MALAWI

National Statistical Office (NSO)

The Fifth Integrated Household Panel Survey

(IHPS)

2024

Basic Information Document

VERSION 1

(JULY 2026)

ACRONYMS

CPI	Consumer Price Index
EA	Enumeration Area
GER	Gross Enrolment Rate
GPS	Geographical Positioning System
HESSEA	Household Expenditure and Small-Scale Economic Activities
IFPRI	International Food Policy Research Institute
IHS	Integrated Household Survey
ISA	Integrated Surveys on Agriculture
LSMS	Living Standards Measurement Surveys
LSMS-ISA	Living Standards Measurement Surveys-Integrated Surveys on Agriculture
MCA	Millennium Challenge Account
MGDS	Malawi Growth and Development Strategy
MIP-1	Malawi 2063 First 10-year Implementation Plan
MoFEPD	Ministry of Finance, Economic Planning and Decentralization
NER	Net Enrolment Rate
NSO	National Statistical Office
PSU	Primary Sampling Unit
SDGs	Sustainable Development Goals
UNHCR	United Nations High Commissioner for Refugees
WB	World Bank

TABLE OF CONTENTS

Chapter 1: Introduction

The purpose of this document is to give comprehensive details on the 2024 Integrated Household Panel Survey (IHPS5), which was conducted by the National Statistical Office (NSO).

The survey, carried out roughly every three years, aligns with national policy goals and international standards, particularly in monitoring the SDGs, MIP-1 and the aspirations of Malawi 2063. Ongoing IHS implementation guarantees access to current socio-economic data, supporting evidence-based development planning and informed policy-making in Malawi, in line with the pillars of agriculture transformation, industrialisation, and urbanisation as envisioned under Malawi 2063.

The panel component was incorporated into the core IHS program to examine trends in poverty, socio-economic, and agricultural characteristics over time through a longitudinal survey. At the baseline, the IHPS sample was selected to be representative at the national, regional, urban/rural levels, and for each of the following six strata: (i) Northern Region — Rural, (ii) Northern Region — Urban, (iii) Central Region — Rural, (iv) Central Region — Urban, (v) Southern Region — Rural, and (vi) Southern Region — Urban. The IHPS 2013 endeavored to monitor all baseline households and individuals who relocated from the original residences between 2010 and 2013, provided they were not servants or guests during the IHS3 survey; were estimated to be at least 12 years old; and were confirmed to reside in mainland Malawi, with the exception of those on Likoma Island.[1] and in institutions, including prisons, police compounds, and army barracks. Once a split-off individual was located, the new household that he/she formed or joined since 2010 was also incorporated into the IHPS sample. In consideration of the tracking regulations, the final IHPS 2013 sample thus comprised a total of 4,000 households that could be traced to 3,104 baseline households.

Due to the increasing number of households from the 2013 database to be tracked and budget constraints, 102 of the 204 baseline EAs were selected for the second round of the IHPS. The IHPS 2016 attempted to track households and individuals who had moved away from the dwellings where they were enumerated in 2013. In line with the tracking rules, the final 2016 database included a total of 2,508 households that could be traced back to 1,908 households from 102 EAs in the IHPS1 (2013) database.

The IHPS4 (2019) sought to track households and individuals who relocated from the residences they occupied during the IHPS 2016.

The analysis was restricted to the national, urban, and rural areas, given that the number of Enumeration Areas (EAs) was reduced by half during IHPS2 (2016). Although the results of the IHPS 2019 cannot be tabulated by region, the stratification of the IHS3 Panel Survey by region, as well as urban and rural strata, was preserved through a proportional allocation of the sample across regions, based on the distribution derived from the 2008 Malawi Census. In line with the tracking rules, the final 2019 database included a total of 3,178 households that could be traced back to 2,370 of the 2016 households, indicating a household-level attrition rate of 5.5 percent.

[1] The exclusion of Likoma Island is rooted in the traditional exclusion of the district for IHS purposes, largely due to logistical considerations.

The IHPS sample underwent a structural change in 2024: a partial refresh, whereby 50 percent of the 102 EAs were retired (and all associated households emanating from those EAs) and replaced with an equivalent set of freshly selected EAs and households within them. There were several motivating factors behind this partial refresh. The first was the desire to alleviate the respondent burden of some households which had been participating in the IHPS across a decade. The second was the desire to boost cross-sectional representativeness of the IHPS with freshly identified and selected households. As longitudinal samples age, their cross-sectional representativeness tends to fade since the underlying population undergoes socio-demographic changes over time. While this deterioration in cross-sectional representativeness is partially offset through implementation practices (e.g., following of split-off households) and weight adjustments (e.g., calibration), incorporating a freshly selected set of households is the most effective strategy to improve cross-sectional representativeness. The third motivation was to carefully curb and regulate the expansion of the panel sample size induced by the individual following rules.

The IHPS5 (2024) which was conducted between July 2024 and February 2024 (with residual tracking spilling over into May) aims to produce representative and reliable estimates at the national level as well as for urban and rural areas. The sample is also designed to provide representative estimates across the three regions (Northern, Central, and Southern), albeit with a lower degree of precision. Furthermore, it aims to provide insights into the longitudinal dynamics occurring within the country since 2010 when the first IHPS was fielded.

Chapter 2: The Survey Instruments

Four types of questionnaires were administered during the IHPS5: The first was the **Household Questionnaire**, a multi-topic survey instrument that is nearly identical in content and structure to the IHPS4. It collected detailed information on household economic activities, demographics, welfare, and other sectoral characteristics. The questionnaire covers a broad range of topics related to poverty dynamics, including consumption, cash and non-cash income, savings, assets, food security, health, education, vulnerability,social protection, savings and domestic tourism. While comprehensive, the IHPS5 household questionnaire intentionally excludes in-depth modules covered extensively in other surveys within the National Statistical Office statistical program, such as maternal and child health topics addressed in the Malawi Demographic and Health Survey.

The **Agriculture Questionnaire** was administered to all IHPS5 households identified as engaging in agricultural or livestock activities. It facilitates detailed agricultural productivity analysis through careful measurement of land ownership and cultivation, labor and non-labor input use and expenditures, and crop and livestock production. Although a primary objective of the agricultural data collection was to generate smallholder production estimates for major crops, the data also allow for disaggregation by gender and by major geographic regions.

The **Fisheries Questionnaire** was designed to collect detailed information from households engaged in fish farming activities.

The **Community Questionnaire** was administered to a group of community-level representatives. For the purposes of the survey, a community was defined as the village or urban locality surrounding the selected enumeration area that residents commonly recognize as their community. The IHPS5 community questionnaire was administered in each sampled EA to a group of knowledgeable individuals, such as the village headperson, school head teacher, agricultural extension worker, religious leaders, local merchants, health workers, and long-term residents. The instrument collected information on a wide range of community characteristics, including religious and ethnic composition, physical infrastructure, access to public services, economic activities, communal resource management, governance and local organization, development projects, and retail prices of essential goods and services.

Table 2.1: Contents of the IHPS5 Household Questionnaire

Module	Description
Module A	This module contains household identifiers, the sample weights, information on household location, date of interview, supervisor and enumerator codes. Additionally, this module contains filters for subsequent modules.

Continued on next page

Table 2.1 – *continued from previous page*

Module	Description
Module B: Household Roster	This module contains the roster of individuals living in the household, their sex, relationship to the household head, age, duration away from the household in past 12 months, number of days meals were taken in the household, where born, how long in this community, and information on the location and level of education of parents of every member, including IDs if in the household and for members aged 10 and above, access to phone and internet. For members over 12, information on religious affiliation, marital status and location of spouses is collected and identifies the ID of the spouse/s of a household member.
Module C: Education	The education module is asked of all individuals over 5 years in age and collects information on self-reported reading and writing ability, school attendance, highest class attended and highest qualification achieved, year and age of beginning school. If the individual is presently attending school, information on the type of school, distance, and costs are collected. For members aged 15 and above, information on techinical/artisan/hands on skills are collected.
Module D: Health	The health module is administered to all individuals and collects information on: illness or injury in the past 2 weeks, diagnosis source, and action taken, and disruption to normal activity; health spending over the past 4 weeks; hospitalization or stay in a traditional healer's in the last 12 months. For individuals over 5 years in age: information on chronic difficulties and disruption to normal activities; chronic illness and diagnosis source. For women aged 12 to 49 years of age, information on births in the last 24 months, prenatal health clinic visits, and where the baby was born and who assisted at birth for last-born child is collected. Questions on disability were also administered.
Module E: Time Use and Labour	The module is administered to all individuals 5 years or older. This module collects information on hours spent yesterday collecting water and wood; hours spent in the last 7 days on agriculture and non-agriculture activities; type of primary and secondary work, employers and wages over the last 12 months; participation in unpaid apprenticeships, casual (ganyu) labour, and other unpaid labour over the last 12 months. Households involved in agriculture: 5 crops were captured in accordance with importance (defined as value addition in terms of non-market (consumption) or market (commercial sales) terms).
Module F: Housing	This module on housing is administered to the household head. It collects information on the characteristics of the dwelling, household fuel use, availability of electricity, telephone and water, toilet and rubbish facilities, and mosquito net use. In an attempt to improve data collected on land rights and ownership, this module contains detailed questions on who owns the property and who has the right to sell or bequeath the property containing their dwelling. New detailed questions were added on use of toilet facilities.

Continued on next page

Table 2.1 – *continued from previous page*

Module	Description
Module F1: Land Roster	This is a new module and it collects information on all agricultural and non-agricultural land for which any household member currently uses, owns or holds use rights for, either alone or with someone else. The gardens in this module feed forward into the different agricultural modules depending on the time they were cultivated.
Module G: Consumption of Food Over Past One Week	This module collects information on all food consumed by the household in the past 7 days: in total and then classified as purchased (with price), own-production, or gift and other sources. Additionally, this module collects information on number of days aggregated food categories were consumed by the household and number of days and meals taken in the household by children and adults.
Module H: Food Security	This module collects information on number of meals taken by adults and children (6-59 months and 5-17 years) in the household and restricted food intake in the past 7 days.
Module I: Non-Food Expenditures	This module collects expenditures on non-food items over the past week and the past 1 month.
Module J: Non-Food Expenditures (3 months)	This module collects expenditures on non-food items over the past 3 months.
Module K: Non-Food Expenditures (12 months)	This module collects expenditures on non-food items over the past 12 months.
Module L: Durable Goods	This module collects information on ownership, quantity owned, age of items, current preserved market value, purchases of items in the last 12 months, and cost of items in the last 12 months for durable goods.
Module M: Farm Implements, Machinery and Structures	This module collects information on household ownership, quantity owned, age of items, perceived market value, item purchases in last 12 months, quantity purchased in last 12 months, asset value, use, and items rental and rental cost, for farm implements and structures. Additionally, for farm structures, information is collected on construction and cost of construction over the past 12 months.

Continued on next page

Table 2.1 – *continued from previous page*

Module	Description
Module N: Household Enterprises	This module collects information on non-agricultural family enterprises or trading businesses, specifically who manages/owns the enterprise, employees, enterprise operation periods, start-up capital and source, customers, business trends, sales revenue, expenditures, and profits.
Module O: Children Living Elsewhere	This module collects information on the age, sex, education, length away from household, current locations, activity status and occupation of children living outside the household. Additional information is collected on remittances to the household from children living outside the household.
Module P: Other Income	This module collects information on household income from interest, pensions, rentals, or other income over the past 12 months.
Module Q: Gifts Given Out	This module collects information on cash, food, or other in-kind items given by the household in the past 12 months.
Module R: Social Safety Nets	This module collects information on receipts and value of social safety nets including cash, food, or other aid from programs. Additionally this module collects information on household member recipients of the aid, decision making for aid received, and number of months aid was received.
Module S: Credit	This module collects information on household credit, specifically where the credit was acquired, who is responsible for the loan, reason credit was obtained, how much was borrowed, timing of loan, and expected pay-off. Additionally this module collects information on attempted credit and reasons for being turned down.
Module T: Subjective Assessment of Well-being	This module collects information on the respondent's assessment of his/her family's situation regarding food consumption, housing, clothing, health care, financial level, and income level. The intended respondent for this module is the head of household. Additionally this module asks the head of household about the number of changes of clothes owned, and bedding type.
Module U: Shocks & Coping Strategies	This module collects information on shocks on the household in the past 3 years such as crop disease, theft of livestock, death of family members and how many times these shocks have happened in the past 3 years. Respondents are asked to rank the 3 most severe shocks experienced within a 12 month period (filtered based on date from those listed as having happened in the past years) and report on the impact of the shock on income, assets, food production, food stocks and food purchases, as well as what was done by the household in response to the shock.
Module V: Child Anthropometry	This module collects weight and height/length measurements as well as observed oedema for children of age 0-60 months. Additionally, this module collects information on child participation in nutrition programs and under five clinics. *Note: The age group was adjusted downwards from 0 for this round.*

Continued on next page

Table 2.1 – *continued from previous page*

Module	Description
Module Y: Domestic Tourism	This module explores the phenomenon of individuals traveling outside their customary environment for various purposes, where the primary motive of the trip does not involve engaging in remunerated activities within the visited location. Such travel occurrences typically involve journeys of at least 30 kilometers from the usual place of residence and are broadly classified under the term 'tourism.' Tourism includes both overnight excursions, where travelers remain away from their residence for a minimum of one night, and day trips, where individuals travel and return within the same calendar day without staying overnight.
Module ZB: Savings	This module offers information regarding whether individuals aged fifteen years and above possess accounts with commercial banks, credit unions, microfinance institutions, village savings organizations, and other financial entities. I also assesses whether they use formal or informal methods of saving.
Module W: Deaths in the Household	This module records information on family members who have died in the past two years and collects information on the type of work previously performed, age at death, and previous illness of the deceased household member. It also collects information on the diagnosis source of cause of death and assets lost due to the death.
Module X: Filter Questions for Agriculture & Fishery	This module contains filter questions on the presence of agricultural, livestock and/or fisheries in the household.

Table 2.2: Contents of the IHPS5 Agricultural Questionnaire

Module	Description
Module B_2: Garden Details (Rainy Season)	This module was adjusted in 2019 and collects information on details of how the garden was acquired, proceeds from renting out the garden, obligations arising from renting in the garden and decision makers as regards the proceeds from the garden. It also collects information on concerns of ownership of a garden. Some information that was collected using this module in 2016 was transferred to the land roster (Module F1 in the household questionnaire) in 2019.
Module AG-C: Plot Roster (Rainy Season)	This module contains the information of agriculture plots owned and/or cultivated by household members during the reference rainy season. More specifically, it reports the location and description and area of the plot.

Continued on next page

Table 2.2 – *continued from previous page*

Module	Description
Module AG-D: Plot Details (Rainy Season)	This module collects detailed plot information (agricultural practices and plot characteristics, use of organic and inorganic fertilizers, use of pesticides/herbicides, and labor inputs) for the reference rainy season. This module also asks a series of questions on sustainable agriculture: trees, cover crops, crop residue disposal, land preparation.
Module AG-E: Affordable Inputs Program (Rainy Season)	This module collects information about quantity/type of ANY input as part of the AIP or say if they were enrolled in E-AIP system and how they were obtained and used during the reference rainy season.
Module AG-F: Other Inputs (Rainy Season)	This module collects information about the inputs used for cultivation and their costs, specifically pesticides and herbicides, during the reference rainy season. It elicits information on the main sources of the input purchased without being enrolled in E-AIP system , any input received for free, any input that was left over from a previous season and own-produced organic fertilizer.
Module AG-G: Crops (Rainy Season)	This module collects information about the crops grown by the household on each plot during the reference rainy season such as the type of crop stand, area of plantation, the amount of seed used and when it was planted, and the details of the harvest.
Module AG-H: Seeds (Rainy Season)	This module collects information about seeds and how they were acquired during the rainy season. More specifically, it elicits information on the main sources of the seed purchased without being enrolled in E-AIP system, any seed received for free, and any seed that was left over from a previous season.
Module AG-I: Sales/Storage (Rainy Season)	This module collects information on the quantity and value of crops sold, the main buyers/outlet, alternative uses, post-harvest losses and storage during the reference rainy season.
Module AG-I_2: Garden Details (Dry/Dimba Season)	This module collects information on details of how the garden was acquired, proceeds from renting out the garden, obligations arising from renting in the garden and decision makers as regards the proceeds from the garden. It also collects information on concerns of ownership of a garden.
Module AG-J: Plot Roster (Dry/Dimba Season)	This module contains the information of agriculture plots owned and/or cultivated by household members during the reference dry (dimba) season. More specifically, it reports the location and description and area of the plot. Enumerators identify whether the plot of land was part of a rainy season or dry season garden.

Continued on next page

Table 2.2 – *continued from previous page*

Module	Description
Module K: Plot Details (Dry Season)	This module collects detailed plot information (agricultural practices and plot characteristics, use of organic and inorganic fertilizers, use of pesticides/herbicides, and labor inputs) for the reference dry (dimba) season.
Module L: Other Inputs (Dry Season)	This module collects information about the inputs used for cultivation and their costs, specifically pesticides and herbicides, during the reference dry (dimba) season. More specifically, it elicits information on the main sources of the input purchased without coupons/vouchers, any input received for free, any input that was left over from a previous season and own-produced organic fertilizer.
Module M: Crops (Dry Season)	This module collects information about the crops grown by the household on each plot during the reference dry (dimba) season such as the type of crop stand, area of plantation, the amount of seed used and when it was planted, and the details of the harvest.
Module N: Seeds (Dry Season)	This module collects information about seeds and how they were acquired during the reference dry (dimba) season. More specifically, it elicits information on the main sources of the seed purchased without coupons/vouchers, any seed received for free, and any seed that was left over from a previous season.
Module O: Sales/Storage (Dry Season)	This module collects information on the quantity and value of crops sold, the main buyers/outlet, alternative uses, post-harvest losses and storage during the reference dry (dimba) season.
Module O_1: Post Harvest Labour (Dry Season)	This is a new module whose questions were added to Module O in 2019. This module collects information on post-harvest labour activities for a particular crop.
Module O_2: Plot Roster Tree Crop Production	This module collects basic information on plots owned and/or cultivated with tree crops by household members during the last 12 months, specifically the area and GPS coordinates of each plot.
Module P: Tree/Permanent Crop Production (Last 12 Months)	This module collects information on crop-stand, area planted, number of trees owned, pre-harvest losses, and amount harvested.
Module Q: Tree/Permanent Crop Sales/Storage (Last 12 Months)	This module collects information on amount sold (value of sales) / given out / used as input for crop by-product / lost / currently in storage.

Continued on next page

Table 2.2 – *continued from previous page*

Module	Description
Module Q_1: Post Harvest Labour (Rainy Season)	This is a new module whose questions were added to Module Q in 2019. This module collects information on post-harvest labour activities for a particular crop.
Module R: Livestock	This module collects information on number currently owned, owners and responsible individuals in the household, inflow/outflow of livestock through various means in the past twelve months, vaccinations, expenditures in the past twelve months on various items.
Module S: Livestock Products	This module collects information on amount produced, sales and expenditures.
Module T: Access to Extension Services	This module collects information on where households receive advice/information on agriculture and how useful the source has been during the last 12 months.

Table 2.3: Contents of the IHPS5 Community Questionnaire

Module	Description
Module CB: Roster of Informants	This module lists the group of informants and their age, sex, positions in community, length of residence in the community, education and language spoken.
Module CC: Basic Information	This module collects basic characteristics of the community, including: population, number of households, major religions, languages spoken, common marriage types, land characteristics and use, number of registered voters and ability to address resource priorities.
Module CD: Access to Basic Services	This module collects information on the community access to and characteristics of transportation networks, markets, ADMARC market, post office, telephone services, churches, schools, health services, and banking services.
Module CE: Economic Activities	This module collects basic information on the primary work activities of community members.
Module CF: Agriculture	This module collects basic information on the prevalence and type of agricultural activities and agricultural facilities.

Continued on next page

Table 2.3 – *continued from previous page*

Module	Description
Module CG: Changes	This module asks respondents to identify changes since 2016 that have made people worse off or better off, such as: drought, flood, changes in prices, changes in access to services, including health facilities, social services, schools, roads, transportation, among others. Additionally, respondent groups are asked to list when these major events occurred and what share of the community they affected.
Module CH: Community Needs, Actions & Achievements	This module asks the respondent group to report on any needs (road and bridge maintenance/construction, school and health center improvement, piped water/boreholes/wells and maize mills construction, orphanage construction, public transportation and law enforcement improvement and the addition of agricultural/fishery/livestock extension services) that community members have expressed during the last 3 years. It then details whether or not the community members took any action to meet these needs and how they went about doing so.
Module CI: Communal Resource Management	This module collects information on communal resources owned by the community and how the rules of access are determined. It further elicits information about how compliance with these rules is enforced among both community members and outsiders.
Module CJ: Communal Organization	This module asks the informed respondent group to report on the presence in the community of listed organizations. It further collects information on the number of specific groups, meeting frequency, size of membership, female and younger adult participation.
Module CK	This module asks the informed respondent group to report on selected items available for sale in the community or nearby at the time of the interview. It further collects information on the prices of these items if they are available in the community.

Chapter 3: Sampling and Weighting Design

3.1 Sampling Design

The IHPS2024 aims to produce representative and reliable estimates at the national level as well as for urban and rural areas. The sample is also designed to provide representative estimates across the three regions (Northern, Central, and Southern), albeit with a lower degree of precision. Furthermore, it aims to provide insights into the longitudinal dynamics occurring within the country since 2010 when the first IHPS was fielded.

The IHPS2024 sampling design builds upon the longitudinal sample of enumeration areas (EAs) and households that were selected and interviewed in previous waves of the IHPS. The details of the original sampling design of the IHPS are documented here. The longitudinal sample has undergone some changes through the course of its lifecycle thus far. The original sample consisted of 204 EAs allocated across six strata corresponding to the urban and rural areas within the three regions. The 204 EAs and associated households were a subsample of those selected for the Third Integrated Household Survey (IHS3) conducted in 2010/11. In 2016, the IHPS sample was cut in half to 102 EAs and all households within those EAs were retained. Furthermore, the IHPS adopted individual "following rules" whereby all individuals who were previous members of any interviewed households are located and interviewed, even when they have "split-off" from the originally sampled households and started their own household or joined another. This includes, for example, when children leave to start their own independent household. As a result of adhering to these individual following rules, the number of households in the IHPS sample has been increasing over time, by nearly 50 percent between 2010 and 2019 (within the 102 EAs).

The IHPS sample underwent another important structural change in 2024: a partial refresh, whereby 50 percent of the 102 EAs were retired (and all associated households emanating from those EAs) and replaced with an equivalent set of freshly selected EAs and households within them. There were several motivating factors behind this partial refresh. The first was the desire to alleviate the respondent burden of some households which had been participating in the IHPS across a decade. The second was the desire to boost cross-sectional representativeness of the IHPS with freshly identified and selected households. As longitudinal samples age, their cross-sectional representativeness tends to fade since the underlying population undergoes socio-demographic changes over time. While this deterioration in cross-sectional representativeness is partially offset through implementation practices (e.g., following of split-off households) and weight adjustments (e.g., calibration), incorporating a freshly selected set of households is the most effective strategy to improve cross-sectional representativeness. The third motivation was to carefully curb and regulate the expansion of the panel sample size induced by the individual following rules.

Under the partial refresh, 51 of the original 102 EAs were retired and 51 retained for the IHPS2024 sample. All associated households residing in or originating from the retained EAs (including split-off households or households that moved out of the EA) were likewise retained and attempted to be interviewed in the IHPS2024. The sample of 51 retained EAs was then complemented by a refreshment sample of 51 EAs. The allocation of the EAs for retention and refreshment was done with a fixed proportion of 50 percent across the six strata according to the

original allocation of 102 EAs inherited from the previous wave of the IHPS (IHPS2019). Table 3.1 presents the distribution and allocation of EAs across the six strata, with the breakdown in terms of the sample of retained versus refreshment EAs. Following the spirit of the original IHPS sample from 2010, the refreshment sample EAs were subsampled from the Sixth Integrated Household Survey (IHS6) whose fieldwork operations coincided with the IHPS2024. The sections that follow provide more details on the selection of both the retained and refreshment samples.

Table 3.1: First-stage allocation of the IHPS2024 sample and its retained and refreshment components across sampling strata.

Stratum	**Retained**	**Refreshment**	**Total**
Northern – Urban	3	3	6
Northern – Rural	3	3	6
Central – Urban	6	6	12
Central – Rural	15	15	30
Southern – Urban	6	6	12
Southern – Rural	18	18	36
Total	**51**	**51**	**102**

3.1.1 Design and selection of the retained sample

The retained sample of 51 EAs was selected from among the 102 EAs that constituted the IHPS2019 sample of EAs following the 50 percent sample cut performed in 2016. The EAs to be retained were selected from within the six strata with equal probability according to the allocation in Table 3.1. Furthermore, the EAs were selected systematically with the EAs first being sorted within strata according to sample type, district, TA, and then EA ID. Sample type was a binary designation applied to the original 102 EAs which historically dictated the schedule by which the sections of the questionnaire were administered (Panel Group A and B). See the documentation from previous IHPS surveys here for more details. It was deemed desirable to ensure that there was a relative balance of these two sample types within the retained sample, which was the case in the 102 EAs from IHPS2019. All households linked to the retained sample of 51 EAs (and interviewed in 2019) were themselves retained and attempted to be interviewed in the IHPS2024.

Following this design, the baseline weights for the retained sample were simply the weights attached to retained households inflated by a factor of 2 which corresponds to the inverse of the probability of an EA being selected from the 102 EAs (a constant 50 percent). These weights served as the baseline and were subsequently adjusted after survey completion, as described in Section 3.2.1 and onward.

3.1.2 Design and selection of the refreshment sample

To preserve full integration between the IHPS2024 and IHS6 surveys, the 51 EAs to be included in the IHPS2024 refreshment sample (see Table 3.1) were randomly selected as a subsample of the first-stage component of the IHS6 sample. More specifically, owing to time constraints and logistical considerations, the decision was made to use as the sampling frame the subset of IHS6 EAs allocated to the second quarter of data collection (Q2). Since the IHS6 sample was designed

as four independent quarterly replicates sharing an identical sampling design, the restriction to its Q2 replicate did not jeopardize the probability nature of the IHPS2024 refreshment sample.

It is important to note that, although both IHS6 and IHPS2024 are stratified samples, their stratification criteria differ.

IHS6 was explicitly stratified by district, yielding 32 sampling strata, with urban and rural areas serving as implicit stratification variables. In addition, IHS6 was designed to generate approximately constant household weights within sampling strata, thereby resulting in a nearly self-weighting sample at the district level. Conversely, IHS6 household weights varied significantly across districts.

For IHPS2024, explicit stratification was defined instead by crossing region (Northern, Central, and Southern) with urban–rural status, yielding six sampling strata. Accordingly, the IHPS2024 sampling strata cut across those of IHS6: each IHPS2024 stratum includes IHS6 EAs and households from different districts, thereby potentially resulting in substantial within-stratum variability in sampling weights for the refreshment sample. To counteract this potential loss in statistical efficiency, the subsampling probabilities of the IHS6 EAs had to be carefully designed and fine-tuned, as detailed in the following procedure.

Denote by p_{Si} the subsampling probability of IHS6 EA i in stratum S ($S = 1, \ldots, 6$) of IHPS2024, and let π_{Di}^{IHS} be the inclusion probability of the same EA in stratum D ($D = 1, \ldots, 32$) of the Q2 quarterly replicate of the IHS6 sample, where the district D belongs to the region identified by stratum S of IHPS2024. Since the selection of the IHS6 sample from the 2018 Malawi PHC frame and the subsequent selection of the IHPS2024 refreshment sample from the IHS6 sample are stochastically independent, the inclusion probability π_{Si}^{REF} of EA i in stratum S of the IHPS2024 refreshment sample from the 2018 Malawi PHC frame can be expressed as the following product:

$$\pi_{Si}^{REF} = \frac{1}{4} \times \pi_{Di}^{IHS} \times p_{Si} \tag{E-1}$$

where the factor 1/4 reflects the restriction to the Q2 quarterly replicate of the IHS6 sample. This equation suggests that appropriate calibration of the subsampling probabilities p_{Si} could lead to the following desirable form for the inclusion probabilities for the EAs of the refreshment sample:

$$\pi_{Si}^{REF} = \frac{1}{4} \times \pi_{Di}^{IHS} \times p_{Si} = n_S \frac{M_{Si}}{M_S} \tag{E-2}$$

where n_S is the number of EAs allocated to stratum S of the IHPS2024 refreshment sample, M_{Si} is the total number of households in EA i of stratum S (i.e., the MOS of the EA in the 2018 PHC frame), and M_S is the total number of households across all EAs in stratum S (i.e., the total MOS of stratum S in the 2018 PHC frame).

Note that Equation (E-2) would associate to each EA selected in the IHPS2024 refreshment sample the same inclusion probability it would have under a direct PPS selection from the 2018 PHC frame. This would, in turn, make the refreshment household sample nearly self-weighting within each of the six (region × urban/rural) IHPS2024 strata.

To achieve this result, the EA subsampling probabilities were set to the following fine-tuned values:

$$p_{Si} = \left(n_S \frac{M_{Si}}{M_S} \right) \Big/ \left(\frac{1}{4} \pi_{Di}^{IHS} \right) \tag{E-3}$$

and the refreshment EA sample was selected from the Q2 IHS6 first-stage sample with unequal probabilities within the six (region × urban/rural) strata.[1]

Fresh household listings were then conducted in the selected EAs to establish up-to-date household-level sampling frames for second-stage selection. From each EA, a "main sample" of 16 households was selected using systematic sampling with equal probability. Furthermore, a "reserve sample" of additional 5 households per EA was selected using the same procedure to facilitate substitution in the event of household nonresponse.

3.2 Weight Calculation

IHPS2024 is the fifth wave of the IHPS panel survey that began with IHPS2010. As such, IHPS2024 supports two types of statistical analyses, each requiring a distinct set of weights: cross-sectional and longitudinal. *Cross-sectional* weights are used to produce estimates representative of the survey's target population at the reference time of the wave. *Longitudinal* weights are instead necessary to estimate gross changes over time or to fit models based on linked data across multiple panel waves.

The methodology used to calculate cross-sectional and longitudinal weights for IHPS2024 had two fundamental objectives: (1) to mitigate potential biases and (2) to improve estimation efficiency. The procedure was designed to produce integrated individual-household weights. As a result, for IHPS2024, individual weights are constant within each household and are equal to the corresponding household weight. This applies to both cross-sectional and longitudinal weights.

The weight calculation procedures for cross-sectional and longitudinal weights are documented separately in the following sections. The main difference is that all responding units in IHPS2024 were assigned a cross-sectional weight, regardless of whether they originated from the retained or refreshment component, whereas longitudinal weights were produced exclusively for the retained component, as only this component can be consistently linked to previous IHPS waves.

The IHPS2024 respondent sample contains 2,763 households and 12,050 individuals. Of these, 1,948 households and 8,475 individuals originated from the retained sample component, whereas 815 households and 3,575 individuals originated from the refreshment sample component. Table 3.2 presents the distribution of respondent households and individuals across the six sampling strata, with the breakdown in terms of retained versus refreshment sample components.

[1] On a technical note, since the direct application of Equation (E-3) would have generated a subsampling probability exceeding one for a specific EA in the Mzimba district within the North – Urban stratum, the quantities p_{Si} were instead used as a synthetic measure of size for PPS selection. This approach yields inclusion probabilities proportional to the target values π_{Si}^{REF} in Equation (E-2), thereby fully preserving the self-weighting property of the refreshment household sample within each stratum.

Table 3.2: Respondent households and individuals in the IHPS2024 sample and its retained and refreshment components across sampling strata.

	Households			Individuals		
Stratum	**Retained**	**Refreshment**	**Total**	**Retained**	**Refreshment**	**Total**
Northern – Urban	108	48	156	505	225	730
Northern – Rural	86	48	134	441	241	682
Central – Urban	202	95	297	943	439	1,382
Central – Rural	608	240	848	2,671	1,028	3,699
Southern – Urban	158	96	254	697	413	1,110
Southern – Rural	786	288	1,074	3,218	1,229	4,447
Total	**1,948**	**815**	**2,763**	**8,475**	**3,575**	**12,050**

3.2.1 Calculation of Cross-Sectional Weights

This section outlines the main procedural steps leading to the construction of the final cross-sectional weights for IHPS2024. The procedure is organized into three stages: (i) weighting steps applied to the retained sample (C1–C3), (ii) weighting steps applied to the refreshment sample (C4–C5), and (iii) adjustment steps applied to the combined sample after integrating the two sample components (C6–C8). Subsequent sections provide detailed descriptions of each step.

- **[C1] Retained sample** – Calculation of base weights.
- **[C2] Retained sample** – Weight-share adjustment for changes in household composition since the previous wave.
- **[C3] Retained sample** – Attrition adjustment.
- **[C4] Refreshment sample** – Calculation of design weights.
- **[C5] Refreshment sample** – Nonresponse adjustment.
- **[C6] Combined sample (retained + refreshment)** – Pooling adjustment.
- **[C7] Combined sample (retained + refreshment)** – Calibration adjustment.
- **[C8] Combined sample (retained + refreshment)** – Trimming adjustment.

Step C8 ended the procedure. At that stage, the cross-sectional weights of respondent households and individuals were ready for dissemination and to be used for general purpose statistical analysis.

Retained Sample – Calculation of Base Weights

Historically, the IHPS household panel survey has adopted the so-called "infinite-contagion" variant of individual-level following rules. Therefore, the retained sample component of IHPS2024 targeted all individuals who responded to the previous wave (IHPS2019) within the retained 51 EAs, as well as all members of their current households, regardless of whether the latter were previously part of the panel. Accordingly, the IHPS2024 retained sample includes both

households that did not change composition from IHPS2019 and those that incorporated "new individuals" who moved in with IHPS2019 panel members from outside the panel.[2]

As usual for panel surveys, the base weight of household k of the retained sample, d_k^{RET}, was initially set equal to the final (i.e., nonresponse-adjusted and calibrated) weight of the originating household k' from IHPS2019, $w_{k'}^{2019}$, and then multiplied by a factor of 2 to reflect the 50% equal-probability subsampling described in Section 3.1.1:

$$d_k^{RET} = 2 \times w_{k'}^{2019} \tag{E-4}$$

Quantitatively, the 1,948 respondent households in the retained IHPS2024 sample originated from 1,338 households that responded to IHPS2019, representing an increase of 610 households (split-offs). In other words, under the "infinite-contagion" model, household splitting dynamics in Malawi resulted in a net increase of 45.6% in the retained sample size over approximately five years. As highlighted in Section 3.1, the attempt to limit this rapid sample expansion was among the primary motivations for retaining only 50% of the panel sample from the previous wave.

Retained Sample – Weight-Share Adjustment for Changes in Household Composition Since the Previous Wave

The "carryover weights" defined in Equation (E-4) were then adjusted to account for changes in household composition over time. Specifically, the weight-share method was used to decrease the weight of households that incorporated "new individuals" from outside the panel. Intuitively, this is because such households could have been included in the panel in multiple ways (either through sample selection at baseline or as a side effect of the following rules in earlier waves) and therefore have a higher probability of inclusion in the current wave than other households. Under the simplifying assumption that the "new individuals" who joined a retained IHPS2024 household originated from the same external household,[3] the weight-share adjustment was calculated as:

$$w_k^{RET,ws} = \begin{cases} d_k^{RET} & \text{if household } k \text{ has no new individuals} \\ d_k^{RET}/2 & \text{otherwise} \end{cases} \tag{E-5}$$

Quantitatively, 1,020 of the 1,948 respondent households in the retained IHPS2024 sample incorporated "new individuals" who were not part of the IHPS2019 panel (about 52% of the retained household sample).

Retained Sample – Attrition Adjustment

The respondent household sample of IHPS2019 within the 51 EAs retained for IHPS2024 comprised 1,515 households. Of these, 1,338 households were successfully followed in 2024, eventually resulting in a total of 1,948 IHPS2024 households owing to household-split events (see

[2]Note that children born to previous panel members between IHPS2019 and IHPS2024 are not considered "new individuals", as they could not have been sampled or included in earlier waves of the panel, even in principle.

[3]This is a standard assumption in the absence of detailed linkage information on origin households.

Table 3.2). The remaining 177 IHPS2019 households ($177 = 1{,}515 - 1{,}338$) were lost to follow-up and thus represent panel attrition. The resulting household attrition rate over the five-year inter-wave period is approximately 11.7%.

To mitigate potential attrition bias, the IHPS2024 retained sample weights were adjusted for attrition. Given the derivation of the retained sample from the previous IHPS2019 wave, rich information was available on both respondent and non-respondent households, making it possible to use the response propensity modeling approach (RPM) to adjust the household weights for attrition. The implemented procedure (also referred to as the "propensity score method") entailed four steps:

- First, a logistic model was developed to estimate household-level response probabilities, using a binary indicator, $resp_k = \text{YES/NO}$, as the dependent variable and multiple variables derived from IHPS2019, z_k, as predictors. Candidate predictors spanned different domains, including territory, socio-demographics, housing, and assets. After careful exploration, a parsimonious subset of powerful predictor variables was selected.[4]
- Second, the logistic model was fitted and used to predict response probabilities, $\hat{p}_k = Pr(resp_k = \text{YES} \mid z_k)$, for all eligible IHPS2024 households (i.e., both respondent and non-respondent).
- Third, deciles of predicted response probabilities were calculated and used to cluster respondent and non-respondent households into ten reasonably homogeneous and non-overlapping classes, P_q, for $q = 1, \ldots, 10$.
- Fourth, the ten response propensity classes, P_q, were treated as Response Homogeneity Groups (RHGs), and household-level attrition adjustment factors, g_k^{attr}, were computed for respondent households as reciprocals of weighted response rates within each class.

Note that the propensity score method summarized here uses the fitted logistic model solely to define the Response Homogeneity Groups. In other words, it does *not* use the predicted response probabilities, $\hat{p}_k$, to compute the attrition adjustment factor, g_k^{attr}. More precisely, in the standard RPM approach, attrition adjustment factors would be calculated as the inverse of the average $\hat{p}_k$ within each class. Instead, the approach adopted here computes them as reciprocals of weighted response rates within each class (i.e., RHG). This choice is generally more robust to model misspecification and less prone to generating extreme weight adjustments, which may lead to unstable estimates.

In formulas, the attrition adjusted weights for the respondent IHPS2024 retained households, $w_k^{RET,attr}$, were obtained by multiplying the weight-share adjusted weights from Equation (E-5) by the attrition adjustment factors, g_k^{attr}, calculated using the propensity score method:

$$w_k^{RET,attr} = g_k^{attr} \times w_k^{RET,ws} = \left(\frac{\sum_{j \in s_{P_q}} w_j^{RET,ws}}{\sum_{j \in r_{P_q}} w_j^{RET,ws}} \right) \times w_k^{RET,ws} \tag{E-6}$$

[4] These variables, selected through an AIC (Akaike Information Criterion) minimization procedure, are: district, household size, number of dependents, marital status of household head, number of rooms per capita, dwelling is owned, dwelling is rented, dwelling has modern roof, dwelling has modern walls, household has electricity, household has improved water source, dwelling permanent construction, dwelling semipermanent construction, household owns mobile phone, number of mobile phones, household has bank account, household owns a car, household owns a fridge.

where $k \in P_q$ for $q = 1, \ldots, 10$, and r_{P_q} and s_{P_q} denote the set of respondent and eligible households within the generic response propensity class P_q of the IHPS2024 retained component. Importantly, as seen from Equation (E-5), the attrition adjustment step could only increase (or, at most, leave unchanged) the weight-share adjusted weights it received as input.

Refreshment Sample – Calculation of Design Weights

The refreshment component of the IHPS2024 sample comprises 51 EAs. These EAs were allocated to the six (region × urban/rural) sampling strata as reported in Table 3.1 and selected within each stratum from the Q2 replicate of the IHS6 first-stage sample with the fine-tuned, unequal subsampling probabilities defined in Section 3.1.2, see Equation (E-3).

Fresh household listings were then conducted in these EAs to establish up-to-date household-level sampling frames for second-stage selection. Fieldwork operations unfolded smoothly, as all 51 targeted EAs turned out to be accessible and were therefore visited. From each EA, 16 households were selected using systematic sampling with equal probability ("main sample"). In addition, 5 reserve households per EA were selected using the same procedure to facilitate substitution in the event of household nonresponse ("reserve sample").

The calculation of design weights for the refreshment households strictly adhered to the sampling design described above. Accordingly, the design weight d_{Sik}^{REF} of household k, belonging to EA i, sampled within stratum S was computed as:

$$d_{Sik}^{REF} = \left(\pi_{Si}^{REF} \times \frac{m}{M_{Si}^{*}} \right)^{-1} \cong \frac{M_S}{m \cdot n_S} \times \frac{M_{Si}^{*}}{M_{Si}} \tag{E-7}$$

where n_S is the number of EAs allocated to stratum S; m is the fixed number of households randomly selected in each sampled EA and targeted in the main sample ($m = 16$); M_{Si} is the total number of households in EA i of stratum S (i.e., the MOS of the EA in the 2018 PHC frame); M_S is the total number of households across all EAs in stratum S (i.e., the total MOS of stratum S in the 2018 PHC frame); M_{Si}^{*} is the total number of households in sampled EA i of stratum S as determined from the listing exercise; π_{Si}^{REF} is defined in Equation (E-1); and the approximate equality in Equation (E-7) rests on the technical detail discussed in footnote 1.

Refreshment Sample – Nonresponse Adjustment

The planned sample size for the refreshment component of IHPS2024 was 816 households ($816 = 51 \times 16$). Of these, 123 households resulted in nonresponse. All but one of these nonrespondent households were successfully replaced using preselected reserve households, yielding 815 completed interviews. These figures correspond, for the refreshment component of IHPS2024, to a substitution rate of approximately 15% and to a residual nonresponse rate after substitution of approximately 0.1 percent ($1 - 815/816$).

To compensate for this modest level of residual nonresponse, a simple ratio adjustment was applied. Within each EA, the design weights in Equation (E-7) were simply divided by the corresponding EA-level response rates:

$$w_{Sik}^{REF,nr} = \frac{d_{Sik}^{REF}}{(m/m_i^r)} \tag{E-8}$$

where $m = 16$ is the planned number of households per EA, and m_i^r is the number of respondent households in EA i. Simply stated, the net effect of this nonresponse adjustment was to multiply by 16/15 the design weights in the EA comprising the only nonrespondent household that had not been replaced, while leaving the weights unchanged in the remaining EAs.

While household substitution improves estimation efficiency by reducing sample loss due to nonresponse, it cannot guarantee the elimination of nonresponse bias. Nonresponse bias may indeed persist due to systematic differences between nonresponding units and their replacements. Unfortunately, such differences (if any) were impossible to assess for the IHPS2024 refreshment sample, due to its cross-sectional nature. In fact, for the refreshment sample, in contrast with the retained sample, no information at all was available for nonrespondents, thereby precluding nonresponse weight adjustments based on explicit modeling approaches. As an alternative, any residual nonresponse bias still affecting the refreshment sample after substitution was addressed through the final calibration adjustment applied to the combined IHPS sample (step C7 of the procedure, see Section 3.2.1).

In the following sections, unnecessary subscripts in Equation (E-8) will be dropped for notational convenience, and the nonresponse adjusted weight of household k in the IHPS2024 refreshment sample will be simply denoted by $w_k^{REF,nr}$.

Combined Sample (Retained + Refreshment) – Pooling Adjustment

The IHPS2024 sample is obtained by combining (or "pooling") the retained and refreshment samples. Users must be able to analyze this combined sample as a whole and compute valid estimates from it without any reference to whether a given record originated from the retained or refreshment component. To this end, a "pooling adjustment" factor was computed and applied to the IHPS2024 weights.

Since the retained and refreshment samples (i) are stochastically independent and (ii) represent the same target population, their original weights, $w_k^{RET,attr}$ and $w_k^{REF,nr}$, cannot be used as they are, as doing so would lead to incorrect inferential results. The issue is immediately evident, since each separate sample would produce unbiased estimates for any population total Y, so using the unadjusted original weights in the pooled dataset would overestimate Y by a factor of 2. The problem arises because each household in the combined sample could have been included in either of the two sample components, but its original weight only reflects the probability of inclusion in the sample where the household was actually selected. Therefore, to ensure valid analysis of the combined sample, the original weights must be adjusted for pooling. This adjustment decreases the original weights to account for the multiple routes of selection of the corresponding units in the combined sample. Accordingly, the pooling adjusted weight of any respondent household k in IHPS2024 was expressed as:

$$w_k^{IHPS,pool} = \begin{cases} \alpha_k^{RET} \times w_k^{RET,attr} & \text{if } k \in \text{RET sample} \\ \alpha_k^{REF} \times w_k^{REF,nr} & \text{if } k \in \text{REF sample} \end{cases} \tag{E-9}$$

where the pooling adjustment factors are such that $\alpha_k^{RET} \in [0,1]$ and $\alpha_k^{REF} \in [0,1]$ and satisfy the condition $\alpha_k^{RET} + \alpha_k^{REF} = 1$, which guarantees unbiased estimation from the combined sample. While different methods exist to calculate these factors, the approach followed for IHPS2024

is generally regarded as statistically efficient and reasonably simple. It defines stratum-specific pooling adjustment factors as follows:

$$\begin{cases} \alpha_{Sk}^{RET} = \dfrac{m_S^{RET}/UWE_S^{RET}}{(m_S^{RET}/UWE_S^{RET}) + (m_S^{REF}/UWE_S^{REF})} & \text{if } k \in \text{stratum } S \text{ of RET sample} \\ \alpha_{Sk}^{REF} = \dfrac{m_S^{REF}/UWE_S^{REF}}{(m_S^{RET}/UWE_S^{RET}) + (m_S^{REF}/UWE_S^{REF})} & \text{if } k \in \text{stratum } S \text{ of REF sample} \end{cases} \tag{E-10}$$

where m_S^{RET} and m_S^{REF} denote the number of respondent households in stratum S for the retained and refreshment samples, whereas UWE_S^{RET} and UWE_S^{REF} denote the corresponding unequal weighting effects based on the original weights. Following Kish's definition, the UWE of a set of weights w is calculated as one plus the relative sample variance of the weights, $UWE(w) = 1 + CV(w)^2$.

The pooling adjustment factors in Equation (E-10) can be interpreted as the relative *effective* sample sizes at the stratum level for the retained and refreshment sample components. Therefore, within each stratum, the pooling adjusted weights from Equation (E-9) assign a relatively larger influence to households coming from the sample component with the larger effective sample size in the stratum. Notably, although the refreshment sample had a significantly smaller size than the retained sample across all strata, it nonetheless exhibited a higher effective sample size in four strata out of six, owing to its smaller UWE. This illustrates the statistical efficiency of the sampling design described in Section 3.1.2, which was specifically intended to contain the weight variability of the refreshment sample.

Following the pooling adjustment, all subsequent weighting steps were applied to the combined IHPS2024 sample as a single integrated dataset, without any further reference to its retained and refreshment components.

Combined Sample (Retained + Refreshment) – Calibration Adjustment

The calibration adjustment implemented for IHPS2024 aimed to provide an additional layer of protection against bias (e.g., residual nonresponse bias after substitution), improve the precision of cross-sectional estimates, and ensure alignment with population totals from trusted external sources.

The calibration process used as calibration benchmarks both (i) individual-level population totals and (ii) household-level population totals. The individual-level totals were derived from official population projections compiled and published by the Malawi NSO. These calibration benchmarks represent projected counts of individuals by region, five-year age class, and sex for mid-2024, thus encompassing 90 independent totals (3 regions × 15 age classes × 2 sexes), as reported in Table 3.3.

Table 3.3: Malawi individuals by region, sex, and five-year age class (source: Official population projections for mid-2024).

	Northern region			Central region			Southern region		
Age class	**Male**	**Female**	**Total**	**Male**	**Female**	**Total**	**Male**	**Female**	**Total**
0–4	183,349	177,362	360,711	627,303	618,211	1,245,514	704,195	692,985	1,397,180
5–9	165,426	165,335	330,761	576,899	576,716	1,153,615	653,043	650,111	1,303,154
10–14	162,452	166,227	328,679	551,242	553,847	1,105,089	566,746	554,007	1,120,753
15–19	148,961	161,524	310,485	494,407	515,087	1,009,494	495,727	517,224	1,012,951
20–24	122,306	140,449	262,755	425,469	458,355	883,824	415,916	469,610	885,526
25–29	100,076	114,714	214,790	362,348	391,155	753,503	337,896	392,807	730,703
30–34	83,302	93,238	176,540	302,603	325,879	628,482	275,376	323,735	599,111
35–39	67,773	76,235	144,008	245,966	267,526	513,492	220,161	264,996	485,157
40–44	55,063	62,884	117,947	199,040	218,505	417,545	175,437	216,178	391,615
45–49	45,848	51,626	97,474	160,607	173,205	333,812	142,913	171,232	314,145
50–54	37,756	41,450	79,206	127,433	134,545	261,978	115,729	133,767	249,496
55–59	29,668	31,993	61,661	95,281	100,509	195,790	88,037	100,802	188,839
60–64	22,567	24,552	47,119	69,252	74,681	143,933	65,895	76,227	142,122
65–69	16,839	18,738	35,577	50,198	57,115	107,313	49,960	60,836	110,796
70+	31,443	36,063	67,506	85,460	114,805	200,265	89,012	124,834	213,846
Total	**1,272,829**	**1,362,390**	**2,635,219**	**4,373,508**	**4,580,141**	**8,953,649**	**4,396,043**	**4,749,351**	**9,145,394**

Because official population projections were not available for the population of households, the second set of calibration benchmarks was derived from IHS6, the large-scale cross-sectional survey conducted simultaneously with IHPS2024. Since IHPS2024 and IHS6 share common field operation protocols and questionnaires, survey estimates based on the IHS6 sample constituted an ideal source of calibration benchmarks for IHPS2024, particularly considering the cross-sectional nature of IHS6 and its significantly larger sample size (12,502 households and 53,641 individuals). As reported in Table 3.4, the second set of benchmarks derived from IHS6 and used to calibrate IHPS2024 represents estimated counts of Malawi households by region, urban/rural status, and 8 household-size classes, thus encompassing 48 independent totals (3 regions $\times$ 2 urban/rural statuses $\times$ 8 household-size classes).

Table 3.4: Malawi households by region, urban/rural status, and household-size classes (source: IHS6 estimates for year 2024, rounded to the nearest integer).

	Northern region			Central region			Southern region		
Household size	**Urban**	**Rural**	**Total**	**Urban**	**Rural**	**Total**	**Urban**	**Rural**	**Total**
1	8,486	23,342	31,828	33,922	80,912	114,834	33,066	120,899	153,965
2	10,613	43,375	53,988	42,541	158,804	201,345	43,678	218,396	262,074
3	21,882	78,941	100,823	69,818	388,351	458,169	76,205	400,548	476,754
4	28,358	96,693	125,051	76,185	430,763	506,948	82,680	417,648	500,327
5	17,622	78,591	96,214	56,777	315,518	372,295	54,957	337,910	392,867
6	12,434	55,313	67,747	47,873	194,860	242,734	30,452	200,306	230,758
7	7,547	36,235	43,783	22,149	109,902	132,051	19,348	112,082	131,431
8+	9,256	46,539	55,794	20,902	69,048	89,950	19,768	69,380	89,148
Total	**116,198**	**459,029**	**575,227**	**370,168**	**1,748,159**	**2,118,327**	**360,155**	**1,877,169**	**2,237,324**

Importantly, a single set of calibration weights was sought that would allow IHPS2024 to simultaneously fulfill all the 138 (= 90 + 48) calibration constraints reported in Table 3.3 and Table 3.4. To solve the calibration problem, the R software ReGenesees (Zardetto, 2015 and 2024) was used. Although the two sets of control totals referred to different units – individuals

and households – the software successfully incorporated both sets of constraints simultaneously, producing a single set of *integrated* calibration weights (i.e., identical weights across members of the same household). On a technical note, a *range-restricted* calibration algorithm was applied to avoid negative or exceedingly large calibration weights. Exact convergence of the calibration algorithm was obtained: all the 138 calibration benchmarks were perfectly matched. By enforcing consistency with both individual-level and household-level population counts from reliable external sources, the IHPS2024 calibration weights also implicitly control for average household size, a critically important population parameter for poverty estimation.

After calibration, the integrated IHPS2024 weight of sampling unit k (household or individual) can be expressed in terms of the pooling adjusted weights in Equation (E-9) as follows:

$$w_k^{IHPS,cal} = g_k^{cal} \times w_k^{IHPS,pool} \tag{E-11}$$

As usual for complex calibration tasks, the calibration adjustment factors, g_k^{cal}, in Equation (E-11) cannot be expressed in closed analytic form: their values were computed through iterative numerical optimization routines.

Combined Sample (Retained + Refreshment) – Trimming Adjustment

The range-restricted calibration procedure described above successfully prevented the generation of negative weights. Nonetheless, a modest number of very small and very large weights were produced. Following extensive diagnostic checks and a careful inspection of the calibration weight distribution, two acceptable threshold values, $w_l = 2.875$ and $w_u = 18{,}605$, were identified, and the decision was made to trim the calibration weights falling outside this interval.

Out of 2,763 respondent households, only 2 calibration weights were below w_l, and only 3 exceeded w_u. The ReGenesees software was used to implement the trimming adjustment in a *calibration-consistent* manner. This allowed the IHPS2024 calibration weights to be restricted to the desired interval $[w_l, w_u]$ while simultaneously (i) preserving all the calibration constraints and (ii) maintaining the individual-household integration property of the weights. In other words, even after trimming, the IHPS2024 cross-sectional weights remained integrated and continued to reproduce, in estimation, all 138 population benchmarks used in the calibration step C7.

The trimming adjustment step completed the cross-sectional weighting procedure for IHPS2024, yielding the final, dissemination-ready weights for both respondent households and individuals. Formally, these trimmed weights can be expressed, for the generic sampling unit k (household or individual), in terms of the calibration adjusted weights in Equation (E-11) as follows:

$$w_k^{IHPS,trim} = g_k^{trim} \times w_k^{IHPS,cal} \tag{E-12}$$

As was the case for the previous calibration step, the values of the trimming adjustment factors, g_k^{trim}, in Equation (E-12) were computed through iterative numerical optimization routines and cannot be expressed in closed analytic form.

3.2.2 Impact of Adjustments on the Weight Distribution

Table 3.5 summarizes the evolution of the sample distribution of the respondent household weights in IHPS2024 along the steps of the weight calculation process schematically illustrated

in Section 3.2.1. Since the variability of the weights tends to negatively affect the precision of survey estimates, Kish's UWE (see Section 3.2.1) is also reported for each set of weights.

Table 3.5: Summary of the sample distribution of IHPS2024 cross-sectional household weights.

3.5.1 Retained sample: $n = 1{,}948$

Household Weights	Min	1st Q	Median	Mean	3rd Q	Max	UWE
Base	16.92	681.08	1,668.22	2,689.16	3,432.22	42,110.76	2.51
Weight-share adjusted	8.46	454.31	1,114.99	1,935.53	2,449.48	42,110.76	2.74
Attrition adjusted	8.77	519.84	1,214.07	2,115.61	2,647.94	42,480.26	2.75

3.5.2 Refreshment sample: $n = 815$

Household Weights	Min	1st Q	Median	Mean	3rd Q	Max	UWE
Design	2,433.70	4,065.27	5,146.99	5,342.20	6,228.07	8,732.11	1.09
Nonresponse adjusted	2,433.70	4,065.27	5,146.99	5,349.82	6,278.20	8,732.11	1.09

3.5.3 Combined sample: $n = 2{,}763$

Household Weights	Min	1st Q	Median	Mean	3rd Q	Max	UWE
Adjusted for pooling	2.87	380.68	1,182.01	1,567.19	2,395.23	14,982.58	1.90
Calibrated	1.97	387.19	1,114.69	1,784.61	2,588.61	20,780.70	2.30
Trimmed	2.88	388.28	1,121.08	1,784.61	2,589.29	18,605.39	2.28

It can be noted from Table 3.5.1 that the base weights of the *retained sample component* were already highly variable at the outset ($UWE = 2.51$). This is expected in a long-running household panel survey, where the inherent weight variability induced by the complex baseline sampling design is further magnified by the cumulative effects of the weight adjustments applied over successive panel waves. The main effects of the two subsequent adjustment steps – for changes in household composition and for attrition – are clearly visible in the table. Consistent with Equation (E-5), the weight-share adjustment shifted the weight distribution to the left. It also significantly increased the dispersion of the weights, inflating the UWE from 2.51 to 2.74. The subsequent attrition adjustment shifted the weight distribution to the right and modestly increased the maximum weight, as expected from Equation (E-6). However, it left the UWE virtually unchanged ($UWE = 2.75$).

As evident from Table 3.5.2, the design weights of the *refreshment sample component* were significantly less variable than those of the retained sample ($UWE = 1.09$). This is consistent with the approximately self-weighting design within strata described in Section 3.1.2. Given the negligible residual nonresponse rate after substitution documented in Section 3.2.1, the impact of the nonresponse adjustment was marginal, only slightly increasing the mean and third quartile of the weight distribution.

Table 3.5.3 highlights the beneficial effects of integrating the retained and refreshment components into the *combined sample*. The pooling adjustment brought the UWE of the combined sample to 1.90, an intermediate level between those of the component samples. The impact of the subsequent calibration adjustment was substantial. First, calibration significantly increased the mean of the weight distribution (+14%), indicating that even after the adjustments for

attrition and nonresponse, the combined sample was still underestimating the total number of Malawi households implied by the external sources (see Section 3.2.1). Second, calibration widened the range of the weights, making the right tail of the distribution more pronounced. As a side effect, the dispersion of the weights increased substantially, with the UWE rising from 1.90 to 2.30: a trade-off that reflects the enhanced bias protection achieved through calibration. The final trimming step carefully shrank the range of the weights, forcing 2 very small and 3 very large calibration weights to fall within the acceptable bounds discussed in Section 3.2.1. At the same time, the remaining weights were minimally adjusted to preserve the 138 calibration constraints. Overall, the trimming step decreased the UWE from 2.30 to 2.28.

3.2.3 Calculation of Longitudinal Weights

As highlighted in Section 3.2, longitudinal weights were produced exclusively for the retained sample component of IHPS2024. This sample, comprising 1,948 respondent households and 8,475 respondent individuals, originated from the IHPS2019 respondent sample through the application of individual-level following rules. Therefore, it can be consistently linked to previous IHPS panel waves to generate longitudinal datasets and underpin longitudinal analysis.

This section lists and concisely describes the main procedural steps that led to the final longitudinal weights of the IHPS2024 retained sample. Since most of the steps are common to both cross-sectional and longitudinal weights, the rest of this section focuses on the key differences between the two weighting procedures. For more details on the methodology, the reader is referred to the corresponding sections of the cross-sectional weight calculation pipeline.

- **[L1] Calculation of base weights:** The base longitudinal weights, d_k^{LONG}, were identical to the base cross-sectional weights in Equation (E-4).
- **[L2] Weight-share adjustment:** The longitudinal weights adjusted for changes in household composition since the previous wave, $w_k^{LONG,ws}$, were identical to the weight-share adjusted cross-sectional weights in Equation (E-5).
- **[L3] Attrition adjustment:** The longitudinal weights adjusted for attrition, $w_k^{LONG,attr}$, were identical to the attrition adjusted cross-sectional weights in Equation (E-6).
- **[L4] Calibration adjustment:** Although no information is generally available for longitudinal populations from external sources, the decision was made to calibrate the longitudinal weights of the retained sample to the same 138 population totals used to calibrate the cross-sectional weights of the combined (retained + refreshment) sample. These 138 calibration totals are reported in Table 3.3 and Table 3.4. The ReGenesees software was used to compute the calibrated longitudinal weights, $w_k^{LONG,cal} = g_k^{LONG,cal} \times w_k^{LONG,attr}$. As for cross-sectional weights, these longitudinal calibration weights were also integrated (i.e., identical across members of the same household). It is important to note that, despite the use of identical calibration benchmarks, the same unit (household or individual) in the retained IHPS2024 sample generally exhibits markedly different longitudinal and cross-sectional calibration weights. This can be appreciated by comparing the "Calibrated" entry of Table 3.5 and Table 3.6.
- **[L5] Trimming adjustment:** A thorough inspection of the calibration weights led to the identification of a single extreme value. The ReGenesees software was used to trim this potentially influential household weight, while simultaneously preserving all 138 calibration constraints, as well as the individual-household integration property of the longitudinal

weights.

The trimming adjustment step L5 completed the longitudinal weighting procedure for IHPS2024, producing the final dissemination-ready longitudinal weights, $w_k^{LONG,trim} = g_k^{LONG,trim} \times w_k^{LONG,cal}$, for both households and individuals.

Table 3.6 summarizes the evolution of the sample distribution of the longitudinal household weights along the steps of the weight calculation process, also reporting their UWE.

Table 3.6: Summary of the sample distribution of IHPS2024 longitudinal household weights.

Retained sample: $n = 1{,}948$

Household Weights	**Min**	**1st Q**	**Median**	**Mean**	**3rd Q**	**Max**	**UWE**
Base	16.92	681.08	1,668.22	2,689.16	3,432.22	42,110.76	2.51
Weight-share adjusted	8.46	454.31	1,114.99	1,935.53	2,449.48	42,110.76	2.74
Attrition adjusted	8.77	519.84	1,214.07	2,115.61	2,647.94	42,480.26	2.75
Calibrated	3.60	543.14	1,396.11	2,531.25	3,141.33	38,522.62	2.83
Trimmed	3.60	543.14	1,396.11	2,531.25	3,149.70	35,436.38	2.82

While calibration and trimming slightly increased the variability of longitudinal weights (the UWE increased by less than 3%, from 2.75 to 2.82), they effectively reduced extreme weights (the maximum weight decreased by approximately 17%).

Chapter 4: Organization of the Survey

Organization of the survey is one of the critical aspect in ensuring collection of high-quality, timely and reliable data. It details institutional arrangements, organizational arrangements, and operational mechanisms with clearly defined roles and responsibilities of different personnel at every organization level.

4.1 Institutional arrangement

The Economic Statistics Division of the National Statistical Office (NSO) conducted the IHS6 in accordance with its mandate under the National Statistical Act (2013). The preparatory process commenced with a formal Stakeholders Consultative Workshop held from 23rd to 24th February 2023 at the Sogecoa Golden Peacock Hotel, Lilongwe, which formally inaugurated the survey alongside the Fourth Integrated Household Panel Survey (IHPS4).

The workshop brought together 46 participants drawn from a broad cross-section of national and international institutions. These included the World Bank, the United Nations Development Programme (UNDP), the United Nations Population Fund (UNFPA), the World Food Programme (WFP), the International Food Policy Research Institute (IFPRI), the Millennium Challenge Account Malawi, the Reserve Bank of Malawi, the National Planning Commission (NPC), the Malawi Economic Justice Network (MEJN), and the Lilongwe University of Agriculture and Natural Resources (LUANAR). Government representation spanned the Ministries of Finance and Economic Affairs, Health, Education, Gender, Labour, Trade and Industry, Tourism, and Local Government, as well as the Environment, Planning and Development department, academia and the Centre for Social Research. The National Statistical Office was represented by the Commissioner of Statistics, senior technical staff, and supporting personnel.

The NSO's Project Steering Committee was responsible for coordinating IHS6 activities and providing institutional-level oversight of the project. Subordinate to the Steering Committee was the Project Implementation Unit (PIU), which served as the technical arm of the World Bank and provided both knowledge-based and financial support.

The survey secretariat was headed by the National Survey Coordinator, who also served as Secretary to the Project Steering Committee. The Coordinator was responsible for documentation, logistical coordination, and reporting to the Steering Committee, the overarching management body, on all associated activities.

Below the Project Steering Committee, the survey structure comprised the following bodies:

1. **Technical Advisory Committee and Working Groups:** These bodies were responsible for designing survey methodologies, developing survey instruments, designing CAPI questionnaires and associated technology, and facilitating knowledge transfer including the training of supervisors and enumerators.
2. **Publicity and Advocacy Committee:** This committee was responsible for public sensitization and awareness of the IHS6, with the aim of enhancing public participation during the survey period.

Prior to enumeration, the Human Resource Management unit was responsible for the recruitment and contract management of temporary field staff. Within the Logistics unit, the Procurement sub-unit managed the purchase of required survey materials; the Transport sub-unit managed field vehicles; and the Finance Department was responsible for the management and disbursement of financial resources to field teams.

During data collection, the Data and ICT unit was responsible for assigning daily work assignments to enumerators, initiating data quality checks, and managing completed questionnaires on the central server. Senior supervisors oversaw survey activities at the national level, including overall field supervision and quality management. Field supervisors were responsible for the daily supervision of enumerators within their jurisdiction and for the administration of the community questionnaire in each sampled enumeration area. Enumerators were responsible for the daily collection of data from assigned households.

4.2 Pre-Enumeration Listing

Pre-enumeration listings were initiated before the start of field work. This was especially necessary for the 51 refreshed Enumeration Areas. Mobile listing teams equipped with printed maps of select EAs were used to record all dwellings and the heads of households in the selected IHPS5 EAs. Household counts per each listed enumeration areas were relayed to the NSO IHPS5 Management, and were recorded. Where applicable, listing forms and maps were transferred directly to field teams for the data collection. The IHPS5 listing was conducted using the CAPI software of the World Bank Survey Solutions.

4.3 IHPS5 pre-testing exercise

The IHS6 pre-testing exercise was conducted shortly after the completion of the pre-enumeration listing. This exercise was carried out in Machinga district, selected to assess questionnaire coherence and flow across the country's major local languages, including Chichewa and Tumbuka.

The primary objectives of the pre-testing exercise were to identify potential issues in the data-collection instruments and operational procedures and to verify the accuracy, reliability, and validity of the survey tools prior to the main data-collection exercise.

Chapter 5: Training of Field Staff

Training of field staff has been one of the core and important aspect during the implementation of the IHPS5 beginning with the pre-enumeration listing exercise up until the actual survey period. This was one of the factors aimed at enhancing production of robust, reliable and accurate data for policy formulation and decision making.

5.1 Training Design

Trainings were undertaken at every stage of survey journey for every activity that required deployment of field staff to carry out data collection as follows:

1. Pre-enumeration listing exercise
2. Pre-testing exercise
3. Enumerators' training for the main survey

Pre-enumeration listing training: This was conducted prior to household listing where both field supervisors and enumerators were jointly trained for 1 week days with field supervisors further provided with an advanced training particularly on assigning the work to enumerators and data quality checks using the CAPI.

Pre-testing training: This was a 2-week training carried out prior to the pre-testing exercise. Similar to the pre-enumeration listing, it targeted both the field supervisors.

Enumerators' training: This was a 4-week long training involving both field supervisors who also acted as data collectors particularly for community questionnaire but also enumerators who were assigned to administer questionnaires for to the households. The training covered theoretical components of the questionnaires, interviewing techniques, survey or research ethical consideration and other cross cutting issues like; (i). HIV and AIDS at workplace, and (ii). Environmental and social safeguards following the World Bank standards. Further, the training also covered the practical aspect of the questionnaire using World Bank's Survey Solutions CAPI platform.

5.1.1 Training materials and methods

All the trainings were conducted using presentations that were prepared by the trainers prior to every training session covering (i). questionnaires, concepts and definitions, (ii). interviewing techniques, (iii). maps, (iv). HIV and AIDS at workplace, (v). Environmental and Social safeguard, and (vi). field ethics. Furthermore using tablets, practical sessions for the CAPI were conducted that centered much on pair mock interviews among the enumerators.

The trainings were held in a classroom instruction but interactive manner in the sense that the trainers would ask questions to the participants to draw their active attention and participation within the entire training period.

5.2 Evaluation of Field Staff

Training evaluation of the field staff or trainees was undertaken at some intervals of the training session but also at the end of the entire training period. In addition to the aforementioned procedure from the previous section of assessing the trainees, trainers provided tests everyday morning before the commencement of the new day tasks. Further, trainers used to pose random questions to the participants everyday during the recap of previous days training material covered.

Chapter 6: Fieldwork

6.1 Organization of Fieldwork

The field work commenced on 23rd July 2024 across all the districts and ended on in February although there some residual tracking activities that continued until May.

In contrast to previous rounds of the panel component where households were visited twice, in 2024, households were each visited once.

Field Supervisors were responsible for administering community questionnaire. In this regard, every sampled EA community questionnaire was administered, where more than 1 sampled EA share the boundaries, one community questionnaire was administered with the fact that the EAs share similar attributes.

6.2 Pre-loaded Information

Relevant information on every household was pre-loaded in the CAPI assignments for each enumerator. The information pre-loaded include basic household information, i.e., household head name, location (district, traditional authority, village and EA code). Individual level information as well as afew garden and plot details captured in the previous round were also uploaded to allow the enumerators to easily track the households.

The pre-loaded information was used at advanced means used by the enumerators and field staff to easily locate and identify the actual intended household since the KML maps could only be used to locate the assigned EA.

However, some element of the pre-loaded information i.e., household head, contact of the household and description of the physical address were editable to allow modification so that the households correspond to their current attributes for easily identification for future follow-ups whenever necessary.

6.3 Fieldwork Monitoring

Fieldwork monitoring was imperative in order to achieve collection of high-quality, accurate and reliable data. Field monitoring was done through manual using routine field supervision by supervisors and digital based particularly using the data quality checks.

The manual field monitoring had supervisors across three levels of hierarchy as follows:

1. **Central Supervisors:** A team from Headquarter went across all the regions immediately the field staff were deployed to the field followed by other periodic supervision exercises.
2. **Field Supervisors:** Responsible for routine monitoring as well, one field supervisor was attached to 4 enumerators. These were also doing the spot checks but also in-person monitoring and provide technical guidance to the enumerators aimed at improving efficient work output.

The digital-based platform apart from being used for quality control checks was also used to

remotely monitor the data collection progress by every enumerator. Both the HQ-based supervisors and field supervisors were given log-in credentials to aid them in accessing the server particularly to the enumerators assigned to their jurisdiction.

Apart from being used as an aspect to achieve high-quality and accurate data, fieldwork monitoring also assisted to get feedback from the field staff on issues that were affecting the data collection which would at some point affected the data quality but also timeframe for data collection.

Chapter 7: Data Management and Description Datasets

7.1 Data Management

7.1.1 CAPI

The IHPS5 was executed utilising Computer Assisted Personal Interview (CAPI) methodologies. All questionnaires (listing, household sector, non-household sector, and community questionnaires) were administered utilising the CAPI software, *Survey Solutions*[1] . The Survey Solutions software was created and is managed by the Survey Unit of the Development Economics Data Group (DECDG) at the World Bank. Each enumerator received a tablet to facilitate the interviews. The implementation of the survey utilising Survey Solutions CAPI was highly successful, facilitating the prompt availability of data from completed interviews.

7.1.2 Data Communication System

During the IHPS5, a highly automated data communication system was utilised. Every field crew had a tablet that could be used as a hotspot, enabling daily tablet synchronisation and internet connectivity. This guaranteed real-time access to the data at the Zomba NSO headquarters. The data managers initially examined the data after the interview was finished and uploaded to the server. In order to look for any more mistakes that the Survey Solutions application could have missed, the data was also downloaded from the server and ran through a STATA dofile error-checking system as well as an auto-reject system[2] also developed by the LSMS. After the STATA dofile was run on the raw dataset, an Excel error file was produced. The relevant field interviewers were then notified of the information in the excel error files so they could take appropriate action. For the length of the survey, this monitoring activity was carried out every day.

7.1.3 Data Cleaning

The data cleaning process was divided into three main parts. The first step was to make sure proper quality control was in place during the fieldwork. The Survey Solutions application, which was used to gather the data, was modified to include validation and consistency checks in order to highlight several errors that were committed during the fieldwork. The range and consistency checks built into the application was informed by the LSMS-ISA experience with the IHS3 2010/11, IHPS 2013, IHPS 2016 and IHPS 2019.

The second cleaning step involved the use of data managers (Headquarters in Survey Solutions). As previously stated, after the interview is finished and uploaded to the server, the data managers review it for abnormalities and extreme values. Depending on the outcome, they can choose to accept or reject the case. The case is synced and sent back to the relevant interviewer's tablet if it is rejected. Additional care was taken to guarantee that the households in the data matched the selected sample and that any differences were suitably assessed and documented. Regular error reports that contained additional errors discovered were sent to

[1]For background and documentation on Survey Solutions, please visit https://mysurvey.solutions/. The software platform is available free of charge and is being developed by the World Bank Development Data Group - Survey Unit. To access Survey Solutions Designer, please visit and sign up as a user at https://designer.mysurvey.solutions/.

[2]For information on how the system works. Visit https://github.com/arthur-shaw/uhis_auto_sort/

the teams. These errors were later corrected after supervisors instructed follow-up visits to the house. The data that had gone through this first cleaning stage was then approved by the data manager. Headquarters evaluates the interview after the Data manager accepts it on the Survey Solutions server and can decide whether to accept or reject it depending on the findings. In addition to using the Survey Solutions platform to review the data, error checking systems in STATA as well as R were also used at this stage.

The final raw data was carefully examined in the third stage following the first and second cleaning phases. Outliers, responses that were out of range, and consistency with other sections and factors were examined for each variable independently. However, great caution was taken to avoid making definitive judgements while correcting any errors. Despite the thorough data cleaning procedure, there may still be a few minor errors in the data that the data cleaning team was unable to identify and/or correct. Should analysts encounter any remaining inconsistencies, the NSO welcomes any queries and is available to provide further clarification.

7.2 Description of Datasets

Table 7.1: Structure of the IHPS 2024 Household Database

File Name	Module Name	Level of Analysis	Identification Variable(s)
HH_MOD_A_FILT	Module A: Household Identification	Household	`y5_hhid`
HH_MOD_B	Module B: Household Roster	Individual	`y5_hhid, PID`
HH_MOD_C	Module C: Education	Individual	`y5_hhid, PID`
HH_MOD_D	Module D: Health	Individual	`y5_hhid, PID`
HH_MOD_E	Module E: Time Use & Labour	Individual	`y5_hhid, PID`
HH_MOD_F	Module F: Housing	Household	`y5_hhid`
HH_MOD_F1	Module F1: Land Roster	Garden	`y5_hhid, gardenid`
HH_MOD_G1	Module G: Food Consumption Over Past One Week	Consumption Item	`y5_hhid, hh_g02`
HH_MOD_G2	Module G: Food Consumption Over Past One Week	Food Group	`y5_hhid, hh_g08*`
HH_MOD_G3	Module G: Food Consumption Over Past One Week	Age Group	`y5_hhid, hh_g10a`
HH_MOD_H	Module H: Food Security	Household	`y5_hhid`
HH_MOD_I1	Module I: Non-Food Expenditures – Over Past One Week & One Month	Consumption Item	`y5_hhid, hh_i02`
HH_MOD_I2	Module I: Non-Food Expenditures – Over Past One Week & One Month	Consumption Item	`y5_hhid, hh_i05`
HH_MOD_J	Module J: Non-Food Expenditures – Over Past Three Months	Consumption Item	`y5_hhid, hh_j02`

Continued on next page...

Continued from previous page

File Name	Module Name	Level of Analysis	Identification Variable(s)
HH_MOD_K	Module K: Non-Food Expenditures – Over Past 12 Months	Consumption Item	y5_hhid, hh_k02
HH_MOD_L	Module L: Durable Goods	Durable Good	y5_hhid, hh_l02
HH_MOD_M	Module M: Farm Implements, Machinery, and Structures	Farm Implement	y5_hhid, hh_m0b
HH_MOD_N1	Module N: Household Enterprises	Household	y5_hhid
HH_MOD_N2	Module N: Household Enterprises	Household Enterprise	y5_hhid, hh_n09a
HH_MOD_O	Module O: Children Living Elsewhere	Child of Head/Spouse Living Elsewhere	y5_hhid, hh_o0a
HH_MOD_P	Module P: Other Income	Income Type	y5_hhid, hh_p0a
HH_MOD_Q	Module Q: Gifts Given Out	Gift Type	y5_hhid, hh_q0a
HH_MOD_R	Module R: Social Safety Nets	Program	y5_hhid, hh_r0a
HH_MOD_S1	Module S: Credit	Loan	y5_hhid, hh_s02
HH_MOD_S2	Module S: Credit	Household	y5_hhid
HH_MOD_T	Module T: Subjective Assessment of Well-Being	Household	y5_hhid
HH_MOD_U	Module U: Shocks & Coping Strategies	Shock	y5_hhid, hh_u0a
HH_MOD_V	Module V: Child Anthropometry	Individual	y5_hhid, PID
HH_MOD_W	Module W: Deaths In Household	Deceased Individual	y5_hhid, hh_w01
HH_MOD_X	Module X: Filter Questions For Agriculture & Fishery Questionnaires	Household	y5_hhid
HH_MOD_ZB	Module ZB: Savings	Individual	y5_hhid,PID
HH_MOD_ZC	Module ZC: Goal Setting	Individual	y5_hhid, PID
HH_MOD_ZD	Module ZC: Generalized Livelihood Efficacy	Individual	y5_hhid, PID
HH_MOD_Z	Module Z: Child Care	Individual	y5_hhid, PID

Table 7.2: Structure of the IHPS 2024 Agriculture Database

File Name	Module Name	Level of Analysis	Identification Variable(s)
AG_META	Agriculture Questionnaire Metadata (Contains time stamps and respondent IDs for each module)	Household	y5_hhid
AG_MOD_B2	Ag-Module B_2: Garden Details – [Rainy Season]	Garden	y5_hhid, gardenid
AG_MOD_C	Ag-Module C: Plot Roster – [Rainy Season]	Plot	y5_hhid, gardenid, plotid
AG_MOD_D	Ag-Module D: Plot Details – [Rainy Season]	Plot	y5_hhid, gardenid, plotid
AG_MOD_E1	Ag-Module E: Affordable Input Program – [Rainy Season]	Household Registration	y5_hhid
AG_MOD_E2	Ag-Module E: Affordable Input Program – [Rainy Season]	Individual-Input Type	y5_hhid, input_id
AG_MOD_F	Ag-Module F: Other Inputs – [Rainy Season]	Input Type	y5_hhid, ag_f0c
AG_MOD_G	Ag-Module G: Crops – [Rainy Season]	Plot-Crop	y5_hhid, gardenid, plotid, crop_code
AG_MOD_H	Ag-Module H: Seeds – [Rainy Season]	Seed Type	y5_hhid, crop_code
AG_MOD_I	Ag-Module I: Sales/Storage – [Rainy Season]	Crop	y5_hhid, crop_code
AG_MOD_I_1	Ag-Module I_1: Harvest Labour [Rainy Season]	Individual-Crop Type	y5_hhid, id_code, crop_code
AG_MOD_I2	Ag-Module I2: Garden Details – [Dry Season]	Garden	y5_hhid, gardenid
AG_MOD_J	Ag-Module J: Plot Roster – [Dry (Dimba) Season]	Plot	y5_hhid, gardenid, plotid
AG_MOD_K	Ag-Module K: Plot Details – [Dry (Dimba) Season]	Plot	y5_hhid, gardenid, plotid
AG_MOD_L	Ag-Module L: Other Inputs – [Dry (Dimba) Season]	Input Type	y5_hhid, ag_l0c

Continued on next page...

Continued from previous page

File Name	Module Name	Level of Analysis	Identification Variable(s)
AG_MOD_M	Ag-Module M: Crops – [Dry (Dimba) Season]	Plot-Crop	y5_hhid, gardenid, plotid, crop_code
AG_MOD_N	Ag-Module N: Seeds – [Dry (Dimba) Season]	Seed Type	y5_hhid, crop_code
AG_MOD_O	Ag-Module O: Sales/Storage – [Dry (Dimba) Season]	Crop	y5_hhid, crop_code
AG_MOD_O_1	Ag-Module I_1: Harvest Labour [Dimba Season]	Individual-Crop Type	y5_hhid, id_code, crop_code
AG_MOD_O2	Ag-Module O_1: Plot Roster Tree Crop Production	Plot	y5_hhid, gardenid
AG_MOD_P	Ag-Module P: Tree / Permanent Crop Production Last 12 Months	Plot-Tree Crop	y5_hhid, gardenid, plotid, crop_code
AG_MOD_Q	Ag-Module Q: Tree/Permanent Crop Sales/Storage Last 12 Months	Tree Crop	y5_hhid, crop_code
AG_MOD_Q_1	Ag-Module Q_1: Harvest Labour [Tree/Perm Season]	Individual-Crop Type	y5_hhid, id_code, crop_code
AG_MOD_R1	Ag-Module R: Livestock	AnimalType	y5_hhid, ag_r0a
AG_MOD_R2	Ag-Module R: Livestock	Household	y5_hhid
AG_MOD_S	Ag-Module S: Livestock Products	By-product	y5_hhid, ag_s0a
AG_MOD_T1	Ag-Module T: Access To Extension Services	Extension Source	y5_hhid, ag_t0a
AG_MOD_T2	Ag-Module T: Access To Extension Services	Extension Source	y5_hhid, ag_t0b

Table 7.3: Structure of the IHPS 2024 Community Database

File Name	Module Name	Level of Analysis	Identification Variable(s)
COM_CA	Module CA: Community Identification	Community	ea_id
COM_CB	Module CB: Roster Of Informants	Informant	ea_id, com_cb01
COM_CC	Module CC: Basic Information	Community	ea_id
COM_CD	Module CD: Access To Basic Services	Community	ea_id

Continued on next page...

Continued from previous page

File Name	Module Name	Level of Analysis	Identification Variable(s)
COM_CE	Module CE: Economic Activities	Community	ea_id
COM_CF	Module CF: Agriculture	Community	ea_id
COM_CG	Module CG: Changes	Community	ea_id
COM_CG1	Module CG: Changes	Community	ea_id
COM_CG2	Module CG: Changes	Event	ea_id, com_cg35a
COM_CH	Module CH: Community Needs, Actions & Achievements	Need	ea_id, com_ch0b
COM_CI	Module CI: Communal Resource Management	Natural Resource	ea_id, com_ci0b
COM_CJ	Module CJ: Communal Organization	Communal Group Type	ea_id, com_cj0b
COM_CK	Section CK: Prices	Item	ea_id, com_ck00a

Table 7.4: Fisheries Module

File Name	Module Name	Level of Analysis
FS_MOD_B	Module B: Filter Questions and Household Labour for fisheries	Household
FS_MOD_C	Module C: Open Catch Fishing	Fish Species
FS_MOD_D	Module D: Acquaculture	Fish Species

7.2.1 Confidential information

Please take note that all names and addresses have been eliminated from the datasets in order to preserve the data's anonymity. Furthermore, the GPS coordinates have been eliminated since they might be used to precisely pinpoint households.

Chapter 8: Using the Data

It is strongly recommended that the end user of the IHPS data familiarize themselves with the questionnaires and manuals while using the IHPS data. The naming of IHPS data files follows the instrument name and module lettering as listed in the questionnaires and variable names, whenever possible, reflect question numbers as presented in relative modules. In the STATA versions of the data, variable labels, whenever possible, perfectly match the question asked in the questionnaires. In some cases it was necessary to modify the variable labels and cross-referencing the questionnaires will be necessary for accurate use of the data.

To increase the efficiency with which the survey instruments were administered, the IHPS instruments make extensive use of skip patterns. End users of the IHPS data must be aware of these skip patterns to properly interpret the data. When referencing the available paper questionnaires note that skip patterns are, in most cases, clearly identified by an arrow followed by a number in parentheses (» 2). The skip codes are explained in detail in the Enumerator Manual.

8.1 File Structure

The file structure of the IHPS data directly reflects the modules in the questionnaires. Where modules in the questionnaire contain data with multiple levels of observation, data files have been divided with additional numeric labels. It is recommended that end users of the IHPS data refer to the questionnaires and manuals when using the data. The index of data files, along with key identifiers relevant for merging data from different modules, are presented in Tables 7.1-7.4.

IHPS data files follow an intuitive naming scheme for easy use by the end user. Each file name gives reference to the instrument component, "HH" (Household), "AG" (Agriculture), "FS" (Fishery) and "COM" (Community) and the specific module as they appear in the questionnaires. For example, file "HH_MOD_B" refers to Household Module B; Household Roster. Similarly, file "AG_MOD_Q", for example, refers to Agriculture Module Q; "Tree / Permanent Crop Production (Over the Last 12 Months)". In modules that contain sub-sections with varying levels of observation, a number has been added to the tail of the file name, "HH_MOD_G1" and "HH_MOD_G2" for example. The numbers are sequential with how the module appears in the questionnaire.

8.2 Linking IHPS Databases Across Rounds

Only households from the 51 "retained EAs" linked across rounds. The information below therefore pertains to those households. The IHPS data include the variables **case_id** and **HHID** as baseline household identifiers, since each of the 1,948 IHPS 2024 households can be mapped to the 2010, 2013, 2016 and 2019 household counterparts.

The variable **y5_hhid** is the unique household identifier in the IHPS 2024 data, and it is composed of a 4-digit renumbered 2019 value plus the lowest IHPS 2019 two-digit roster ID code (identified by the variable hh_b06_1 in the HH_MOD_B of the IHPS 2024 database) for the baseline sample members that were found in that household in 2024.

At the **individual-level**, the IHPS data from 2010, 2013, 2016, 2019 and 2024 can be merged using the variable **PID** (without using any other variable for individual level merges across time). **PID** is a unique individual identifier that is assigned to a given individual the first time he/she joined the panel sample, whether in 2010, 2013, 2016 or 2019. Given the attrition at the household and individual levels, the merges across the rounds of the IHPS data will not be perfect.

A special scenario encountered in the IHPS involved individuals moving from one IHPS baseline household to another. These individuals are identified in ta **individualmover** in HH_MOD_B. The baseline household these individuals are listed in the variable **moverbasehh**.

To replicate the attrition statistics reported in the BID and report, the users should consult the ancillary data file "**IHPS2024MemberDatabase**" that has been made available. The file contains all 7,099 IHPS 2019 sample individuals; the variable **eligible_tracking** that identifies those that were tracking eligible in 2024 in terms of their projected age in 2024 and their relationship to household head in 2019 in accordance with the tracking protocol; the variable status that identifies their final tracking outcome in 2024; the variable **whynotfound** that identifies the reason for being unable to interview a given tracking-eligible individual in 2024 (cases that migrated outside of Malawi, moved to an institution such as a police compound or army barracks, and other special cases); and the variable **specialcase** that details the unique reason for not being able to interview IHPS 2024 individuals that are marked as "special case" for the variable **whynotfound**.

8.3 Crop Unit Measures

Participants were free to report in any unit they were most comfortable with while gathering data on agricultural quantities (such as the amount of crop harvested, etc.). Frequently, instead of using normal measurements like kilos, litres, etc., respondents gave quantities in non-standard units like "heap," "chigoba," or "basket." For the examination of production estimations, a database of conversion factors is accessible.

8.3.1 Unit Conversion Factors

Quantities in several units must be translated into a common unit using conversion factors in order to be used and compared. The NSO also administers a nation-wide Market survey as part of the IHS program, which gathers conversion factors for the Non-Standard Units.

To change the quantity to kilos or litres, the user only needs to multiply the amount by the conversion.

Though there are still some gaps where conversion factors are not accessible, the conversion factors provided in the data files cover the bulk of item/crop-unit combinations seen in the data set. These gaps are being filled, and new conversion factors will be published as soon as they are available.

Chapter 9: Appendices

Appendix 1: How to Obtain Copies of the Data

The data are available through the NSO web site: https://www.nsomalawi.mw/ or through the LSMS-ISA website: https://microdata.worldbank.org/

Users do not need to obtain the permission of the NSO to receive a copy of the data but will be asked to fill in a data access agreement. In this agreement, users agree to: (a) cite the National Statistical Office as the collector of the data in all reports, publications and presentations; (b) provide copies of all reports, publications and presentation to the National Statistical Office (see address below); and (c) not pass the data to any third parties for any reasons.

The Commissioner of Statistics,
National Statistical Office
Off Chimbiya Road
P. O. Box 333, Zomba, Malawi
Email: commissioner@nso.gov.mw