Sleman HDSS Guidebook

Volume 1:
Wave 1-5 (2015-2019)

UNIVERSITAS GADJAH MADA
FAKULTAS KEDOKTERAN,
KESEHATAN MASYARAKAT, DAN KEPERAWATAN

Sleman HDSS Guidebook Volume 1:

Wave 1-5 (2015-2019)

Septi Kurnia Lestari, S.Gz., M.Med.Sc.PH, Ph.D.
Tri Atmaja Huda, S.Stat.
Ratri Kusuma Wardani, S.Gz.
Kadharmestan Gilang Pratama, S.Stat.
Putri Tiara Rosha, SKM., MPH.
Fachriyan Rizal Maulana, S.Kom.
Rahayu Kia Sandi Cahaya Putri, S.Stat.
Nur Fadlilah Nurul Hikmah, S.Stat.

Sleman Health and Demographic Surveillance System (Sleman HDSS)
Faculty of Medicine, Public Health, and Nursing (FK-KMK)
Universitas Gadjah Mada
Yogyakarta, Indonesia
2023

Title: Sleman HDSS Guidebook Volume 1: Wave 1-5 (2015-2019)

Supervisor
dr. Ifta Choiriyyah, MSPH, Ph.D.
dr. Prima Dhewi Ratrikaningtyas, M.Biotech.

Authors
Septi Kurnia Lestari, S.Gz., M.Med.Sc.PH., Ph.D.
Tri Atmaja Huda, S.Stat.
Ratri Kusuma Wardani, S.Gz.
Kadharmestan Gilang Pratama, S.Stat.
Putri Tiara Rosha, SKM., MPH.
Fachriyan Rizal Maulana, S.Kom.
Rahayu Kia Sandi Cahaya Putri, S.Stat.
Nur Fadlilah Nurul Hikmah, S.Stat.

Design Cover and Illustration
Feby Nurul Wahyuni, S.Gz.

Layout
Septi Kurnia Lestari, S.Gz., M.Med.Sc.PH., Ph.D.
Florida Agata Suni, S.KM.,MPH

Proofreader
Naufal Farah Azizah, S.KM.
Sarah Gumush M.Sc.

Published by
Fakultas Kedokteran, Kesehatan Masyarakat, dan Keperawatan
(FK-KMK) UGM
Faculty of Medicine, Public Health, and Nursing (FK-KMK) UGM
Farmako Street, Sekip Utara, Yogyakarta
Yogyakarta, Indonesia 55281
Telp./Fax. (+62 274) 560300
Email: iro.fk@ugm.ac.id

Publish in collaboration with
Sleman Health and Demographic Surveillance System (HDSS)
Radiopoetro Building 1st Floor, Farmako Street, Sekip Utara, Yogyakarta 55281
Email: hdssjogja.fkkmk@ugm.ac.id | Website: https://hdss.fk.ugm.ac.id

This book can be cited as Lestari, S. K., Huda, T. A., Wardani, R. K., Pratama, K. G., Rosha, P. T., Maulana, F. R., Putri, R. K. S. C., & Hikmah, N. F. N. (2023). Sleman HDSS Guidebook Volume 1: Wave 1-5 (2015-2019). Fakultas Kedokteran, Kesehatan Masyarakat, dan Keperawatan UGM.

Table of Contents

List of Figures

List of Tables

Preface

The Sleman Health and Demographic Surveillance System (HDSS) is an ongoing longitudinal study conducted annually since 2015, with the aim of monitoring demographic, social, and health changes in the Sleman Regency, Special Region of Yogyakarta, Indonesia. It is funded by the Faculty of Medicine, Public Health, and Nursing (FK-KMK) at Universitas Gadjah Mada (UGM), Indonesia, in collaboration with the Sleman Regency government. HDSS Sleman has successfully followed over 5,000 households, encompassing nearly 20,000 household members, and collected over 17 million data cells during five years from 2015 to 2019. The primary objective of Sleman HDSS is to provide high-quality data that can serve as evidence-based policy support That can be utilised by a wide range of stakeholders, including academics, students, policymakers, and other interested parties.

This comprehensive guidebook offers insights into the research methods, data collection processes, modules and questionnaires used, data management practices, and data weighting techniques. It also provides in-depth information on each module implemented by Sleman HDSS, including any modifications made during the five data collection cycles. Additionally, the guidebook includes a compilation of the instruments employed in the data collection process.

The creation of this guidebook was prompted by the challenges faced by Sleman HDSS data users in accessing relevant information about Sleman HDSS data. By addressing frequently asked questions and incorporating information gleaned from inquiries, we have endeavoured to ensure that data consumers gain a better understanding of Sleman HDSS data and recognise the potential benefits offered by the Sleman HDSS dataset. We express our sincere gratitude to all individuals who contributed to the development and completion of this guidebook, as their support has been invaluable in bringing this publication to fruition.

Sleman, November 2023

Chairperson of Sleman HDSS

dr. Ifta Choiriyyah, MSPH, Ph.D

NIP. 111198412201101201

Sleman HDSS

The Health and Demographic Surveillance System (HDSS) is a surveillance system that periodically collects data on population transitions, health status, and social changes within a specific timeframe. In 2014, the Faculty of Medicine, Public Health, and Nursing at Universitas Gadjah Mada (UGM), supported by the Sleman Regency government, initiated the implementation of HDSS in Sleman Regency, known as Sleman HDSS.

The vision of Sleman HDSS is to support scientific development in the health sector through innovative, excellent, and high-quality, world-class research based on population data. To achieve this vision, Sleman HDSS has established the following missions:

1. Efficiently collecting high-quality population-based longitudinal data.
2. Supporting the sustainability of health research at UGM.
3. Supporting education and training activities in the field of population-based longitudinal study.
4. Encouraging health policy making based on scientific evidence.
5. Supporting scientific evidence-based community service activities.

Based on its vision and missions, Sleman HDSS aims to accomplish the following objectives:

1. Establish a surveillance site to gather data on demographics, epidemiology, ecological factors (risk factors), and health service utilization.
2. Foster networks and partnerships to support independent and sustainable research.
3. Generate quality demographic and health data as a foundation for further research, evidence-based policymaking, community service, and educational activities.
4. Organise activities that enhance education in the field of population-based longitudinal research.
5. Contribute to improving population health through evidence-based community service.

Sleman HDSS conducts annual data collection, with the first wave occurring in 2015. Over the course of 2015-2019, Sleman HDSS carried out five waves of data collection. In the initial wave, Sleman HDSS involved 5,147 selected households from 216 clusters (census blocks) in all sub-districts of the Sleman Regency. During this baseline survey, Sleman HDSS collected basic demographic information, including age, gender, education, occupation, health insurance ownership, births, and deaths.

In subsequent waves, Sleman HDSS updated demographic data and collected additional health-related information, including communicable and non-communicable diseases,

behavioural risk factors, and healthcare service utilisation. Alongside health and demographic surveys, Sleman HDSS conducts various supporting activities. These activities include community service, advocacy for the Sleman Regency government, and facilitating opportunities for researchers to conduct nested research and utilise HDSS data under relevant regulations.

Ethical Considerations

Ethical Clearance

Sleman HDSS data collection began after receiving approval from the Committee for Medical and Health Research Ethics at the Faculty of Medicine, Public Health, and Nursing, UGM. Annual continuing reviews are conducted, and the list of ethical reviews conducted on Sleman HDSS can be found in **Table 1**.

Informed Consent

Before conducting any data collection, Sleman HDSS requests the consent of the participants to take part in the survey. The participants are given detailed information about Sleman HDSS, their rights, and the potential benefits and risks of participating in the survey. If they agree to participate, they are asked to sign an informed consent form.

Table 1. List of ethical reviews conducted on Sleman HDSS

Release date	Type	Ethical clearance number
22 July 2014	Ethical clearance	KE/FK/842/EC
18 August 2015	Continuing review	KE/FK/1037/EC
19 October 2015	Amendment	KE/FK/1037/EC 18 August 2015
11 March 2016	Amendment	KE/FK/1037/EC 18 August 2015
27 April 2017	Ethical clearance	KE/FK/0492/EC/2017
19 May 2017	Amendment	KE/FK/0492/EC 27 April 2017
06 October 2017	Amendment	KE/FK/0492/EC 27 April 2017
12 February 2018	Amendment	KE/FK/0492/EC 27 April 2017
05 March 2018	Amendment	KE/FK/0492/EC 27 April 2017
04 May 2018	Continuing review	KE/FK/0434/EC/2018
26 June 2018	Amendment	KE/FK/0434/EC 4 May 2018

Release date	Type	Ethical clearance number
01 March 2019	Amendment Approval	KE/FK/0434/EC 4 May 2018
10 May 2019	Continuing review	KE/FK/0526/EC/2019
10 October 2019	Amendment	KE/FK/0526/EC 10 May 2019

* The text in bold indicates the primary ethical clearance issued annually.

Confidentiality

Sleman HDSS ensures the confidentiality of respondents' identities and data. The data released by Sleman HDSS undergoes processing to remove any personally identifiable information.

Research Method

Sleman HDSS is a prospective longitudinal survey involving a panel of households representing the population of Sleman Regency. In each wave, Sleman HDSS collects demographic and health information from these households. However, the composition of Sleman HDSS respondents is dynamic as family members may move out of the HDSS household or pass away. Additionally, individuals born or moving into the HDSS household are also considered Sleman HDSS respondents.

The eligibility criteria for Sleman HDSS respondents are residents who have lived or will reside in Sleman Regency continuously for six months. The minimum sample size required to describe the demographic and health conditions in Sleman Regency was calculated using **Formula 1**.

$$n\_h = (z\text{^}2\, r(1-r) f k)/(p\_n\, e\text{^}2\,)$$

(Formula 1.)

with:

n_h = number of selected households,
z = expected level of confidence, which is 95%, thus z=1.96,
r = key indicator estimation, calculated from the survey, which is infant mortality (in Special Region of Yogyakarta) rate of 25 per 1000 live births,
f = design effect (*deff*) sample, assumed to be 2.0,
k = the multiplier of the expected non-response rate of 10%,
p = the proportion of the target population to the total population, which is the basis for the r parameter, namely the proportion of children aged 0-4 years in Sleman Regency = 7.89%,
$\_n$ = the average household size (number of household members per household),

e = the achieved margin error of 9%.

Based on the calculation, 4,890 households (rounded up to 5,000 households) are required as samples. The sample frame utilised for this study was derived from the 2010 Population Census conducted by Statistics Indonesia, consisting of a list of census blocks and households. Employing a two-stage stratified random sampling design, census blocks were designated as the first stratum, while households served as the second stratum. The selection of census blocks for the sample frame were based on a stratification approach that considered the wealth index (determined through Polychoric Principal Component Analysis) within Sleman Regency and the region (urban or rural). This stratification was implemented to achieve a representative sample that accurately reflects the region's distribution of urban and rural areas.

To compile the household sample frame, the Sleman HDSS team utilised the updated list of the heads of households (Kepala keluarga-KK) derived from the 2010 Population Census (SP2010). Subsequently, the team verified the presence of each household within the selected census blocks by consulting local leaders at the hamlet (Rukun Warga-RW) and the neighbourhood (Rukun Tetangga-RT) levels.

Table 2. Sampling procedure

Steps	Unit	Number of strata unit *h*		Sampling method	Sampling probability	Sampling fraction
		Population	Sample			
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Census block	B_h	b_h	PPS-with replacement, size M_{hi}	M_hi/M_h	$b\_h\ M\_hi /M\_h$
2	Household	$M\_hi\text{^}up$	$\_m = 25$	Systematic random sampling	$1 /(M\_hi\text{^}up\ )$	$25 /(M\_hi\text{^}up\ )$

with:

Index *h*: 1, 2, 3, 4, 5, 6 (*wealth index* strata and urban/rural), *i*: 1, 2,…n_h (census block)

B_h the total population of the census block in the strata *h,*

b_h the number of census block samples at strata *h,*

M_h the number of households resulting from SP2010 in the strata *h,*

M_hi the number of households resulting from SP2010 in strata h of the census block *i,*

$M\_hi\text{^}up$ the number of households as a result of Sleman HDSS updated data strata *h* census block *i.*

The sampling was conducted according to the procedure outlined in **Table 2**. In the initial stage, census blocks were selected from the sample frame based on a probability proportional to size. This resulted in a total sample of 184 census blocks for urban areas and 32 census blocks for rural areas. Statistics Indonesia carried out this stage, providing a list of households in the selected census blocks to Sleman HDSS.

In the second stage, a maximum of 25 households were selected from the sample frame of households in 216 census blocks. Systematic random sampling was used, with implicit stratification based on the education levels of heads of households. The HDSS Sleman team performed the household sampling. During the baseline survey, HDSS Sleman collected data from 19,724 residents in 5,147 selected households across 216 clusters (census blocks) in all sub-districts of Sleman Regency. Some census blocks did not meet the maximum sample size of 25 households because many houses in those areas were used as boarding houses or home industries, making them ineligible as respondents for Sleman HDSS. Further details regarding Sleman HDSS' design and research methodology are available in Dewi, Choiriyyah (1).

Number of respondents

Table 3 illustrates the changes in the number of Sleman HDSS households in each data collection wave. The number of households decreased due to non-participation during the interview period, households dropping out, or household mergers. Households that could not be reached in the previous wave were included in the in-scope population for the next wave. Failed interviews were considered dropouts (**Table 4**) and were not revisited in subsequent waves.

Table 3. Number of Sleman HDSS Respondents

	Wave 1 (2015)	Wave 2 (2016)	Wave 3 (2017)	Wave 4 (2018)			Wave 5 (2019)		
				Panel 1	Panel 2	Total	Panel 1	Panel 2	Total
In scope population		5,147	5,041	4,873*	393	5,266	4,799	264	5,063
Interviewed	5,147	4,999*	4,761*	4,701*	264	4,965	4,648*	244	4,892
Cannot be met		42	109	99			155		
Drop-out		106	174	165			144		
Merged households		3	7	1			3		

	Wave 1 (2015)	Wave 2 (2016)	Wave 3 (2017)	Wave 4 (2018)			Wave 5 (2019)		
				Panel 1	Panel 2	Total	Panel 1	Panel 2	Total
Split households		-	3	83			90		
Dropout rate		2.06%	3.48%	3.47%			2.90%		
Wave Response rate		97.07%	94.31%	96.45%	67.18%	94.27%	94.98%	92.42%	96.56%
Retention rate panel 1		97.07%	92.36%	91.32%			90.25%	92.42%	

* Including merged households and split households.

In Wave 4, aside from revisiting the in-scope population, Sleman HDSS recruited a new sample (refreshment sample). The following protocol was adhered to for the addition of the new sample: households that could not be reached in Waves 2 and 3 were revisited in Wave 4. If a household refused to participate, it was replaced with a refreshment sample using the following protocol: starting from the house adjacent to the Sleman HDSS household to be replaced (if multiple houses, choose the one closest to the main front door), and if the replacement sample also refused, another house was selected in a clockwise order. This approach was implemented to adjust the number of Sleman HDSS samples closer to the required minimum sample size of 5,000. The refreshment sample numbers in each wave are shown in columns labelled as "Panel 2" in Table 3.

Table 4. Number of respondents by reasons for dropping out per wave in 2016-2019

	Wave 2 (2016)	Wave 3 (2017)	Wave 4 (2018)	Wave 5 (2019)
Passed away (single-person household)	14	15	9	22
Refused to participate	26	54	82	45
Moved out of study area	61	75	63	58
Unable to communicate due to health problems (single-person household)	5	21	8	11
Cannot be met		9	3	8
Total	106	174	165	144

Data Collection

Data collection for Sleman HDSS has been conducted annually since 2015. Data collection periods for the core modules, the individual panel, and verbal autopsy are provided in **Table**

5, **Table 6**, and **Table 7**, respectively. Every year, data collection activities begin with interviews for the core modules and the adult individual panel, followed by fieldwork for verbal autopsy.

Table 5. Sleman HDSS core module data collection periods between 2015-2019

Wave	Year	Start	End
1	2015	15 January 2015	28 April 2015
2	2016	1 February 2016	17 June 2016
3	2017	13 March 2017	22 May 2017
4	2018	26 February 2018	16 May 2018
5	2019	4 March 2019	29 May 2019

Table 6. Sleman HDSS individual panel module data collection periods between 2015-2019

Wave	Year	Start	End
1	2015	-	-
2	2016	-	-
3	2017	13 March 2017	22 May 2017
4	2018	26 February 2018	16 May 2018
5	2019	4 March 2019	29 May 2019

Table 7. HDSS Sleman verbal autopsy data collection periods between 2015-2019

Wave	Year	Start	End
1	2015	27 July 2015	5 August 2015
2	2016	08 October 2016	29 October 2016
3	2017	18 October 2017	4 December 2017
4	2018	25 October 2018	16 December 2018
5	2019	24 October 2019	18 November 2019

Questionnaire Module

The core modules encompass demographic modules, such as socio-demographic profiles of new household members (referred to as ART) and updates on socio-demographic profiles of existing household members (referred to as PART), alongside health-related modules, including infectious diseases (PM), healthcare facility utilisation (AKS), and mental health (KJ). Data collected from the individual panel primarily pertains to non-communicable diseases and

their associated risk factors. Verbal autopsy data are diligently gathered for each reported death during the core data collection process, utilising the WHO Verbal Autopsy Form. Sleman HDSS adheres to a predefined data collection schedule of each module, which was developed by the Sleman HDSS scientific advisory board (refer to **Table 8**). Furthermore, Table 9 presents a comprehensive list of modules employed in Sleman HDSS Wave 1-5.

Table 8. The list of questionnaire modules used in Sleman HDSS

Topic	Modul	Interval	2015	2016	2017	2018	2019	2020	2021
Households	Sosio-demographic	1 year	x	x	x	x	x	x	x
	Nutrition/Consumption	2 years	x	x		x			x
	Socioeconomic Status	3 years	x			x			
Maternal and Child Health, Reproductive Health	Reproductive Health	1 year		x		x	x	x	x
	Maternal and Child Health,	1 year		x		x	x	x	x
	Child Health (Vaccination, Breastfeeding and Complementary Food)			x			x		x
Infectious Diseases	Infectious Diseases	1 year		x	x	x	x	x	x
	Behavioral Factors (Mosquito Repellents & HIV/AIDS)			x					
Health Service	Health Service	3 years		x	x		x	x	x
Non-communicable Diseases	Non-communicable Disease (NCD)	3 years		x	x			x	
	Mental Health	5 years					x	x	x
	Quality of Life	5 years			x				
	Risk Factors of NCD				x	x		x	
	Injury	5 years		x	x		x		x
	Anthropometry* and blood pressure Measurement	1 year				x			
Cause of Death	Verbal Autopsy	1 year	x	x	x	x	x		x
COVID 19	Social economic impact							x	x
	Vaccine								x
Health Literacy	Health Literacy								x

Table 9. Sleman HDSS questionnaire module by data collection wave

Module Code	Module Name	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
	Cover	Yes	Yes	Yes	Yes	Yes
ART	Demographic data: Household member	Yes	Yes	Yes	Yes	Yes
ARTB	Demographic data: New household member		Yes	Yes	Yes	Yes
PART	Update demographic data: Household member		Yes	Yes	Yes	Yes
KSM	Household food and beverage consumption	Yes				
PRP	Average Household Expenditure	Yes				

Module Code	Module Name	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
HRT	Asset ownership and household facilities	Yes			Yes	
KR	House or residential building characteristics	Yes			Yes	
KLH	Birth	Yes				
KMT	Death	Yes				
KRP	Reproductive health (Parity)		Yes		Yes	Yes
PM	Infectious disease		Yes	Yes	Yes	Yes
PTM	Non-communicable disease		Yes	Yes		Yes
DIF	Disability		Yes			
HIV	Knowledge of HIV/AIDS		Yes			
MB	Unhealthy food		Yes			
INS	The behaviour of using mosquito repellents		Yes			
AKS	Health service		Yes	Yes		Yes
KH-SF12	Quality of life			Yes		
KJ	Mental health					Yes
PH	Hirschsprung disease					Yes
KAI	Maternal, child health, and immunisation		Yes		Yes	Yes
ASM	Breastfeeding and complementary feeding		Yes			Yes
PTMV2	Non-communicable disease/NCD (Individual panel)			Yes		
AGH	Anthropometry and risk factors of NCD				Yes	
KJS	Individual mental health (Individual panel)					Yes
VA	Verbal Autopsy (Cause of death)	Yes	Yes	Yes	Yes	Yes

Cover Module

Level	:	Household
Sampling target	:	All households
Interviewee	:	Main respondents or related household members
Waves	:	All waves
Question code	:	KL (location information), KRT (household information), KPD (data collection information)

The cover module encompasses essential information regarding location details (home address and coordinates), household information (head of household's name, main respondent's name, telephone number, number of household members), and data collection particulars (interviewer code, supervisor code, interview date). This module's questionnaire was derived from similar surveys or surveillance initiatives. The demographic data collection instrument utilised in Sleman HDSS is adapted from various sources, including the WHO Study on Global Ageing and Adult Health (SAGE) (2), Purworejo HDSS, Indonesian Demographic and Health Survey (3), and *Riset Kesehatan Dasar Indonesia* (Indonesian Basic Health Research) (4).

Demographic Data: Household Member Module

Level	:	Individual
Sampling target	:	All household members
Interviewee	:	Main respondents or related household members
Waves	:	All waves
Question code	:	ART

The definition of Sleman HDSS household members includes two criteria:

1. Individuals who have resided in the household for a minimum of six consecutive months and share or eat from the same kitchen. OR
2. Individuals who have lived in the household for less than six months but plan to remain in the household for at least the next six months and share or eat from the same kitchen.

The Sleman HDSS questionnaire employs two types of household member modules, i.e., for the new household members (referred to as ART) and for the existing household members (referred to as PART). The ART module focuses on collecting basic demographic information, such as gender, date of birth, marital status, education, occupation, religion, ethnicity, and ownership of health insurance. For existing respondents, any changes in their demographic data are captured using the PART module, including the status of the head of the family,

presence (present, deceased, or migrated), marital status, education, occupation, ownership of health insurance, and pregnancy outcome.

In the Sleman HDSS dataset, all variables from the ART and PART modules are stored with variable names starting with the code 'ART.' The household module is adapted from questionnaires used in similar surveys or surveillance, while the demographic data collection instrument draws from the WHO SAGE (2), Purworejo HDSS, Indonesian Demographic and Health Survey (3), and Indonesian Basic Health Research (4).

Pregnancy and Birth

Female household members aged 10-54 years are asked if they are pregnant (question art22). The outcomes of the pregnancies are inquired from members of the existing household in Wave 2. Two sets of questions are utilised: 1) for those previously recorded as pregnant in the previous wave (question art25), and 2) for those not recorded as pregnant (question art30).

Both sets of questions collect information on pregnancy outcomes (miscarriage, stillbirth, or live birth), delivery date, gestational age, birth attendant, method of delivery, and the newborn's length and weight. The difference between the two sets of questions is that the second set begins with a question asking whether the respondent was pregnant between the previous and the current data collections. There is a possibility that the respondent may have experienced pregnancy multiple times with merely one pregnancy being recorded by the Sleman HDSS. Consequently, respondents who were recorded as pregnant are also interviewed using the second set of questions.

In Wave 1 (2015), data on household members born within the two years prior to the interview were collected using the Birth module (variable code: klh). The collected data includes pregnancy outcomes, birth attendant, gender, date of birth, and the length and weight of the newborn. It is important to note that not all babies recorded in this module are Sleman HDSS respondents. Only babies born alive and meeting the criteria for household members receive a respondent's identity number (keyid).

Mortality

The mortality module (variable code: kmt) was utilised in the first wave to record deaths in Sleman HDSS households from January 2013 until the time of the first wave of interviews. The data collected includes name, gender, date of death, and age at death.

However, household members recorded in the kmt module are not considered part of the Sleman HDSS household since they do not meet the definition of a household member. Therefore, they were not assigned a respondent's identity number (keyid). Starting from the second wave, the incidence of death among Sleman HDSS household members is recorded in the ART module as one of the reasons why the respondent is no longer in the household.

Migration

The migration to and from Sleman HDSS households are recorded in the ART module. In-migration refers to new household members entering or becoming part of an existing household through migration or birth. In contrast, out-migration refers to household members who no longer reside in the household, possibly due to death or migration. Information on in-migration is obtained from the ART module, and all household members joining Sleman HDSS from the second wave onwards (excluding the refreshment sample) are counted as in-migration. Identifying out-migration occurs through questions about the presence of household members in the ART module.

Household Food and Beverages Consumption Module

Level	:	Household
Sampling target	:	Household
Interviewee	:	Main respondents
Wave	:	1
Question code	:	KSM

The Household Food and Beverages Consumption Module gathers data on the quantity of food consumed by all household members in the previous week. The recorded data represents the actual amount of food consumed by the household. This module utilises an instrument adapted from the 2014 *Survei Sosial Ekonomi Nasional 2014* (National Socio-Economic Survey)(5).

The food consumption recorded in this module only concerns the household's needs, including purchases, own production, and gifts received from others. However, it does not include food consumed for home business purposes, given to other parties, wasted, damaged, or leftover foods provided to pets. The dataset for the KSM module also includes derived variables that provide information on daily consumption and per capita consumption. Consumption per day is calculated by dividing the seven days of consumption in the previous week, while per capita consumption per day is obtained by dividing the daily consumption by the number of family members.

Average Household Expenditure Module

Level	:	Household
Sampling target	:	Household
Interviewee	:	Main respondents
Wave	:	1
Question code	:	PRP

The module on average household expenditure collects data on the average expenditure over the past six months for both food and non-food items. The instrument used to collect data on average expenditures is adapted from the 2014 National Socio-Economic Survey (5), as well as the 2013 Indonesian Basic Health Research Questionnaire (6). **Table 10** lists the categories of non-food consumption.

Table 10. Description of non-food expenditure

Type of expenditure	Details
Housing	All expenses related to building or renovating a house (e.g., painting, tiling, purchasing building materials, etc.).
Electricity	All expenditures for paying the electricity bill.
Water bill	All expenses related to drinking water or water for daily needs (bathing, washing, and cooking).
Education	Education-related expenses, including tuition fees, books, graduation fees, uniform fees, shoes, bags, extracurricular activity, academic course fees, monthly class dues, etc.
Clothes	Clear enough.
Health	Expenditures for medical purposes, buying medicines and paying for health insurance.
Party/social	Expenditures related to social activities, for example social gatherings, gifts, death donations, holiday allowance, and "*angpao*".
Transport	Expenditure related to transport including gas, bus, taxi, and motorcycle taxi.
Cigarette/tobacco	Clear enough.
Communication (phone credits)	Communication-related expenses such as phone credit, telephone subscriptions, and prepaid internet
Internet services	Internet provider subscription fees such as Telkom Speedy, Citranet, Jogja Medianet.
Housekeepers and/or babysitters	Monthly salary for housekeeper and/or babysitter.

Type of expenditure	Details
Others	Other expenses not mentioned above including soap, laundry soap, and toothpaste.

Asset Ownership and Household Facilities Module

Level	:	Household
Sampling target	:	Household
Interviewee	:	Main respondents
Wave	:	1&4
Question code	:	HRT

The asset module records various assets owned by Sleman HDSS households, including televisions, refrigerators, motorcycles, buses, livestock, and land ownership (such as rice fields, yards, and gardens). It also captures information on land use (cultivated or not) and whether crops are for self-consumption or sale. The instrument used for collecting data on household goods and facility ownership is adapted from the 2014 National Socio-Economic Survey (5) and the 2013 Indonesian Basic Health Research Questionnaire (4).

House or Residential Building Characteristic Module

Level	:	Household
Sampling target	:	Household
Interviewee	:	Main respondents
Wave	:	1&4
Question code	:	KR

The module on house characteristics collects information on home ownership status and the characteristics of the respondent's house, such as the type of floor, roof, frame, and wall. Additionally, it includes questions about the main sources of water and lighting. The instrument used for collecting data on house characteristics is adapted from the 2014 National Socio-Economic Survey (5) and the 2013 Indonesian Basic Health Research Questionnaire(4).

Reproductive Health Module (Parity)

Level	:	Individual
Sampling target	:	15-49-year-old household members (question on parity for female respondents only)
Interviewee	:	the relevant respondents
Wave	:	2, 4,&5
Question code	:	KRP

The module on births, miscarriages, and contraceptive use primarily focuses on two aspects: 1) the history of births and miscarriages (parity) and 2) the use of contraceptives. In Wave 2, parity information is collected from all female household members aged 15-49 years. Furthermore, in the fourth and subsequent waves, parity information is only collected from new female household members aged 15-49 years. The contraceptive section collects data on reproductive-age men and women in the second and fourth waves. It includes questions about the type of contraceptives used, sources of family planning services, service providers, and payment methods for contraceptive services. The contraceptive use module is adapted from the 2013 Basic Health Research Questionnaire (4)

Infectious Disease Module

Level	:	Individual and household
Sampling target	:	Household members
Interviewee	:	Main respondents and the relevant respondents
Wave	:	2, 3, 4, &5
Question code	:	PM

The module on infectious diseases captures information about the history of infectious diseases experienced by family members during a specified period. This module includes recording various infectious diseases such as upper respiratory infections (URI), malaria, dengue haemorrhagic fever, leptospirosis, and pulmonary tuberculosis. This module has household-level questions that serve as a filter to identify if at least one household member has been diagnosed with or suffered from any of the specified diseases at a particular time. If the respondent gives an affirmative answer to the household-level question, they are then asked to identify household members who had the disease. The infectious disease module is adapted from the 2013 Basic Health Research Questionnaire (4).

Non-Communicable Disease (NCD) Module

There are two types of non-communicable disease modules, i.e., regular module for household members and module for individual panel respondents.

Regular Module

Level	:	Individual and household
Sampling target	:	All household members
Interviewee	:	Main respondents and the relevant respondents
Wave	:	2 (complete regular module); 2,3,5 (regular module: injury)
Question code	:	PTM (regular module)

The regular module on non-communicable diseases focuses on capturing data regarding non-communicable diseases such as hypertension, diabetes mellitus, stroke, heart disease, cancer, asthma, and injuries. This module contains household-level questions that serve as a filter to identify if at least one household member has been diagnosed with or suffered from any of the specified diseases at a particular time. The next question asks the respondent to name any household members that suffer from the disease if they answered affirmatively to the household-level question. The non-communicable disease module is adopted from the 2013 Basic Health Research Questionnaire (4).

Individual Panel Module

Level	:	Individual
Sampling target	:	Individual panel respondent
Interviewee	:	Individual panel respondent
Wave	:	3
Question code	:	Q4

This module is adapted from the Chronic Conditions and Health Services Coverage module of the Study on Global Aging and Adult Health (SAGE) (2). It addresses non-communicable diseases, including stroke, angina, diabetes mellitus, chronic obstructive pulmonary disease (COPD), and hypertension. The module includes questions about whether the respondent has been diagnosed with the condition and whether they have ever had any signs or symptoms related to stroke, angina, or chronic obstructive pulmonary disease (COPD).

Anthropometry and Risk Factors of NCD Module

Level	:	Individual
Sampling target	:	Individual panel respondent
Interviewee	:	Individual panel respondent
Wave	:	3,4 (tobacco consumption); 4 (fruit and vegetable consumption, physical activity, waist circumference, hip circumference, and blood pressure measurement)
Question code	:	E (fruit and vegetable consumption), F (physical activity), G (tobacco consumption), H (waist circumference, hip circumference, and blood pressure measurement)

The module on non-communicable disease risk factors collects data on fruit and vegetable consumption, physical activity, tobacco consumption, blood pressure measurements, as well as hip and waist circumference measurements. This module is adapted from the WHO Stepwise Approach to Chronic Disease Risk Factor Surveillance (7). Anthropometric measurements use a measuring tape distributed by the Yogyakarta Nutritionist Association.

Blood pressure measurements are conducted using the OMRON Automatic Blood Pressure Monitor Series HEM-7130-L, HEM 7130, HEM 7203, and HEM 7200, manufactured by OMRON Healthcare Manufacturing Vietnam CO., LTD, Binh Duong Province, Vietnam. Food photo books from the Center for Applied Health Technology and Clinical Epidemiology, Health Research and Development Agency, and Ministry of Health of the Republic of Indonesia, were used to assist respondents in estimating vegetable and fruit consumption (8).

Disability Module

Level	:	Individual and household
Sampling target	:	Household members
Interviewee	:	Main Respondents
Wave	:	2
Question code	:	DIF

The module asks about specific conditions (speech impairment, quadriplegia, intellectual disability, deafness, and blindness) that affect members of households. These conditions are defined in Table 11. This module employs a household-level question as a filter to identify households where at least one member exhibits the specified condition. If the respondent gives an affirmative answer to the household-level question, they are then asked to identify household members who had the condition. The disability module utilised in this study is adapted from the 2013 Basic Health Research Questionnaire (4).

Table 11. Definition of types of disability

Type of disability	Definitions
Speech impairment	This is a form of impairment, denoting a condition wherein the dysfunction of the speech organs is attributable to disorders of the organs, including the throat, vocal cords, lungs, mouth, and tongue, either from birth, accidents, or illnesses (9, 10).
Physical disability	Physical disability is an absence of or significant difference in a person's body structure or function (11).
Intellectual disabilities	Intellectual Disabilities are disorders of thinking function or a person's ability to learn limitation due to below average intelligence levels, including slow learning, mental disabilities, and Down syndrome (9, 12).
Deafness	Deafness is a condition where an individual has no ability to hear at the same level as someone with typical hearing, characterized by hearing

Type of disability	Definitions
	thresholds of 20 dB or higher in both ears, and is referred to as having hearing impairment. This impairment can range from mild to profound, and may impact one or both ears, resulting in challenges in perceiving conversational speech or loud auditory stimuli (13).
Blindness	The term "blindness" encompasses individuals with both low vision and legal blindness (14). 1. Low vision refers to individuals whose visual impairment cannot be completely corrected with glasses, contact lenses, refractive surgery, or other surgical interventions. 2. Legal blindness is a condition that is met when an individual's central visual acuity, even with the best correction (using glasses or contact lenses) at a distance, is 20/200 or worse in the better-seeing eye. Alternatively, legal blindness can also be determined by the presence of visual field restriction, where the widest diameter is 20 degrees or less in the better-seeing eye.

Knowledge of HIV/AIDS Module

Level	:	Individual
Sampling target	:	Main Respondent
Interviewee	:	Main Respondents aged 18–24-year-old.
Wave	:	2
Question code	:	HIV

To assess respondents' general knowledge about HIV/AIDS, the following points are examined:

1. Condom usage for HIV prevention.
2. Being faithful reduce the risk of contracting HIV.
3. Modes of HIV transmission.
4. Ways to identify people with HIV.

The questionnaire for this module is adapted from the Ethiopia Demographic and Health Survey (15).

Unhealthy Food Module

Level	:	Individual
Sampling target	:	Main Respondent
Interviewee	:	Main Respondent
Wave	:	2
Question code	:	MB

This module focuses on gathering information regarding the frequency of consuming unhealthy foods, which are known to increase the risk of non-communicable diseases. Examples of unhealthy foods include sweet foods/beverages, salty foods, fatty/cholesterol-rich/fried foods, grilled foods, processed meat/chicken/fish with preservatives, foods high in monosodium glutamate (MSG), coffee, non-coffee caffeinated drinks, salted fish, and instant noodles. The module's structure is similar to the unhealthy food module used in the 2013 Basic Health Research in Indonesia (4).

Mosquito Repellent Usage Module

Level	:	Household
Sampling target	:	Household
Interviewee	:	Main Respondent
Wave	:	2
Question code	:	INS

This module aims to collect information about the usage of mosquito repellents by the Sleman HDSS households. The questions within this module address the frequency of using different mosquito repellents (such as sprays or electric devices) and when the household typically uses insect repellents. This module is adapted from the 2013 Basic Health Research Indonesia (4).

Access to Healthcare Services Module

Level	:	Individual
Sampling target	:	All household members (section 1) and main respondent (section 2)
Interviewee	:	Main Respondents or the relevant household member
Wave	:	section 1=2,3,5; section 2=2
Question code	:	AKS

This module comprises two sections. The first section includes questions about the respondent's medical history, primary healthcare facility, payment method, and reasons for not seeking treatment. The second section collects respondents' opinions on the quality, accessibility, and satisfaction levels of primary health centres (*Puskesmas*), public hospitals, private hospitals, and the National Health Insurance system (*Jaminan Kesehatan Nasional*).

In Wave 5, additional questions were introduced to confirm non-communicable diseases reported in the previous wave and to gather more detailed information about healthcare utilization. This includes the duration and cost of treatment, treatment frequency, and sources of financing for inpatient and outpatient conditions over the past 12 months. This module is adapted from the questionnaire used in the 2013 Basic Health Research in Indonesia (4).

Quality of Life Module

Level	:	Individual
Sampling target	:	Individual panel respondents
Interviewee	:	Individual panel respondents
Wave	:	3
Question code	:	K

To assess the respondents' quality of life over the past four weeks, we employ the SF-12v2 questionnaire, developed explicitly by Optum® Health Survey. This questionnaire aims to capture individuals' perspectives on their health and well-being. It consists of eight domains, each addressing different aspects (16):

1. Physical functioning (PF) - limitations in moderate physical activities.
2. Role-physical (RP) - limitations in work or other activities due to physical problems.
3. Bodily pain (BP) - interference of bodily pain with regular work.
4. General health (GH) - perception of overall health condition and illness resistance.
5. Vitality (VT) - feeling energised versus experiencing fatigue and lack of energy.
6. Social functioning (SF) - limitations in social activities due to physical or emotional issues.
7. Role emotional (RE) - limitations in usual role activities caused by emotional problems.
8. Mental health (MH) - general mental well-being, encompassing depression and anxiety.

Furthermore, these eight domains are combined into two component scores to assess the quality of life:

1. Physical Component Summary (PCS)
    a. Physical functioning (PF)
    b. Role-physical (RP)
    c. Bodily pain (BP)
    d. General health (GH)
2. Mental Component Summary (MCS)
    a. Vitality (VT)
    b. Social functioning (SF)

c. Role emotional (RE)
d. Mental health (MH)

To interpret the results, it is important to compare the average score of each component within specific age groups. The mean value varies across age groups. Above-average scores indicate better health status, while below-average scores suggest poorer health status. Quality of life is categorised as 'average,' 'below average,' or 'above average' based on these comparisons (16, 17).

Mental Health Module

There are two mental health modules: 1. Mental health of household members and 2. Self-reported mental health of the individual panel respondent.

Mental health in the households

Level	:	Households
Sampling target	:	All household members
Interviewee	:	Main respondents
Wave	:	5
Question code	:	JK

This module aims to identify cases of severe mental or psychiatric disorders within households. The main respondents are asked to provide information on whether any family members have ever been diagnosed with a mental disorder, including details about treatment history. The questionnaire used in this module is adapted from the 2018 Basic Health Research in Indonesia (6), which aims to estimate the prevalence of severe mental disorders such as depression, bipolar disorder, schizophrenia, and psychosis.

Self-reported mental disorder

Level	:	Individual
Sampling target	:	Individual panel respondents
Interviewee	:	Individual panel respondents
Wave	:	5
Question code	:	SRQ

The Self-Reported Questionnaire (SRQ 20) is a screening tool developed by the World Health Organization (WHO) to assess mental health problems in developing countries (18). It consists of approximately 20 questions about respondents' experiences with emotional and mental

disorders leading to mental distress (neurotic conditions). The Indonesian version of this tool is adapted from the 2013 Basic Health Research in Indonesia (4).

Under-Five Children Health Module

Level	:	Individual
Interviewee	:	Main caregivers
Wave	:	2&5
Question code	:	KAI (pregnancy record) and KIM (Immunisation record)

Anthropometric and immunisation record

Sampling Target	:	0-59-month-old household member

This module focuses on households with members aged 0-59 months. The primary caregiver is interviewed, and the questionnaire collects information on various aspects of infant and child health, including anthropometric measurements at birth (body length, weight, head circumference), history of body weight measurements, immunization history, reasons for incomplete vaccination, and vaccine side effects. The questionnaire used in this module is based on the 2013 Basic Health Research in Indonesia (4).

Information regarding immunization history, infant anthropometrics, and administration of vitamin A supplements is primarily obtained from the Maternal and Child Health book (*Buku Kesehatan Ibu Anak/KIA*). If respondents cannot present the KIA book or the information within it is incomplete, the data is recorded based on the respondent's memory.

The completeness and timeliness of vaccinations are assessed based on the child's age at the time of receiving each vaccine, as indicated in **Table 12**. In the fifth wave, additional information is collected on six non-compulsory immunization programmes, including influenza, rotavirus, and varicella.

Table 12. Type of vaccine

Type of vaccine	Details
Hepatitis B	The Hepatitis B vaccine is ideally administered within 12 hours of birth, preceded by an injection of vitamin K1. The vaccination can be carried out using either a monovalent Hepatitis B vaccine or a combination vaccine.
The bacille Calmette-Guérin (*BCG*) *vaccine*	The BCG vaccine is administered to protect infants from Tuberculosis Disease. It is most effective when given to infants aged 0-2 months. The BCG vaccine is administered subcutaneously in the right upper

Type of vaccine	Details
	arm. The resulting wounds and scars indicate the successful administration of the BCG vaccine.
Oral Polio Vaccine (OPV)	The Oral Polio Vaccine is administered to prevent polio in babies. This vaccine is given by administering drops into the baby's mouth. It is usually administered concurrently with the injected DPT-HB vaccine at the ages of two, four, and six months.
Inactivated Polio Vaccine (IPV)	The Inactivated Polio Vaccine (IPV) is typically administered at two months of age. It is injected into the right thigh simultaneously with the DPT/HB or DPT/HB/Hib injections.
Pentabio®/DPT-HB-Hib Vaccine	Combo vaccines, such as the DPT-HB-Hib vaccine, combine different vaccines to provide simultaneous protection against multiple diseases in a single shot. The DPT-HB vaccine is administered in three doses: DPT-HB1 at 2-3 months old, DPT-HB2 at 3-4 months old, and DPT-HB3 at 4-6 months old.
DPT-HB Combo/DPT-HB-Hib Booster Vaccine	This booster vaccine is administered to children aged 18 months to 24 months. It serves as a booster to the previous DPT-HB/DPT-HB-Hib vaccine series. The vaccine is injected into children who have already received 1-3 doses of the DPT-HB-Hib vaccine.
Measles Rubella Vaccine	The Measles Rubella vaccine provides immunity against measles. The first dose is administered when the baby is nine months old, and a booster dose is given at 5-7 years old. It is common for children to experience fever within one week after receiving this vaccine. The vaccine is injected into the thigh or upper left arm.
Measles Rubella Vaccine Booster	The Measles/MR vaccine booster is administered to children aged at least 18 months to 24 months. This booster vaccine is given after the child has received the previous Measles Rubella vaccine at nine months old.

Breastfeeding and complementary feeding history

Sampling Target : 0-23-month-old household member

This module aims to record information about the early initiation of breastfeeding (*Inisiasi Menyusu Dini/IMD*), breastfeeding practices, and complementary feeding (MP-ASI). Questions about the early initiation of breastfeeding are included only in the fifth wave. Additional questions related to complementary feeding are also incorporated. The questionnaire used is adapted from the 2013 Basic Health Research in Indonesia (4).

Hirschsprung Disease

Sampling target	:	0-59-months old household member
Wave	:	5
Question code	:	PH

Hirschsprung disease, or megacolon, is a genetic disorder characterised by the absence of certain nerve cells, which leads to bowel problems and functional intestinal obstruction in infants (19). The questionnaire in this module is developed based on the typical symptoms of Hirschsprung disease, including failure to have a bowel movement for more than a week, a swollen stomach, difficulties in swallowing or eating, and diarrhoea accompanied by fever (20).

Verbal Autopsy Module

Level	:	Individual
Sampling target	:	Respondents who were recently deceased.
Interviewee	:	Close relative or final caregiver of a recently deceased person
Wave	:	2,3,4,5
Question code	:	VA

A verbal autopsy is a method used to determine the cause of death based on information about the symptoms and circumstances leading to the person's demise. Health information and descriptions of events prior to death are obtained through interviews with individuals close to or serving as the final caregiver for the deceased. The questionnaire set used for verbal autopsies is translated from the 2016 WHO Verbal Autopsy Form 1.5.1.22. (21). Sleman HDSS utilises two versions of the WHO questionnaire: the 2014 version for verbal autopsies conducted in 2016 and the 2016 version for verbal autopsies from 2017 to 2019 (22). InterVA software version interva4_04_20170606 was used to analyse the data in 2015-2019 (23), while InterVA-5 software version 5.0 2018-02-22 was released in 2016 and used to analyse data in 2017-2021 (24).

Updates to Sleman HDSS Questionnaire

The Sleman HDSS employs a standardised set of questions for certain topics to ensure data consistency across waves. However, there are instances where modifications are necessary to enhance data quality or due to unavoidable changes in measurement instruments. Those modifications are as follows:

Cover

In waves 1, 4, and 5, the question 'KL10. GPS device used' underwent a change in answer choices. Initially, the options were '1. Android' and '2. Manual map,' which were subsequently replaced by '1. Android' and '2. Garmin' in waves 4 and 5.

Household members

There were additions to the answer choices in certain questions, such as 'art07', which in the third wave included '5. *Pisahan* (separated) ' and "6. *Rujuk* (get back together)', and 'art18', which included the additional answer choice '13. *Pelajar* (students)' regarding the respondent's main job.

Household Wealth Index

The household wealth index module was included in waves 1 and 4. It comprises sub-modules covering average household expenditure, household assets, ownership of goods, livestock ownership, land ownership, and house characteristics. However, in wave 4, average household expenditure was excluded from the household wealth index. Additionally, an extra answer choice was added to the variable 'hrt02' regarding fuel type used, namely '9. Bright gas 5 kg', '10. Bright gas 12 kg', and '11. Blue gas 5 kg.

Sleman HDSS Questionnaire Catalogue and Codebook

To facilitate the search for research variables and determine their inclusion in each wave's questionnaires, Sleman HDSS provides a comprehensive catalogue and codebook document. This resource contains the questions from each questionnaire module, spanning from the first wave to the most recent one. For further details, please refer to the following link:

https://hdss.fk.ugm.ac.id/codebook/.

Sleman HDSS Respondent Type

Sleman HDSS has three types of respondents:

1. Main Respondents

   Main respondents refer to household members over 18 years old who possess a comprehensive understanding of the household's situation and its members. Typically, the main respondents are the heads of households or their spouses.

2. Individual Panel Respondents

Individual panel respondents are representatives chosen from each household (1 person per household) aged over 25 years. Respondents are chosen using the Kish Grid Sampling method, 26–28, to ensure equal chances of selection for eligible household members. Since 2018, the e-HDSS application, developed by Sleman HDSS to facilitate data collection, includes a feature to select respondents based on the Kish Grid sampling method. Consequently, for new households, individual module respondents are selected automatically (see the e-HDSS manual). As for existing households, individual module respondents were previously selected in HDSS wave 3 (2017).

3. Primary Caregiver

The primary caregiver is an individual with detailed knowledge of the daily conditions of infants/children whose information is collected by Sleman HDSS. Primary caregivers may include mothers, grandmothers, or household assistants.

Table 13 shows the modules and the corresponding types of respondents who provide answers.

Table 13. Sleman HDSS modules list and respondents.

Module Code	Module Name	Main Respondent	Individual Panel	Main Caregiver
	Cover	Yes		
ART	Demographic data: Household member	Yes		
ARTB	Demographic data: new household member	Yes		
PART	Update demographic data: Household member	Yes		
KSM	Household food and beverage consumption	Yes		
PRP	Average Household Expenditure	Yes		
HRT	Asset ownership and household facilities	Yes		
KR	House or residential building characteristics	Yes		
KLH	Birth	Yes		

Module Code	Module Name	Main Respondent	Individual Panel	Main Caregiver
KMT	Death	Yes		
KRP	Reproductive health (Parity)	Yes		
PM	Infectious disease	Yes		
PTM	Non-communicable disease	Yes		
DIF	Disability	Yes		
HIV	Knowledge of HIV/AIDS	Yes		
MB	Unhealthy food	Yes		
INS	The behaviour of using mosquito repellents	Yes		
AKS	Health service	Yes		
KH-SF12	Quality of life		Yes	
KJ	Mental health	Yes		
PH	Hirschsprung disease	Yes		
KAI	Maternal, child health, and immunisation			Yes
ASM	Breastfeeding and complementary feeding			Yes
PTMV2	Non-communicable disease/NCD (Individual panel)		Yes	
AGH	Anthropometry and risk factors of NCD		Yes	
KJS	Individual mental health (Individual panel)		Yes	
VA	Verbal Autopsy (Cause of death)	Yes		

Respondents' ID Number

Each respondent participating in the Sleman HDSS study is assigned four types of research ID numbers:

1. IDRT Number

   The IDRT number represents the household ID and indicates the household location. It consists of seven digits, and **Figure 1.** provides a detailed explanation of each digit.

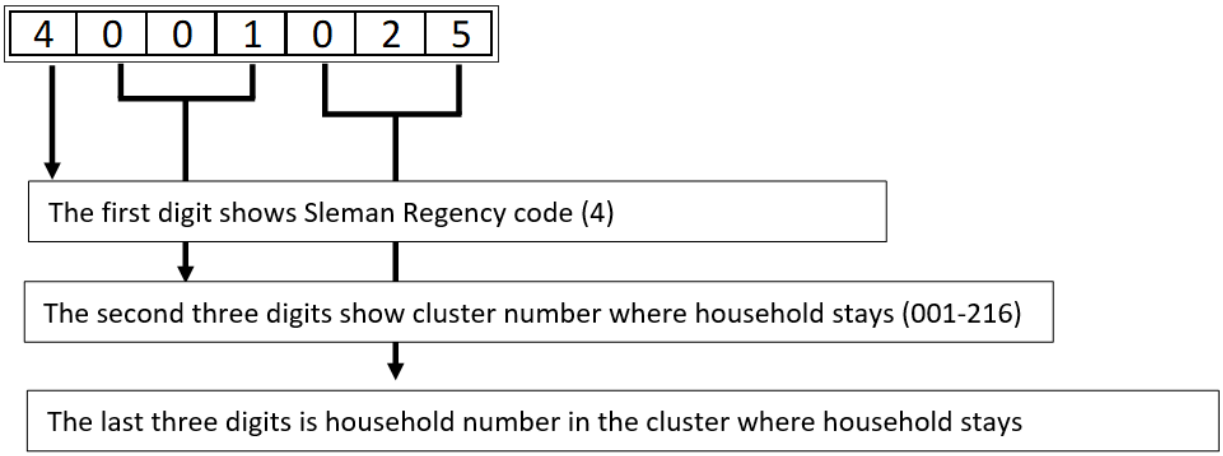


Figure 1. Sleman HDSS Respondents IDRT

2. NOART

   NOART is a two-digit serial number assigned to household members for identification within the household. In Sleman HDSS, the numbering follows the order of the respondent's name on the family register (see **Table 14**).

3. IDART

   IDART is a combination of IDRT and NOART (see **Table 14**).

4. Keyid

   Keyid is a permanent unique survey identity number assigned to each respondent. In wave 1, Keyid (referred to as art03b in the dataset) is the same as IDART (see the example in Table 14 and Table 15). However, IDART may change in subsequent waves if a respondent moves to another Sleman HDSS household (merged household) or forms a new household (split household). In such cases, the respondents receive new IDRT, NOART, and IDART, while the Keyid remains the same.

Table 14. List of Household A members and their ID numbers.

IDRT	NOART	IDART	Keyid	Name	Relationship with the household head
4019001	01	401900101	401900101	Gilang	Head of household
4019001	02	401900102	401900102	Rahayu	Spouse
4019001	03	401900103	401900103	Fadilah	Child
4019001	04	401900104	401900104	Atmaja	Grandchild

Table 15. List of Household B members and their ID numbers.

IDRT	NOART	IDART	Keyid	Name	Relationship with the household head
4019002	01	401900201	401900201	Alfianto	Head of household
4019002	02	401900202	401900202	Putri	Spouse
4019002	03	401900203	401900203	Anisa	Child

Changes in Respondents ID Number

The Sleman HDSS households can undergo changes that lead to changes in respondents' IDRT, NOART, and IDART.

Split Households

A split household occurs when one or more household members live separately from the registered household and establish a new household within the same Sleman HDSS working area cluster. Alternatively, if two different households share a house but have a separate kitchen, it is also considered a split household.

For example, in Table 14, Fadilah and Atmaja, members of family A, decide to form new families and move out. Consequently, they receive new IDRT, NOART, and IDART (Table 17), while the ID numbers for Gilang and Rahayu, who remain in the household, remain unchanged (Table 16).

Table 16. List of Household A members and their ID numbers after it split.

IDRT	NOART	IDART	Keyid	Name	Relationship with the household head
4019001	01	401900101	401900101	Gilang	Head of household
4019001	02	401900102	401900102	Rahayu	Spouse

Table 17. List of Household C (used to be part of Household A) members and their ID numbers.

IDRT	NOART	IDART	Keyid	Name	Relationship with the household head
4019003	01	401900301	401900101	Fadilah	Head of household
4019003	02	401900302	401900102	Atmaja	Child

Merged Household

Merged households occur when both households have previously participated in the HDSS study during the previous cycle. This event causes changes in the household and household member ID numbers. We need to identify the primary household to determine the new ID numbers. In order to determine the primary household (which determines the IDRT and the head of the household), it is necessary to consider several criteria:

1. Which household lived in the house that the two families now occupy? For example, suppose a member of household B moves into a house occupied by household A. In that case, household A is considered the primary household.
2. Who is the head of the household, or a household member designated as the head of the merged household? For example, households A and B declare their intention to merge, and a member or the head of household B assumes the role of the merged household's head; household B becomes the primary household.

To illustrate a 'merged households' event, let us consider the case of households A (**Table 14**) and B (**Table 15**) merging. In this scenario, Mr. Gilang serves as both the head of household A and the head of the merged household comprising households A and B. As a result, household B members' IDRT, NOART, and IDART follow the order of household A members (see **Table 18**).

Table 18. List of Household D (merging of Household A and B) members and their ID numbers.

IDRT	NOART	IDART	Keyid	Name	Relationship with the household head
4019001	01	401900101	401900101	Gilang	Head of household
4019001	02	401900102	401900102	Rahayu	Spouse
4019001	03	401900103	401900103	Fadilah	Child
4019001	04	401900104	401900104	Atmaja	Grandchild
4019001	05	401900105	401900201	Alfianto	Child
4019001	06	401900106	401900202	Putri	Daughter in law
4019001	07	401900107	401900203	Anisa	Grandchild

Another scenario involves a few household members moving to another Sleman HDSS household. For example, if only Atmaja from household A moves to household B, Atmaja receives new IDRT, NOART, and IDART based on Household B's identity numbers. In Table 16, Table 17, Table 18, Table 19, we can observe that respondents who receive new IDRT, NOART, and IDART keep their original Keyids.

Table 19. List of Household E members (Household B with an additional member from Household A) and their ID numbers.

IDRT	NOART	IDART	Keyid	Nama	Hubungan dengan KK
4019002	01	401900201	401900201	Gilang	Kepala Keluarga
4019002	02	401900202	401900202	Rahayu	Istri
4019002	03	401900203	401900203	Fadilah	Anak Kandung
4019002	04	401900204	401900104	Atmaja	Famili lain

Respondent Identity Numbers in Data Processing

Sleman HDSS data is stored in separate datasets by questionnaire module, data collection wave, and question level (the individual and household levels). IDRT is a unique identifier for merging the household-level dataset with other datasets. Meanwhile, Keyid (art03b) is used as unique identifiers for merging individual-level datasets within the household level.

Missing Data Codes

Missing data are common in research due to filtering or skipping patterns, and respondents refusing to answer certain questions. Properly defining and explaining missing data is crucial as it can impact calculations, analysis, and interpretation. Sleman HDSS applies the following codes to missing data:

-1 : “Do not know.”
-2 : “Refuse to answer.”
-3 : “Technical error.”
-4 : “Sphygmomanometer cuff does not fit.”
98 : “Unknown” missing.

Variable Naming

Sleman HDSS employs unique variable names, ensuring that ensuring that variables do not have the same combination. Variables that code for identical information and are asked in different waves retain the same names. Variable names in the Sleman HDSS data survey follow the format:

xxxYY_dd

“xxx” = module code e.g., “ptm” for non-communicable disease module

“YY” = question number e.g., “01”

“_dd” = indicates a dummy variable, which is used for variables such as the type of insurance owned (art20). Each insurance type is coded as a dummy variable (Yes/No answer), with art20_01 representing JKN PBI insurance and art20_02 representing JKN non-PBI.

Most Sleman HDSS variables, including occupation, education, and the infectious disease module, are time-varying, meaning they get updated each wave. Variable naming is differentiated based on the dataset type in which the variable is stored. In the wide dataset, the variable naming follows the pattern of xxxYY_dd_ (dummy variable) or xxxYY_ (non-dummy variable). In the long dataset, the pattern is xxxYY_dd_w or xxxYY_w, respectively. The element 'w' refers to the wave number. The examples below illustrate the naming of time-varying variables:

In a long dataset format: ART07_ → ends with “_”

Given variable means question number 07 in the Household Member module (ART). Table 20 shows how the variable is displayed in a long dataset.

Table 20. Long dataset variable display.

ID	Siklus	ART07_
40302	1	Ya
40302	2	Tdk
40503	1	Tdk
40503	2	Tdk

In a wide dataset format: ART07_1 → ends with “_” and wave number.

Given variable means question number 07 in the Household Member module (ART), in Wave 1. Table 21 shows how ART07_1 applies to the same respondent as Table 19 but in a wide dataset.

Table 21. Wide dataset variable display.

ID	ART07_1	ART07_2
40302	Ya	Tdk

ID	ART07_1	ART07_2
40503	Tdk	Tdk

Generated Variables

In addition to the original variables (variables corresponding to the questionnaire), the Sleman HDSS dataset also includes generated variables. These variables result from processing several initial/ raw variables following a specific manual.

Socioeconomic Status (SES)

Variable name in dataset: **sespct**

Sleman HDSS measures the Household Socio-Economic Status as Wealth Index. This index is derived from asset ownership and housing characteristics variables using the Principal Component Analysis (PCA) method (25). The resulting output classifies households into five distinct quintiles: low, lower-middle, middle, upper-middle, and high. Several variables used in determining the wealth index include the following:

1. kr04 : widest wall type
2. kr07 : most roof type
3. kr03 : widest flooring type
4. kr08 : main source of light
5. kr01 : house or residential building ownership status
6. hrt05 : does this household have a rice field?
7. hrt09 : does this household have a garden/field?
8. hrt13 : does this household have a lawn?
9. hrt01a : property ownership: television/tv
10. hrt01b : property ownership: cable tv (tv subscription services)
11. hrt01c : property ownership: refrigerator
12. hrt01d : property ownership: air conditioner/AC
13. hrt01e : property ownership: washing machine
14. hrt01f : property ownership: cell phone/ mobile phone
15. hrt01g : property ownership: water heater
16. hrt01h : property ownership: laptop/netbook/tablet
17. hrt01i : property ownership: desktop personal computer/PC
18. hrt01j : property ownership: bicycle

19. hrt01k : property ownership: motorcycle
20. hrt01l : property ownership: car
21. hrt01m: property ownership: truck
22. hrt01n : property ownership: bus
23. hrt01o : property ownership: 12 kgs or more, gas cylinder as the main type of fuel used in the kitchen

Quality of Life

Quality of life was measured using the SF-12v2® Health Survey (four-week recall). It covers eight domains: general health (GH), physical functioning (PF), role physical (RP), bodily pain (BP), vitality (VT), social functioning (SF), role emotional (RE), and mental health (MH). Each domain is scored based on the responses to specific questions. For example, physical functioning is scored based on questions about moderate activities and climbing stairs, while role emotional is scored based on questions about how emotional problems affect work or other activities. There are 12 items/questions in the SF-12 form, ten of them have a 5-option response. The other 2 questions use a 3-option response. The main results of SF-12v2® are psychometrically-based physical component summary (PCS) and mental component summary (MCS) scores.

The SF12 scores can be obtained by using PRO CoRE, the official software from OPTUM, Inc. (17) The analysis steps conducted by this software are as follows (16, 26):

1. Check the item values and exclude any that are out of range (for example, negative or greater than 5).
2. Reverse the values of four items (k1, k5, k6a, and k6b). By doing so, a higher item response value indicates better health for all SF-12® items.
3. Add up the values of all items to get the raw scale score of each domain, i.e., PF (Items k2a+k2b), RP (Items k3a+k3b), BP (item k5), GH (item k1), VT (item k6b), SF (item k7), RE (items k4a+k4b), and MH (items k6a+k6c).
4. Apply **Formula 2** to transform the raw scale score to a 0-100 scale. Save the results as new variables (pf, rp, bp, gh, vt, sf, re, mh).
5. Standardize the 0-100 score using a z-score transformation. In this process, each score is subtracted by the mean of the 0-100 score observed in the 1998 general U.S. population, then divided by the corresponding standard deviation (SD).
6. Calculate the norm-based score (NBS). NBS is the result of linear transformation of scores to get a mean of 50 and SD of 10 in the general U.S. population. This is achieved by multiplying each domain z-score by 10 and adding the resulting product to 50. Save

the results as new variables (pf_nbs, rp_nbs, bp_nbs, gh_nbs, vt_nbs, sf_nbs, re_nbs, mh_nbs).

7. Calculate NBS Physical Component Summary (PCS) and the Mental Component Summary (MCS). First, aggregate the standardized score of related domains (PCS=PF, RP, BP, GH and MCS=VT, SF, RE, MH) using weights (factor score coefficients) from the 1998 general U.S. population. Second, PCS and the MCS scores are then rescaled to have a mean of 50 and a standard deviation of 10 in the US general population.

$$\text{Transformed scale} = \frac{(\text{actual raw score} - \text{lowest possible raw score})}{\text{possible raw score range}} \times 100$$

(Formula 2)

Fruit and Vegetable Consumption

Variable name in dataset: **porsicat**

To determine the average consumption of fruit and vegetables, Sleman HDSS utilised probing questions to collect relevant information. The following steps outline how the portion of fruit and vegetables consumed by respondents is calculated:

1. The number of servings of fruit that are often consumed.
2. The name of the fruit that is most often consumed in a week.
3. Weight (grams) of fruit consumed in one sitting.
4. Frequency of consumption of the fruit per day.

A similar set of questions is used to gain information about vegetable consumption. Respondents' answers to the above questions are then processed to obtain the average portion of fruit and vegetables consumed. The following are the steps to calculate the portion of fruit and vegetables consumed by respondents:

1. Calculate the total consumption by multiplying the weight of fruit (or vegetable) by the frequency of consumption.
2. Convert the total consumption into servings based on the WHO standard, where 1 serving equals 80 grams. Divide the total consumption by 1 (one) WHO serving to obtain the number of servings.
3. Sum up the servings of consumed fruit and vegetables.

4. Classify fruit and vegetable intake as “inadequate” if consumption is < 5 servings daily (7).

Physical Activity Level

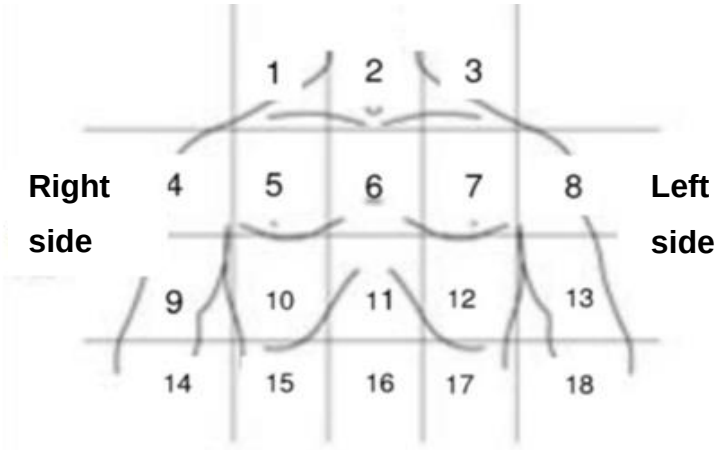
Variable name in dataset: **phyact**

Physical activity level is measured using the GPAQ V2 questionnaire (7). Collected data are stored as variables, i.e., f1, f2, f3a, f3b, f4, f5, f6a, f6b, f7, f8, f9a, f9b, f10, f11, f12a, f12b, f13, f14, f15a, and f15b. Following, the data are processed according to the GPAQ manual (27). During this process, additional variables are generated, namely: p1, p2, p3a, p3b, p4, p5, p6a, p6b, p7, p8, p9a, p9b, p10, p11, p12a, p12b, p13, p14, p15a, p15b, ptotal, ptotalday, phyact, and metsedent. The outcome of this entire process is the "phyact" variable, which categorises individuals into three levels of physical activity:

1. High: Individuals achieving a minimum of 3000 MET-minutes/week through walking, moderate or vigorous-intensity activities, or accumulating 1500 MET-minutes/week from vigorous-intensity activity within at least three days.
2. Moderate: Individuals achieving a minimum of 600 MET-minutes/week through walking, moderate or vigorous-intensity activities, or engaging in moderate activities for at least 30 minutes each on five or more days.
3. Low: Individuals who engage in no activity or some activity but do not meet the criteria for the previous two categories.

Angina

Questions related to angina symptoms are adapted from the Rose Questionnaire (28), which is also used in the Global Ageing and Adult Health (SAGE) study conducted by the World Health Organization (WHO)(2).



Angina is assessed based on the following criteria:

1. Experiencing chest pain on the left side (regions 2, 3, 6, 7, 8, 11, 12, 13, 16, 17, and 18) when walking uphill, on stairs, or while hurrying on level ground within the last 12 months.
2. Being diagnosed with angina and receiving angina medications or treatments.

The variables related to angina are recorded using the following questionnaire items: q4014, q4015a, q4015b, q4016, q4017, q4018, q4019, q4020, q4021. The angina status is stored as a variable named "angina_w3".

Stroke

In Wave 2, a stroke is identified when the main respondents report that a household member has been diagnosed with a stroke or has received medical and/or traditional treatments. In Wave 3, a stroke is identified when respondents report being diagnosed with a stroke and having taken medications or undergone treatments within the last 12 months.

The variables associated with stroke are based on the following questionnaire items: ptm05, ptm05a, and ptm06. The generated variable representing stroke status in Wave 2 is captured by the variable "stroke_w2."

The variables associated with stroke are based on the following questionnaire items: q4010, q4011a, and q4011b. The generated variable representing stroke status in Wave 3 is captured by the variable "stroke_w3".

Chronic Obstructive Pulmonary Disease (COPD)

In Wave 2, COPD is assessed when a household member was reported to be showing these symptoms in the past month: increased breathing difficulty during activities or increased breathing difficulty in line with age or experiencing asthma, wheezing, or COPD; all of which are accompanied by chronic cough/productive phlegm. COPD-related variables are based on questions ptm14, ptm14a, ptm19a, ptm19b, and ptm19c. The generated variable representing COPD status in Wave 2 is "**ppok_w2**".

In Wave 3, a respondent is categorised as having COPD when they report having been diagnosed with it and taking medication and/or treatment within the last 12 months. Variables of COPD are based on questions q4025, q4026a, q4026b, q4027, q4028, q4029, q4030,

q4031, q4032a, q4032b, and q4032c. The generated variable representing COPD status in Wave 3 is "**ppok_w3**".

Asthma was assessed only in Wave 2 using the following criteria: The main respondents stated that the household member experienced symptoms of shortness of breath accompanied by wheezing when exposed to conditions such as allergens (drugs, food, pollen, etc.), cigarette smoke, cold air, dust, fatigue, flu/infection, or stress; and symptoms worsened at night/early in the morning; or symptoms were relieved with medications; and symptoms started at the age of less than 40 years old. Asthma related variables are based on questions ptm14, ptm14a, ptm15a, ptm15b, ptm15c, ptm15d, ptm15e, ptm15f, ptm15g, ptm15h, ptm16a, ptm16b, ptm16c, and ptm16d. The generated variable representing asthma status is captured by the variable "**asma_w2**".

Hypertension

Hypertension in Waves 2 and 3 is determined based on the reported diagnosis of hypertension. The variables associated with hypertension are derived from questions ptm01, ptm01a, and ptm02 in Wave 3, or q4060, q4061a, and q4061b in Wave 3. The generated variables for hypertension in Waves 2 and 3 are **"hiper_w2" and " hiper_w3",** respectively.

Mental Health

The individual mental health module utilises the Self Reporting Questionnaire (SRQ) developed by the World Health Organization (WHO) (18). It consists of 20 questions addressed to selected respondents. The aim is to assess the respondent's mental health problems during the past month. This instrument validity in Indonesian setting has been tested and it has been used in The Basic Health Research (RISKESDAS) Indonesia. A study In Indonesia reported that with a cut-off point of 5/6, SRQ had a positive predictive value of 60% and a negative predictive value of 92% (29). If a subject answers "yes" to six or more questions out of the 20, they are considered to have emotional, mental disorder, or distress that may lead to mental problems. The generated variable representing mental health status is “mental”.

In addition to the total score, subscale scores can be calculated for specific symptoms, such as symptoms of depression. These symptoms are indicated by answering "yes" to all the following questions (variable name: depresi):

1. srq06: Does [NAME] hands shaking?
2. srq09: Does [NAME] feel unhappy?
3. srq10: Does [NAME] cry more than usual?

4. srq14: Is [NAME] unable to play a useful part in life?
5. srq15: Has [NAME] lost interest in things?
6. srq16: Does [NAME] feel that she/he is a worthless person?
7. srq17: Does [NAME] has the thought of ending his/her life been on [NAME] mind?

Symptoms of anxiety: answer “yes” to all the following questions (**variable name: cemas**):

1. srq03: Does [NAME] sleep badly?
2. srq04: Is [NAME] easily frightened?
3. srq05: Does [NAME] feel nervous, tense, or worried?

Somatic symptoms: answer “yes” to all the following questions (**variable name: somatik**):

1. srq01: Does [NAME] often have headaches?
2. srq02: Is [NAME] appetite, poor?
3. srq07: Is [NAME] digestion, poor?
4. srq019: Does [NAME] has uncomfortable feelings in his/her in stomach?

Cognitive symptoms: answer “yes” to all the following questions (**variable name: kognitif**):

1. srq08: Does [NAME] have trouble thinking clearly?
2. srq12: Does [NAME] find it difficult to make decisions?
3. srq13: Is [NAME] daily work suffering?

Symptoms of decreased energy: answer “yes” to all the following questions (**variable name: energi**):

1. srq08: Does [NAME] have trouble thinking clearly?
2. srq11: Does [NAME] find it difficult to enjoy your daily activities?
3. srq12: Does [NAME] find it difficult to make decisions?
4. srq13: Is [NAME] daily work suffering?
5. srq18: Does [NAME] feel tired all the time?
6. srq20: Is [NAME] easily tired?

Cause of Death

Sleman HDSS uses the Verbal Autopsy (VA) method to determine the cause of death in the population. Data related to the conditions before death are gathered from interviews using a VA questionnaire. The cause of death analysis is performed by the InterVA software. However,

beforehand, the data should be prepared in accordance with the InterVA guidelines. This preparation process is carried out using Stata software.

Once the data are ready, the analysis in the InterVA software can be conducted following these steps:

1. Input the prevalence level of Malaria and HIV in the study population. In our case, the prevalence of both diseases is “Low”.
2. Import the VA data.
3. Enter the desired file name for the analysis results and proceed to execute the program.
4. The analysis conducted using the InterVA software will yield potential causes of death along with other variables, including malaria prevalence (malprev), HIV prevalence (hivprev), pregnancy-related causes (pregstat), likelihood of pregnancy-related causes (preglik), cause of death 1 (cause1), likelihood of cause 1 (lik1), cause of death 2 (cause2), likelihood of cause 2 (lik2), cause of death 3 (cause3), likelihood of cause 3 (lik3), indeterminate likelihood (indent), most likely circumstantial category (comcat), and likelihood of comcat (comnum).
5. Import the InterVA analysis results into Stata and merge them with the initial VA data. The cause of death variables derived from InterVA include malprev, hivprev, pregstat, preglik, cause1, lik1, cause2, lik2, cause3, lik3, indent, comcat, and comnum. In the Sleman HDSS dataset, known causes of death for deceased individuals are recorded in the cause1, cause2, cause3, lik1, lik2, and lik3 variables.

Weighting

A large population study typically does not employ random sampling methods but instead adopts complex sampling techniques. Complex sample selection involves combining various existing methods and generally consists of multiple steps (30, 31). Sleman HDSS used a two-stage stratified design sampling method that results in a study sample that can closely represent the study population.

In the analysis, the use of a complex sampling method should be taken into account. One way to do it is by implementing weighting. Weighting is a method employed to re-establish the representativeness of a population based on samples by assigning additional "values" to underrepresented characteristics. The following are some additional objectives of weighting (32):

1. Counterbalancing the probability of unequal selection.
2. Adjusting for non-response units.

3. Aligning the sample distribution with key variables of interest, such as age, race, and gender, to match the known population distribution.

The Sleman dataset includes two types of weights variables: post-stratification weight and design weight. The choice of which weight to use depends on the data type and the specific analysis conducted by researchers. It is recommended that researchers consult with a statistician to ensure proper application of the weighting techniques.

Design Weight (Sampling Weight)

PSU: Cluster

Strata: Rural/Urban

In sample selection, the ideal scenario entails all sample units (in Sleman HDSS, households) having an equal probability of being selected. However, it is not always the case. In data analysis, design weights compensate for unequal selection probabilities (33), thus reducing bias caused by the sampling design. Design weight is based on the sampling procedure described in Formula 1 on page 10. The weight for each household in the selected census block is calculated as the inverse of the overall sampling fraction. The overall sampling fraction is determined by multiplying the sampling fractions at each sampling stage. The formula is denoted as follows:

$$f\_hij = f\_1 \times f\_2 = (b\_h\, M\_hi)/M\_h \times 25/(M\_hi\text{^}up) \quad (2)$$

The index f_hij is the overall sampling fraction of households $j = 1, 2, 3, \ldots, n\_$, while hi refers to strata of h and census block of i. Therefore, the weight for each household in the selected census block is written as the formula below:

$$W\_hij = 1/f\_hij = (M\_h\, M\_hi\text{^}up)/(25 b\_h\, M\_hi) \quad (3)$$

The Statistics Indonesia (BPS Indonesia) provided this weight. The sampling of the census block corresponds to the strata (Rural/Urban) and the proportion of rural and urban areas in Sleman Regency (variable name: dsweight).

Post-stratification Weight

In addition to bias caused by sampling design, discrepancies can arise due to sampling error and non-response error. When using data sourced from samples instead of the entire census population, sampling error can result in statistical estimates that do not accurately represent the target population. Similarly, the non-response error occurs when selected sample units

with specific characteristics refuse to participate, leading to the under-representation of those units and the over-representation of other units with different characteristics (34-36). Both types of errors can introduce bias in the analysis.

To address these issues, post-stratification weighting is applied to enhance population representativeness in the sample. The Sleman HDSS uses population data of Sleman Regency, based on age group, gender, and sub-district, to prepare this weighting. The steps of post-stratification weighting, based on age group and gender, require the following data:

1. Respondent group indicators based on age and gender (variable name: strata). There are 15 age categories with an interval of 5 years (ranging from 0-4 years to over 70 years old) and two gender categories, resulting in 30 categories.
2. Population of Sleman Regency in the year of data collection (variable name: fpc). This data is obtained from the demographic website of Sleman Regency (https://kependudukan.jogjaprov.go.id/).
4. Population of Sleman Regency for each age group and gender (variable name: postwgt). This data is also obtained from the demographic website of Sleman Regency (https://kependudukan.jogjaprov.go.id/).

Table 22. Number and percentage of HDSS Sleman respondents and Sleman Regency population by age group and gender in 2019

Age groups and sex (Variable name: strata)	Sleman HDSS Sample		Sleman Regency Population	
	Number (Variable name: obs_hdss5)	Percentage	Number (Variable name: postwgt)	Percentage
0-4 years old (L)	584	2.92	34,520	3.22
5-9 years old (L)	707	3.54	41,014	3.83
10-14 years old (L)	749	3.75	41,795	3.90
15-19 years old (L)	796	3.98	39,503	3.69
20-24 years old (L)	783	3.92	37,428	3.49
25-29 years old (L)	657	3.29	36,310	3.39
30-34 years old (L)	561	2.81	35,943	3.36
35-39 years old (L)	688	3.44	43,164	4.03
40-44 years old (L)	734	3.67	41,556	3.88
45-49 years old (L)	727	3.64	39,874	3.72

Age groups and sex (Variable name: strata)	Sleman HDSS Sample		Sleman Regency Population	
	Number (Variable name: obs_hdss5)	Percentage	Number (Variable name: postwgt)	Percentage
50-54 years old (L)	708	3.54	36,120	3.37
55-59 years old (L)	608	3.04	29,886	2.79
60-64 years old (L)	565	2.83	25,042	2.34
65-69 years old (L)	400	2.00	17,465	1.63
Over 70 years old (L)	659	3.30	30,942	2.89
0-4 years old (P)	592	2.96	32,790	3.06
5-9 years old (P)	665	3.33	39,016	3.64
10-14 years old (P)	728	3.64	39,434	3.68
15-19 years old (P)	757	3.79	37,775	3.53
20-24 years old (P)	767	3.84	36,077	3.37
25-29 years old (P)	588	2.94	36,488	3.41
30-34 years old (P)	555	2.78	37,199	3.47
35-39 years old (P)	714	3.57	43,527	4.06
40-44 years old (P)	783	3.92	42,344	3.95
45-49 years old (P)	739	3.70	40,845	3.81
50-54 years old (P)	777	3.89	38,346	3.58
55-59 years old (P)	698	3.49	33,068	3.09
60-64 years old (P)	607	3.04	26,831	2.51
65-69 years old (P)	367	1.84	18,519	1.73
Over 70 years old (P)	735	3.68	38,092	3.56
Total	19,998	100.00	1,070,913	100.0

* Display of number and percentage of respondents for weighting calculation.

In Stata software, post-stratification weight is applied with the following command:

```
svyset: svyset, poststrata(strata) postweight(postwgt) fpc(fpc)
```

Example: Stroke prevalence in the Sleman HDSS' individual panel (wave 3) with and without weighting.

There are 4,502 respondents in the individual panel who answered questions about stroke; among them 55 respondents had stroke. Therefore, the stroke prevalence without weighting equals 1.22%. After applying post-stratification weight to the analysis, stroke prevalence decreased to 0.88% (see **Table 23**).

Table 23. Tabulation output of stroke_w3

Stroke	Number of respondents	Prevalence	Number of respondents with weighting	Prevalence with weighting
No	4,447	98.78	4,437	99.12
Yes	55	1.22	55	0.8785

For analysis by sub-district, the post-stratification weight is constructed based on the following data:

1. Respondent group indicators based on age group (**range_agex**), gender (**art04**), and sub-district (**kl01**) (variable name= **stratakec**). There are 15 age categories with an interval of 5 (five) years (from 0-–4 years to over 70 years old), 2 (two) sex categories, and 17 sub-districts. Thus, all together, there are 510 categories.
2. Sleman Regency's population number in the year of data collection (variable name= **fpc**).
3. Population of Sleman Regency for every age group and gender (variable name= **postwgtkec**).

Sleman HDSS Data Release

Every year, Sleman HDSS releases a dataset that includes compiled data from the first wave up to the latest wave. Each dataset is assigned a release number, indicating its version. The data undergo continuous cleaning and synchronisation between waves. Therefore, it is recommended that data users always utilise the latest version and include the corresponding release number in their publications.

The release number of the Sleman HDSS dataset consists of three digits, with the following explanations:

1. The first digit indicates a significant change in the dataset, such as the addition of the latest data or data corrections that result in changes in the sample size.
2. The second digit represents a data change that does not affect the sample size.
3. The third digit indicates a minor change, such as adding a new variable.

Table 24. Sleman HDSS dataset version list

Version Number	Date of Release
1-0-0	24 November 2017
2-0-0	29 December 2017
3-0-0	02 July 2018
3-1-0	11 February 2019
5-0-0	13 June 2017
5-1-0	15 Mei 2017
6-0-0	19 December 2017
7-0-0	12 December 2018
8-0-0	26 February 2019
9-0-0	30 December 2020
9-1-0	02 January 2020
9-1-1	09 Market 2020

Sleman HDSS Data Utilisation

Lecturers, researchers, and students can utilise Sleman HDSS data for their final assignments, such as thesis, dissertation, and manuscript. They can request secondary data, aggregated data, or nested research based on the Sleman HDSS dataset.

Aggregate Data

Aggregate data is the result of processing raw Sleman HDSS data. It includes aggregated information such as disease prevalence, the number of respondents with specific conditions, or birth rates.

Secondary Data

Users can utilise Sleman HDSS data for secondary data analysis. Prior to applying for data requests, users need to comprehend Sleman HDSS's data collection methods, variable availability, and data structure. Additionally, it is crucial to ensure the novelty of the research plan. Therefore, they should also review past and ongoing research based on Sleman HDSS

data. Once their application is approved, the Sleman HDSS data management team will prepare a dataset containing the requested variables. Below are the general conditions of use:

1. Only the academic community of FKKMK UGM can access the latest Sleman HDSS data.
2. Non-FKKMK UGM data users can access Sleman HDSS data two years after the data collection.
3. Non-FKKMK UGM data users can access the latest data if they collaborate with FKKMK lecturers or researchers.
4. Data users must commit to publishing their work in reputable journals.

Nested Research

Nested research collects primary data, which is not yet available in the Sleman HDSS, from some or all Sleman HDSS respondents. A nested study can use both secondary data from Sleman HDSS and solely primary data. If necessary, nested research can also recruit additional respondents from the Sleman HDSS area. Conducting nested research in Sleman HDSS offers several advantages as follows:

5. Simplified research sample selection.
6. Improved time, cost, and resource efficiency, as Sleman HDSS can aid in data collection.
7. Enhanced data analysis by enabling researchers to integrate Sleman HDSS data into their research.

Researchers wishing to use HDSS data or conduct nested research must comply with current regulations. They need to complete a data request form and submit all required documents to Sleman HDSS. The Data Utilisation Division of Sleman HDSS will review the request, while the Data Management Team will process requests for aggregate and secondary data that meet the assessment criteria. Nested research requests will be forwarded to the Data Collection Division. This entire process usually takes between 2 to 4 weeks. Further details regarding HDSS Sleman collaboration schemes are available on the HDSS Sleman website at https://hdss.fk.ugm.ac.id/skema-kerjasama/.

Data Collection Instruments

Data collection for HDSS is conducted through direct interviews using electronic questionnaires (CAPI—Computer-assisted personal interview). Enumerators conduct face-to-face interviews using tablet computers equipped with interview data collection applications. In the first wave, data was collected using the Commcare application, which is based on Open Data Kit. In wave 2, paper-based questionnaires were used, and data were entered into

Epidata software. For wave 3 and subsequent cycles, the Sleman HDSS programming team developed the e-HDSS software, which became the primary tool for data collection.

e-HDSS

The e-HDSS application was developed by the Sleman HDSS Programming Team in 2016. Since the third wave (2017), e-HDSS has been utilised for data collection in the main waves. This application caters to the specific needs of Sleman HDSS data collection and management.

The process of data collection using e-HDSS consists of four stages (**Figure 2**). The first stage involves research planning, where researchers determine the modules, questionnaires, and instruments to be used in the survey. Research assistants then verify the language and instrument compatibility with the characteristics of the Sleman HDSS research area. If necessary, the instrument undergoes reliability and validity testing. Once everything is finalised, the questionnaire is sent to the programmer. The data manager prepares the latest version of the socio-demographic data of the respondents, which serves as the basic information for e-HDSS.

The second stage is e-HDSS development. The programmer developed e-HDSS based on the latest questionnaire and respondents' socio-demographic data. Once the development phase is completed, a beta version of e-HDSS will be released and undergo a testing phase. During this phase, a research assistant will act as a user and assess the instruments' functionality according to the provided guidelines. Any identified errors will be addressed and fixed by the programmer. This iterative process continues until the final version of e-HDSS is achieved.

Within the e-HDSS system, there are three types of accounts: administrator, supervisor, and enumerator. The supervisor can create enumerator accounts, allowing enumerators to access the e-HDSS database. Subsequently, the supervisor assigns lists of respondents to each enumerator. The enumerators download the relevant respondent data from the server to their tablets and commence the data collection process.

e-HDSS is a hybrid application that can be utilised on various platforms, including the web and Android. It is installed on tablet computers for data collection purposes, while the web version is accessed on a PC or laptop for data management activities. Once the enumerators have collected the interview data, it is uploaded to the server. The data manager then downloads the data onto their computer and initiates the data cleaning process.

Notably, the e-HDSS application was registered as an Intellectual Property Rights (IPR) at the Ministry of Law and Human Rights of the Republic of Indonesia on February 28, 2020, with the IPR number 000191054, acknowledging it as a computer program invention.

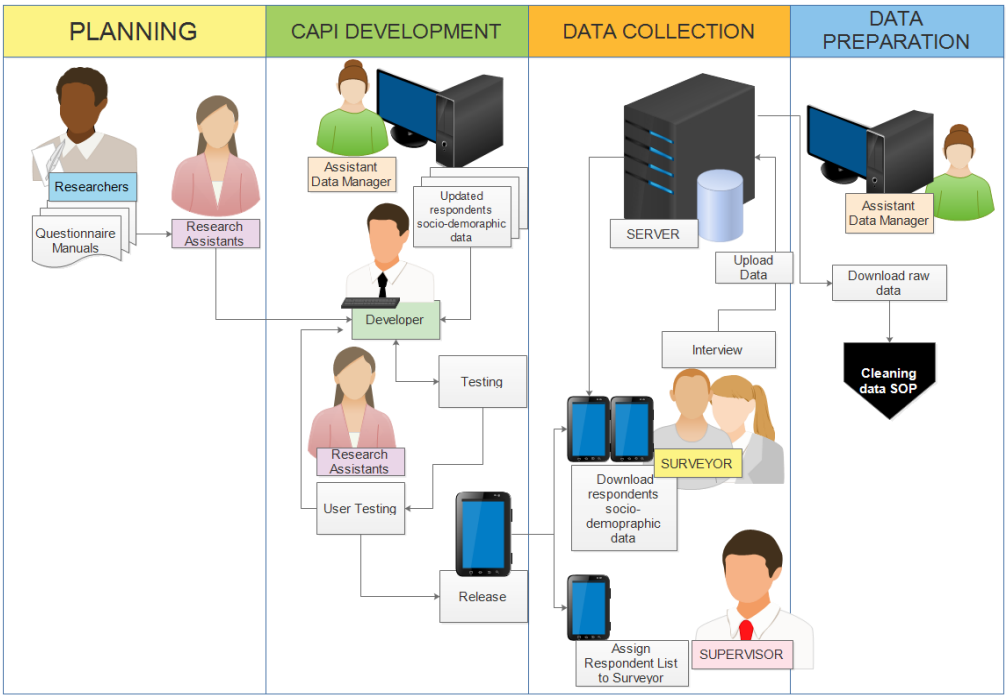


Figure 2. Data collection and management using e-HDSS

Table 25. e-HDSS features

Platform	Multiplatform
Form	Modular template Used standard meta data Support various question types Validation rules Skipping pattern Offline data input Recall data Household and individual level questionnaire Automatic calculation Filter questions
Special feature of longitudinal study	List of household members Split household record Merge household record Add new household Kish sampling Generate new individual respondent id automatically Generate new household respondent id automatically
Server	Data storage

	Editing and deleting data on the server
Daily report	
Display image/media	
User roles for access (multi-user)	
Import/Export	Upload data
	Upload forms
	Export the specified part
	Data labels were fully importable
	export as csv.
Security	Server encryption
	De-identification process

Reference

1. Dewi FST, Choiriyyah I, Indriyani C, Wahab A, Lazuardi L, Nugroho A, et al. Designing and collecting data for a longitudinal study: the Sleman Health and Demographic Surveillance System (HDSS). Scandinavian Journal of Public Health. 2018;46(7):704-10.
2. World Health Organization. WHO SAGE survey manual: the WHO study on global AGEing and adult health (SAGE). Geneva: World Health Organization. 2006.
3. Wilopo SA, Hakimi M, Dasuki D, Wahab A, Ng N, Emilia O. Purworejo Demographics Surveillance System Indonesia 1997 [Available from: http://www.indepth-network.org/dss_site_profiles/PurworejoProfile.pdf.
4. National Institute of Health Research and Development (NIHRD) MoHI. Kuesioner Riset Kesehatan Dasar 2013 (Basic Health Research 2013 Questionnaire) 2013.
5. Badan Pusat Statistik Indonesia. Kuesioner Survei Sosial Ekonomi Nasional 2014 (National Socio-Economic Survey Questionnaire). Jakarta (IDN): Badan Pusat Statistik Indonesia; 2014.
6. National Institute of Health Research and Development (NIHRD) MoHI. Riset Kesehatan Dasar 2018 (Basic Health Research 2018) 2018.
7. World Health Organization. WHO STEPS surveillance manual: the WHO STEPwise approach to chronic disease risk factor surveillance. World Health Organization; 2020.
8. Kementerian Kesehatan Republik Indonesia (Indonesia Ministry of Health). Buku Foto Makanan (PORSIMETRI) (Food Book Photos): Kementerian Kesehatan Republik Indonesia (Indonesia Ministry of Health); 2014. Available from: http://sikap.pemkomedan.go.id/sipolan/porsimetri.pdf.
9. Undang-Undang Tentang Penyandang Disabilitas (Laws on Persons with Disabilities), 8 (2016).
10. Regional Health Office of the Special Region of Yogyakarta. Mengenal Penyandang Disabilitas Lebih Dekat (bagian 1)-Let's Get to Know People with Disabilities More Closely (Part 1) 2018 [Available from: https://dinkes.jogjaprov.go.id/berita/detail/disabilitas-ragam-jenis-yuk-mengenal-penyandang-disabilitas-lebih-dekat-bagian-1.
11. Centers for Disease Control and Prevention. Disability and Health Overview 2020 [Available from: https://www.cdc.gov/ncbddd/disabilityandhealth/disability.html#:~:text=According%20to%20the%20World%20Health,%2C%20walking%2C%20or%20problem%20solving.
12. Centers for Disease Control and Prevention. Facts About Intellectual Disability 2022 [Available from: https://www.cdc.gov/ncbddd/developmentaldisabilities/facts-about-intellectual-disability.html
13. World Health Organization. Deafness and hearing loss 2023 [Available from: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
14. National Library of Medicine National Center for Biotechnology Information. Blindness 2023 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK448182/.
15. Central Statistical AE, Macro ORC. Ethiopia Demographic and Health Survey 2005. Addis Ababa, Ethiopia: Central Statistical Agency/Ethiopia and ORC Macro; 2006.
16. Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. How to score version 2 of the SF-12 health survey (with a supplement documenting version 1): QualityMetric Incorporated Lincoln, RI; 2005.
17. Optum. PRO CoRE 1.3 Smart Measurement® System: QualityMetric Incorporated Lincoln, RI; 2018.
18. Beusenberg M, Orley JH. A User's guide to the self reporting questionnaire (SRQ). Geneva: World Health Organization. Division of Mental, Health; 1994.
19. Langer JC. Hirschsprung Disease. In: Mattei P, editor. Fundamentals of Pediatric Surgery. New York, NY: Springer New York; 2011. p. 475-84.
20. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Hirschsprung Disease 2015 [Available from: https://www.niddk.nih.gov/health-information/digestive-diseases/hirschsprung-disease.

21. Nichols EK, Byass P, Chandramohan D, Clark SJ, Flaxman AD, Jakob R, et al. The WHO 2016 verbal autopsy instrument: An international standard suitable for automated analysis by InterVA, InSilicoVA, and Tariff 2.0. PLOS Medicine. 2018;15(1):e1002486.
22. World Health Organization. Verbal autopsy standards: The 2016 WHO verbal autopsy instrument 2016.
23. Byass P, Chandramohan D, Clark SJ, D'Ambruoso L, Fottrell E, Graham WJ, et al. Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. Global Health Action. 2012;5(1):19281.
24. Umeå University. InterVA-5 User Guide Version 5.0 2018-02-22. 2018. 2018 [Available from: http://www.byass.uk/interva/products.htm.
25. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Health Policy and Planning. 2006;21(6):459-68.
26. Ware JE, Keller SD, Kosinski M. SF-12: How to score the SF-12 physical and mental health summary scales: Health Institute, New England Medical Center; 1995.
27. World Health Organization. Global Physical Activity Questionnaire (GPAQ) Analysis Guide. Geneva: World Health Organization; 2012. Available from: https://www.who.int/docs/default-source/ncds/ncd-surveillance/gpaq-analysis-guide.pdf.
28. Rose GA. The diagnosis of ischaemic heart pain and intermittent claudication in field surveys*. Bulletin of the World Health Organization. 1962;27(6):645-58.
29. Ganihartono I. Psychiatric morbidity among patients attending the Bangetayu community health centre in Indonesia. Buletin Penelitian Kesehatan. 1996;24(4).
30. Johnson DR. Using weights in the analysis of survey data 2008. Available from: https://pages.nyu.edu/jackson/design.of.social.research/Readings/Johnson%20-%20Introduction%20to%20survey%20weights%20(PRI%20version).pdf.
31. Kaminska O. Guide to Using Weights and Sample Design Indicators with ESS Data Contents 2020. Available from: https://www.europeansocialsurvey.org/docs/methodology/ESS_weighting_data_1_1.pdf.
32. UN- Statistics Division. Designing Household Survey Samples: Practical Guidelines. Studies in Methods Series F. 2005;98.
33. Lavallée P, Beaumont J-F. Why We Should Put Some Weight on Weights. Survey methods: Insights from the field (SMIF). 2015.
34. Arnab R. Chapter 15 - Nonsampling Errors. In: Arnab R, editor. Survey Sampling Theory and Applications: Academic Press; 2017. p. 469-504.
35. Tyrer S, Heyman B. Sampling in epidemiological research: issues, hazards and pitfalls. BJPsych Bulletin. 2016;40(2):57-60.
36. Kempf-Leonard K. Encyclopedia of social measurement. 2004.

UNIVERSITAS GADJAH MADA
FAKULTAS KEDOKTERAN,
KESEHATAN MASYARAKAT, DAN KEPERAWATAN