File: IN021.CMD2015.v1

**Overview**

Type: Continuous	Valid cases: 692091
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-319766	Maximum: 692091
	Mean: 346046
	Standard deviation: 199789.6

**Description**

A sequential number uniquely identifying each record in the data file

**CountryId (CountryId)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 692091
Format: numeric	Invalid: 0
Decimals: 0	
Range: 356-356	

**Description**

ISO 3166-1 numeric code of the country in which the surveillance site is situated

**CentreId (CentreId)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 692091
Format: character	Invalid: 0
Width: 5	

**Description**

An identifier issued by INDEPTH to each member centre of the format CCCSS, where CCC is a sequential centre identifier and SS is a sequential identifier of the site within the centre in the case of multiple site centres

**IndividualId (IndividualId)**

File: IN021.CMD2015.v1

**Overview**

Type: Continuous	Valid cases: 692091
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-167625	Maximum: 302739
	Mean: 151415.8
	Standard deviation: 87451.4

**Description**

A number uniquely identifying all the records belonging to a specific individual in the data file. This number is not be the same as the identifier used by a contributing centre to identify the individual.

**Sex (Sex)**

File: IN021.CMD2015.v1



**Sex (Sex)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 692091
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-2	

**Description**

Sex of the individual. 1 for Male and 2 for Female

**DoB (DoB)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 691983
Format: character	Minimum: NaN
Width: 23	Maximum: NaN

**Description**

The date of birth of the individual. Format: YYYY/MM/DD

**EventCount (EventCount)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 692091
Format: numeric	Invalid: 0
Decimals: 0	
Range: 2-8	

**Description**

The total number of events associated with this individual in this data set

**EventNr (EventNr)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 692091
Format: numeric	Invalid: 0
Decimals: 0	
Range: 1-8	

**Description**

A number increasing from 1 to EventCount for each event record in order of event occurrence

**EventCode (EventCode)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 692091
Format: character	Invalid: 0
Width: 3	

**Description**

A code identifying the type of event that has occurred.

**EventDate (EventDate)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 692091
Format: character	Minimum: NaN
Width: 23	Maximum: NaN

**Description**

The date on which the event occurred. Format: YYYY/MM/DD

**ObservationDate (ObservationDate)**

File: IN021.CMD2015.v1

**Overview**

Type: Discrete	Valid cases: 692091
Format: character	Minimum: NaN
Width: 23	Maximum: NaN

**Description**

Date on which the event was observed (recorded), also known as surveillance visit date. Format: YYYY/MM/DD

**LocationId (LocationId)**

File: IN021.CMD2015.v1

**Overview**

Type: Continuous	Valid cases: 692091
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-60789	Maximum: 127874
	Mean: 63918.7
	Standard deviation: 36964

**Description**

Unique identifier associated with a residential unit within the site and is the location where the individual was or became resident when the event occurred. This identifier is not be the same as the identifier used internally by the contributing centre.

**MotherId (MotherId)**

File: IN021.CMD2015.v1

**Overview**

Type: Continuous	Valid cases: 34225
Format: numeric	Invalid: 657866
Decimals: 0	Minimum: 52
Range: 51-164853	Maximum: 302718
	Mean: 150347.7
	Standard deviation: 87104.4

**Description**

The IndividualId of the mother. Only provided for BTH events.

**DeliveryId (DeliveryId)**

File: IN021.CMD2015.v1

**Overview**

## **DeliveryId (DeliveryId)**

File: IN021.CMD2015.v1

Type: Continuous  
Format: numeric  
Decimals: 0  
Range: 1-8581

Valid cases: 34225  
Invalid: 657866  
Minimum: 1  
Maximum: 47  
Mean: 1.4  
Standard deviation: 1.1

### **Description**

The RecNr of the delivery event associated with this birth

