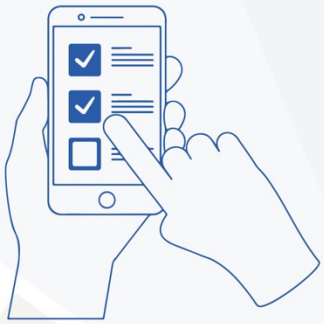
REPUBLIC OF RWANDA

RWANDA POPULATION AND HOUSING CENSUS
16-30 AUGUST 2022
BE COUNTED BECAUSE YOU COUNT

NISR
NATIONAL INSTITUTE OF STATISTICS OF RWANDA



5th

POPULATION AND HOUSING CENSUS

Rwanda, 2022

Thematic Report

POST ENUMERATION SURVEY REPORT

"Be counted because you count - Ibaruze kuko uri uw'agaciro"

Ministry of Finance and Economic Planning

National Institute of Statistics of Rwanda

Fifth Rwanda Population and Housing Census, 2022

Thematic Report

POST ENUMERATION SURVEY REPORT
July 2023

EXECUTIVE SUMMARY

The National Institute of Statistics of Rwanda (NISR) conducted the Post Enumeration Survey (PES2022) from 10th to 30th September 2022 whose objective was to evaluate the coverage and content errors of the 2022 Population and Housing Census (RPHC2022) data that were conducted in August 2022. The use of modern techniques and tools based on Data Science drastically shortened the timeline for obtaining matching results, three weeks compared to six months in PES2012 that used traditional methods for matching process.

The planning and implementation of PES2022 were undertaken by staff from both the department of Statistical Methods, Research and Publication (SMRP) and the department of Data Revolution and Big Data (DR&BD) who were not involved in the RPHC2022 to guarantee the independence of the PES2022 vis a vis the Census.

To achieve PES2022 objectives, 180 EAs were selected countrywide (6 EAs per district) from a universe of 24,339 EAs countrywide, which corresponds to a sampling fraction of 0.74%. A brief and short questionnaire was developed with two sections namely a section about demographic characteristics of household members and another one about household members' addresses (No-movers, out-movers and in-movers) with reference to Census night.

For PES2022 smooth implementation, 210 enumerators were recruited and trained in navigating through allocated EAs using digital maps and GPS in addition to other modern data collection techniques. For the first time, PES listing and data collection relied on technology where data collection was conducted using Computer Assisted Personal Interview (CAPI), field monitoring through a dashboard and matching using data science techniques and tools.

The field work at district level was supervised by a Team Leader whose responsibilities were to ensure the completeness and highest quality of collected information among others. The daily monitoring of fieldwork (listing and data collection) was performed by a team of 10 Monitors based at NISR premises supported by real time dashboard.

To conduct the PES2022 analysis, a computer-based algorithm was developed using the Python programming language to facilitate the matching process. This algorithm significantly reduced the time required to produce results compared to the PES2012 analysis. Specifically, the PES2022 matching results were generated within three

weeks, whereas similar operations in 2012 took six months to complete. This highlights the advancements in computational tools and methodologies over the past decade.

Field visits and telephone calls were done during reconciliation to find additional matching records and understanding reasons behind the unmatched data. a comparison of Census and PES records was done at individual level. We call this process matching and it was done basing on variables that are most likely to facilitate an optimal identification of people in both RPHC2022 and PES2022 datasets. These are: household identification (HHID), names, age, sex, marital status and relationship to the head of household.

Matched and unmatched records were analyzed to estimate the true population size of Rwanda and coverage rates using a statistical technique called "Dual-System Estimation (DSE)." Furthermore, the analysis of matched records helped to estimate content errors of the census records.

PES Key Findings

A. Estimation of Coverage rates

The 2022 Post Enumeration Survey finds that:

- The national level Net Coverage Error Rate is 1.3%. This indicates that, on average, there is a 1.3% discrepancy between the actual data and the reported data at the national level. The highest NCER is observed in the City of Kigali with 3.14%, while the lowest is observed in Eastern Province with 0.88%. Urban areas, have a higher NCER of 2.26% (above the national average), compared to 0.97% in rural areas. Among different age groups, the group of 15 to 29 years has the highest NCER of 2.14%, whereas those aged between 45 to 59 have the smallest NCER at 0.87.
- The Census Omission Rate (COR) is 1.51% at national level. The highest COR is observed in Urban areas with 2.45%, and lower in rural areas (1.17%).
- The difference of COR between Males and Females is 0.58 percentage points with 1.80% for males and 1.22 % for females. Among age groups, the highest COR is observed in the group of 15 to 29 years with 2.34%, while the lowest is observed in the group of 45 to 59 years with 1.12. By Provinces, the highest COR is observed in the City of Kigali with 3.45% while the lowest is observed in Northern Province with 1.09%.

- The Rate of Erroneous Inclusion (REI) at national level is 0.19 %. The highest REI is observed in the City of Kigali with a rate of 0.31%, while the lowest is in Northern Province with 0.07%. The age group of 30 to 44 years has the smallest REI at 0.08% whereas the highest REI is observed in the group of 60 years and above, with a rate of 0.27%. The difference of REI between males and females is 0.09 percentage points, while similar in Urban and Rural areas.

B. Estimation of Content Errors

- The highest NDR in the category of Marital Status is observed for the characteristic "Married to one wife/husband officially" with a value of 1.03%. The lowest NDR is in the category of Relationship to the Head of Household, specifically for the characteristic "Father-in-law/Mother in law," and also in the category of Health Insurance, particularly for the characteristic "NGOs," with a NDR of 0.00%.
- Rate of Agreements: The rate of agreements is more than 90% in all characteristics. If a post-enumeration survey reveals that more than 90% of respondents agree on specific characteristics like Relationships to the Head of Households, Sex, Marital Status, Age groups, or any other variables of interest, it indicates that the data collected in the census was accurate and reliable.
- Index of inconsistency: The highest index of inconsistency is observed among characteristics of Relationship to the head of households and Health Insurance. For Relationship to HH head and Health Insurance, the IOI is highest for "Unknown relationship and Employer with a value of 100. This suggests significant inconsistency or variability in the data for individuals whose characteristics are "Unknown relationship" and "employer" provided relationship to the head of households and Health Insurance. The lowest IOI is in the characteristics of age groups: Specifically, the IOI for the "60+ years" age group is 3.17. This is much lower than the IOI values for the other characteristics mentioned, indicating that there is less inconsistency or variability in the data for this age group.
- Aggregate Index of Inconsistency: The highest Aggregate Index of Inconsistency is observed in characteristics of Health Insurance with a value of 33.01%. The lowest Aggregate Index of Inconsistency is observed in characteristics of Sex and Age groups with 3.34% and 4.20% respectively.
- Gross Difference Rate (GDR): the highest GDR is observed in characteristics of Marital Status with 10.5%, while the lowest GDR is in characteristics of Sex and Age groups with 1.67% and 3.35% respectively.

Contents

List of Figures

List of Tables

LIST OF ABBREVIATIONS

AII : Aggregate Index of Inconsistency

CAPI : Computer Assisted Personal Interview

CI : Confidence Interval

COR : Census Omission Rate

DRBD : Data Revolution and Big Data

DSE : Dual System Estimation

EA : Enumeration Area

EU : European Union

GDR : Gross Difference Rate

GDR : Gross Difference Rate

GIS : Geographic Information System

HHID : Household Identification

ICT : Information and Communication Technology

ID : Identification Code

NCER : Net Coverage Error Rate

NDR : Net Difference Rate

NISR : National Institute of Statistics of Rwanda

PES : Post Enumeration Survey

RA : Rate of Agreement

RPHC : Rwanda Population and Housing Census

SMRP : Statistical Methods, Research and Publication

Foreword

The Government of Rwanda, through the National Institute of Statistics of Rwanda (NISR) conducted the 2022 Rwanda Population and Housing Census to provide updated demographic, social and economic indicators for policy formulation and planning to support the national development agenda and track the implementation of national, regional, continental and global development goals, such as the First National Strategy for Transformation (NST1 2017-2024); the AU Agenda 2063; and the Sustainable Development Goals (SDGs).

The Population and Housing Census in Rwanda dates back to the year 1978 when the first ever modern census was implemented. The second, third, and fourth censuses were carried out in 1991, 2002, and 2012 respectively. The 2022 Population and Housing Census marks therefore the Fifth in the series following the United Nations Recommendations to conduct a census every ten years.

Considering census' crucial importance for the planning process, the Government of Rwanda has made the Population and Housing Census a priority to be undertaken every 10 years and adopted the use of technologies for timely delivery of census results for use.

Results of the 2022 Population and Housing Census provide population numbers from national to the lowest administrative level, as well as demographic and socio-economic indicators at both national and district levels. The census remains the only national data collection exercise that can provide the lowest levels of disaggregation to support decentralised decision making across the country.

I would like therefore to take this opportunity to thank all stakeholders that contributed to the success of the 2022 Population and Housing Census. They include Ministries and other Government institutions, international organizations such as the World Bank (WB), the European Union (EU), the United Nations Population Fund (UNFPA), One-UN, UN Women, UNICEF, UNECA, the United Kingdom AID (UKAID), the African Development Bank (AfDB), the USAID, Enabel, PARIS 21 and others for their support in diverse ways.

My special thanks go to the local government leaders from the province to the village levels who contributed a lot to the success of the 2022 Population and Housing Census. Exceptional gratitude to all enumerators, teachers and young people who collected the information and all respondents for their cooperation and dedication. The National Institute of Statistics of Rwanda (NISR) deserves special appreciation for the excellent operational and coordination of all census activities.

I finally recommend that the invaluable information contained in different thematic reports of the 5th Rwanda Population and Housing Census be used as updated evidence for all decision and policy making for the national, regional and global development programs.

Dr. Uzziel NDAGIJIMANA
Minister of Finance and Economic Planning

Acknowledgements

The National Institute of Statistics of Rwanda (NISR) is pleased to release the results of the Fifth Population and Housing Census (RPHC5). The execution of different Census phases: preparatory works, data collection, data processing, tabulation and data analysis spans a period of about four years between 2020 and 2023.

NISR is pleased to publish the main indicators report as the main flagship report of the RPHC5. This report will be followed by several thematic reports and districts profile reports. The main indicators report covers several issues mainly population size and distribution, education, settlement, population projections to mention but a few. NISR hopes that the results in this report supplemented by the upcoming thematic reports would meet the demand of census data users across the board.

On this occasion, I would like to extend my sincere gratitude to the Government of Rwanda and development partners for availing financial, logistical and technical support. The NISR would like to appreciate all stakeholders who worked tirelessly with us to ensure that the 2022 Rwanda Population and Housing Census operation was successful.

Special recognition also goes to the Ministry of Finance and Economic Planning, Ministry of Defense, Ministry of Local Government, Ministry of Education, Ministry of Foreign Affairs, Ministry of ICT and Innovation, Ministry of Interior, Ministry of Health, Ministry in Charge of Emergency Management, the Rwanda National Police, Rwanda Correctional Services, Rwanda Biomedical Center (RBC), Rwanda Information Society Authority (RISA), Rwanda Utilities Regulatory Authority (RURA), Rwanda Public Procurement Authority (RPPA), Office of Government Spokesperson (OGS), and Rwanda Broadcasting Agency (RBA) for the direct involvement in awareness campaign, logistical and data collection operations.

I also wish to express my appreciation to the local government authorities and NISR staff for their excellent operational organization and to the tens of thousands of enumerators and supervisors for their painstaking efforts throughout the data collection phase.

Finally, the people of Rwanda, residents, and visitors, your cooperation was crucial towards the success of the census.

Thank you.

MURANGWA Yusuf
Director General,
National Institute of Statistics of Rwanda

Chapter 1 : Introduction

The National Institute of Statistics of Rwanda (NISR) conducted the Post Enumeration Survey (PES-2022) for the fourth time. The main objective of this survey was to collect high-quality information for the evaluation of the quality of data gathered during the Rwandan Population and Housing Census (RPHC2022). Post Enumeration Surveys are crucial in assessing the accuracy and completeness of the data collected in a census.

The specific objectives of the PES2022 were as follows:

- To evaluate the accuracy of RPHC2022 records by providing any reader with relevant information on coverage and content errors on selected key census variables at different levels of estimation including national, provincial, urban and rural;
- To identify implementation issues which may require improvement in future population and Housing censuses as well as other surveys;

Even though censuses are not expected to include sampling errors as sample surveys do, there are risks of various types of non-sampling errors that may arise during census operations. They are generally categorized into coverage and content errors. Therefore, evaluation of the magnitude and direction of these errors is necessary to respond to questions about the quality of census results. In addition, such evaluations provide users with the basis for deciding either the errors are relatively small and not likely to affect conclusions to be drawn from the census data or that the errors are relatively large and inferences should be made with caution. The importance of PES results calls therefore to be cautious in its planning, implementation and analysis and in an independent context vis a vis the census.

In this regard, NISR established a dedicated team composed of staff from the department of SMRP and staff from the department of Big Data and Data Revolution to manage the PES project, from planning to the report writing. The team selected 180 Enumeration Areas (EAs) that were allocated in all districts: 6 EAs for each district.

The data collection went on in a smooth way and successfully completed on time despite some slight challenges that were identified and immediately managed. The use of Modern technologies for data collection and transmission as well as the field monitoring played a vital role in speeding up and achieving the quality of collected data.

The process of Matching PES records and Census as well as the data analysis was also very important phase of PES implementation. It required robust tools and well trained personnel. It is in this context that NISR established a team composed of data scientists to prepare the ground and set up an adequate environment for this process. A dedicated algorithm for matching was developed through Python programming language in collaboration with the UK office of National Statistics that also trained the NISR staff in techniques of automatic and manual matching.

Using the Dual System Estimation (DSE) method, the true population was estimated as well as other coverage rates. Content errors were also computed, basing on the matched records of PES and Census.

Therefore, this report presents the salient features of PES2022, including its scope, methodology, the information gathered and discuss the results of data analysis including estimates of coverage rates and content errors rates. It also present lessons learnt and provides recommendations for the future censuses.

Chapter 2 : Survey methodology

Chapter two of this report provides the reader with details about the PES2022 methodology including Choice of PES procedure, Questionnaire designing, Training Manual Preparation, Sample design, **Recruitment and training of Enumerators, PES listing and data collection,** Weighting and estimation methods, Estimation of coverage rates, Estimation of content errors, Matching of PES and Census records.

2.1. Choosing the PES procedure

There are three procedures for evaluating census coverage in a PES (Design and Implementation of a Post Enumeration Survey: Developing Country Example, 1993). The three procedures are known as A, B and C respectively.

Procedure A: This procedure attempts to construct the households as they existed at the time of the Census. The respondent must identify all persons in the PES household as of the Census reference night. Based on this information, the number and percentage of non-movers and movers (out-movers) are estimated. In procedure A, the matching of non-movers and out-movers is relatively simple and inexpensive because the search is limited to the sample areas and their adjacent areas. The weakness of procedure A lies in the fact that it is very difficult and expensive to identify out-movers, especially out-mover households, given that they are no longer at the sample address and the information is reported by proxy respondents. Therefore, there is strong possibility of underestimation of the number of out-movers, and since movers are more likely to be missed by the Census this leads to underestimation of the Census omissions.

Procedure B: This procedure tries to identify all persons currently in the PES household at the time of PES interview date. The respondent must identify all persons in the PES household at the time of PES interview and he/she is asked to provide the addresses where the persons were living on census night. The persons would later be matched to the corresponding census records. Based on this information, the number and percentage of matched non-movers and movers (in-movers) would be easily estimated. But even though procedure B provides better estimates of the number of non-movers and in-movers than procedure A, it is not providing the number of out movers. This is complicated by the fact that addresses in developing countries are often inadequate. One problem with procedure B is that one is not always sure whether a failure to match indicates an actually omitted persons or an incorrectly located persons.

Procedure C is a combination of procedure A and B. The goal of this procedure is to identify all current members of the household, as of the PES date. In addition, any other resident as of the census reference night. At the end, these persons were classified as non-mover, out-mover or in-mover with regard to household membership status as of the Census night.
Therefore, only residents as of the Census night (non-movers) and the persons who left the household (out-movers) were matched to Census records. The matching and estimates of the numbers of non-movers and in-movers are based on procedure B. The matching and estimates of the numbers of out-movers were based on procedure A. In sum procedure C is a combination of procedures A and B. Consequently, it is the procedure C which has been followed in the current PES Survey. It is this procedure C that guided the design of PES2022 questionnaire and developing the matching procedures and techniques.

2.2. Questionnaire design

The PES questionnaire was designed in conformity to procedure C of coverage analysis. It is also consistent with the Dejure enumeration basis of the 2022 Population and Housing Census. Its questions allow to capture information that is needed to estimate non-movers, in-movers, out-movers, and their demographic characteristics such as location, age, sex, relationship to the head of household and marital status. It also contains a question about the use of Health insurance that was used to estimate some content errors.

The PES2022 questionnaire was designed to be compatible with the recommendation of the UN Statistical Division[1]. Concerning content analysis, the questionnaire comprises several census data items that are compared with collected Census data in order to measure the extent of variability between PES responses and the corresponding census responses. These data items include sex, age, Relationship to Head of household, etc. It is worth noting that definitions and categories that were used to design the PES2022 questionnaire are identical to that applied in the RPHC2022 questionnaire.

Unlike the similar surveys conducted in previous decades that used paper based questionnaires, the PES2022 data collection was done through a Computer Assisted Personal Interview (CAPI) on smartphones.

[1] Post Enumeration Surveys, Operational guidelines- 2022-census

2.3. Training Manual Preparation

The field interviewer/supervisor manual was prepared as a guide to all field workers. It was developed as a reference document that explains in detail the questionnaire and the listing procedures to equip the field-workers with comprehensive knowledge on PES concepts and definitions. The roles and functions of enumerators and supervisors at all levels have been specified in detail. The manual includes in addition definitions of relevant concepts and data items to be collected, a description of the operational processes that must be followed in listing roads, housing units and households have been included. The method of defining the EA boundaries on the ground as well as its compatibility with the pre-prepared map has also been explained. The census authority has reviewed the manual with regard to the definitions of census data items that are included in the PES questionnaire.

2.4. Sample Design

To ensure that the PES results are representative as much as possible, a stratified Single-stage sample design was used. The sampling frame was based on the cartographic data from the census mapping done in 2020 with a total of 24,339 enumeration areas, where 180 EAs were drawn for the PES stratified by district, sector and urban/rural. Within each district the sample of EAs was selected systematically with an estimated 6 EAs per district. The population in institutions, floating and homeless populations was not included in this evaluation.

a) **Sampling frame**

One of the preparatory activities for a population and housing census is to prepare and update maps. The drawing of EAs is partly justified by the subsequent use of maps for other purposes and particularly as frame for inter-censal sample surveys. The frame has to have an up to date listing of EAs and addresses. The listings are needed, among other reasons, for the coding of location names. It is in this context that, the NISR conducted the mapping operation in 2020 whose 24,339 EAs which was used as the basis of the population and Housing Census, Post Enumeration Survey-2022 as well as a sampling frame to be used during the inter-sensual surveys. Below is the sample size and distribution used in PES-2022 derived from the 2020 population mapping of EAs.

b) **Sample size and allocation**

The literature reviews of previous Post Enumeration Surveys carried out in Rwanda (1991, 2002 and 2012) revealed that the adopted sample size was 120 EAs for three previous surveys. As such, it was deemed appropriate and logical to maintain this

sample size for the present PES-2022. However, it was not practical due to significant changes in village boundaries, urban/rural definitions between Census-2012 and Census-2022. Additionally, while previous PES reports provided estimates at the national level, the current PES report offers estimates at the province level.

It is in this regard that we decided to use the current Census mapping (2020 mapping) which has updates of the current EAs. Nonetheless, the sample size was independently calculated based on a finite population of 2020-Census mapping of EAs and the assumed margin of error of 0.0742.

Using the above sample estimates, the national level sample size for the current PES-2022 is 180 EAs, where the sample size at the domain level (District) was estimated to be 6 EAs per district. The reason of choosing 180 sample size for PES-2022 compared to 120 for the previous Censuses is to minimize the high risk of anticipated errors due to changes of EAs. The above stated sample size was illustrated in Yamane's formula:

$$n = \frac{N}{1 + N(d^2)},$$

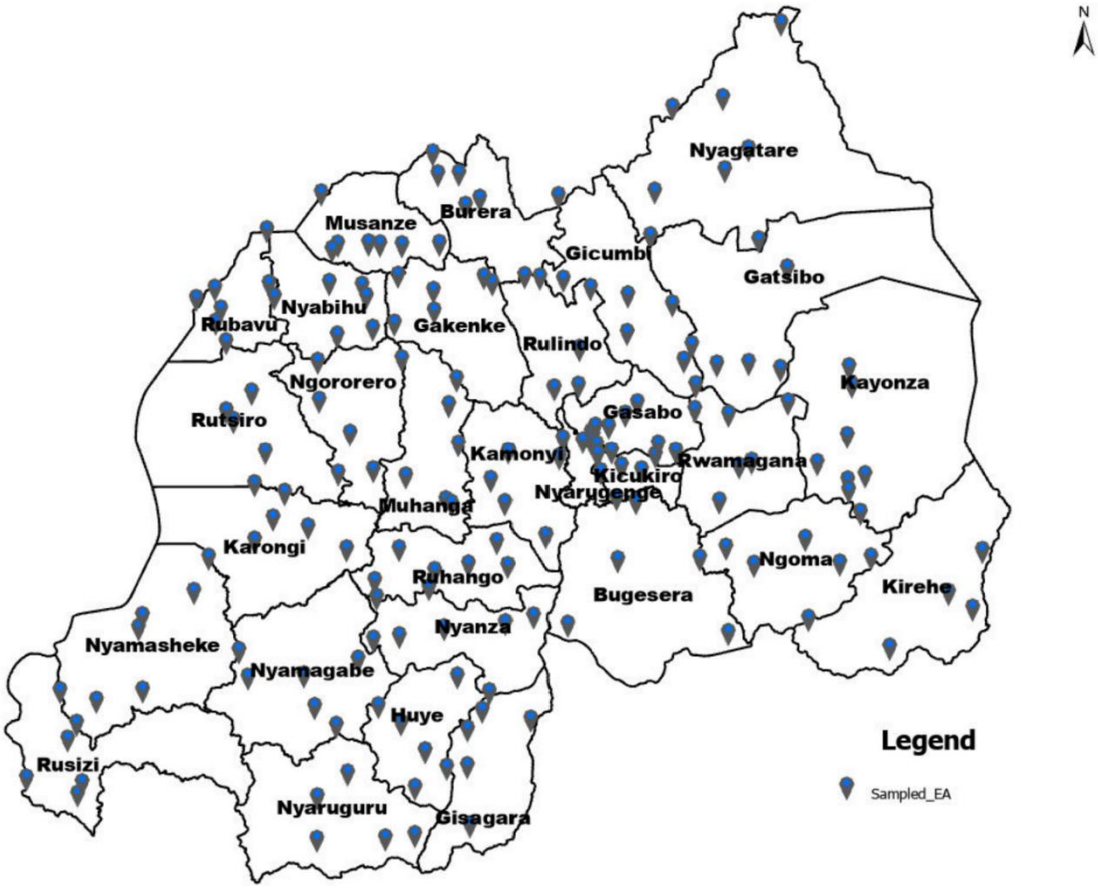
d is the margin of error (0.0742) and
$N = 24{,}339\ (EAs\ population\ for\ RPHC2020)$

Therefore,

$$n = \frac{24{,}339}{1 + 24{,}339 * (0.0742)^2} = 180$$

Thus, the sample size used is a single stage stratified sampling. In each sampled EA, the whole universe within a cluster (all HHs & HH members) was enumerated and used to estimate the national level population (Census population). To estimate census population, the weighting procedure included probability proportional to size criteria since the whole universe from each sampled EA differs in size. This was done using STATA. The map below shows distribution of sampled Enumeration Areas for PES-2022 from a sampling frame of 24,339 EAs.

Figure 2.1: Distribution of sampled EAs for PES-2022



2.4.1. Accuracy of the sample size used in PES-2022

According to experience from 2012 Rwanda Post Enumeration Survey, the estimates of census coverage computed from the sample size of 120 EAs whose sampling fraction was 0.72 percent were high (i.e; the estimated population was 11.2 million, and the Net Coverage Rate was 99.25%,).

In addition, with the same sample size, the estimated coverage errors were also minor (i.e; Gross Coverage Error Rate: 1.92%, and Census Omission Rate: 1.33%). Furthermore, Kosovo PES in 2011 and South Africa PES in 2012 whose sampling fractions were 0.50% and 0.67% respectively, also produced reliable estimates.

Considering the fact that the sample size of Rwanda PES-2022 (180EAs) whose sampling fraction is 0.74% that is greater than 0.72% for Rwanda PES-2012, 0.67% for South Africa PES-2012, and 0.50% for Kosovo PES-2011, the estimates of PES2022 for the 5th RPHC are also anticipated to achieve the acceptable precision levels. The computed margin of errors and estimation of confidence intervals for the above key coverage rates examine it.

Estimation of margin of errors and confidence intervals for PES-2022 key coverage rates.

A margin of error is a statistical measurement that accounts for the difference between actual and projected results in a random survey sample. In simpler terms, the margin of error allows you to gauge the level of unpredictability in data and research outcomes.

The results from Post Enumeration Survey 2012 indicate that the Net coverage rate, Gross-coverage error rate, and Census omission rate were 99.25 percent, 1.92 percent, and 1.33 percent respectively. Extrapolating the above information to the actual PES-2022, we would have the following forecasts at 95% level of confidence.

Margin of error for Net coverage rate (99.25) percent

$$d = z * \sqrt{\frac{p(1-p)}{n}} = z * \sqrt{\frac{pq}{n}}$$

Where "d" is the margin of error $(\pm \text{within value})$ used to estimate the confidence interval, "z"=Coefficient of normal distribution at 95% level of confidence (1.96), "p" is the percentage of coverage rate or content error , and "n" is the sample size.

$$d = 1.96 * \sqrt{\frac{0.99 * 0.01}{180}} = 0.0145358$$

Margin of error for Gross coverage error rate (1.92) percent

$$d = 1.96 * \sqrt{\frac{0.019 * 0.981}{180}} = 0.0199449$$

Margin of error for Census omission rate (1.33) percent

$$d = 1.96 * \sqrt{\frac{0.0133 * 0.9867}{180}} = 0.016735$$

Therefore, at 95% level of confidence, the Net Coverage Rate, Gross Coverage Error Rate, and Census Omission Rate, are forecasted to be between $\pm$:
$(98.00 \leq \text{ncr} \leq 100.00);\ (0.09 \leq \text{gcer} \leq 3.89);\ \text{and}\ (0.34 \leq \text{cor} \leq 3.00)$ Percent respectively.

Table 2-1: Stratum distribution of the sample and Universe by urban and rural

Stratum	Sample	Universe
Urban	52	7,944

Rural	128	16,395
Total	180	24,339

Source: Census Mapping Operation, 2020

2.5. Recruitment and training of Enumerators

About 210 persons were recruited for the PES fieldwork. A six-day training program was administered to all potential field personnel where NISR staff acted as trainers. In each end of the training session, a test was given to the trainees to imbed them with comprehensive knowledge on PES data collection techniques and principles. In addition, a field test of the planned data collection approach was conducted in 24 randomly selected EAs. The field test was held in areas that were not a subset of PES areas in order to avoid repeated interviews in the same PES EAs. Face-to-face interviews on the doorstep with 200 households were conducted.

2.6. PES Listing and Data collection.

The field-work started with the listing operation that took place in 180 Enumeration Areas for the period of four days, from 10th to 14th September 2022. It was followed by enumeration activities that were successfully finalized within a period of 15 days as planned from 16th to 30th September 2022. To make it successful, NISR has undertaken the PES Publicity as part of the overall Census publicity through mass media, social media, meetings and conferences, and through different advocacy and awareness materials such as banners, brochures and posters among others.

PES listing and enumeration operations were carried out in the way that make them independent from census operations. In that context, the following measures were taken:

- Ensure the listing of all Households in EA is done in a different way as it was done in census listing;
- Enumerators in PES must have not worked in Census listing and/or Enumeration operations in the census;
- Ensure that census enumerators are not influencing the PES listing and enumeration operations. To achieve it, PES enumerators were always reminded that confidentiality of the PES records is a must.

To mitigate any risks related to long absence of enumerator, it was planned one staff that can intervene in the case of any enumerator has any reason to abandon the PES listing or enumeration. In the same way, to mitigate any risk related to the

smartphone problems, at least three smartphones were ready to replace anyone with any problems making it not properly used.
To ensure the completeness and good quality of completed questionnaires as well as coverage of HHs in all EAs, a team of 5 monitors, 6 zone supervisors and Data Analysts were monitoring the data transmissions and collaborated with National Coordinators. The PES-2022 face-to-face interviews used a mobile phone as the main mode of data collection.

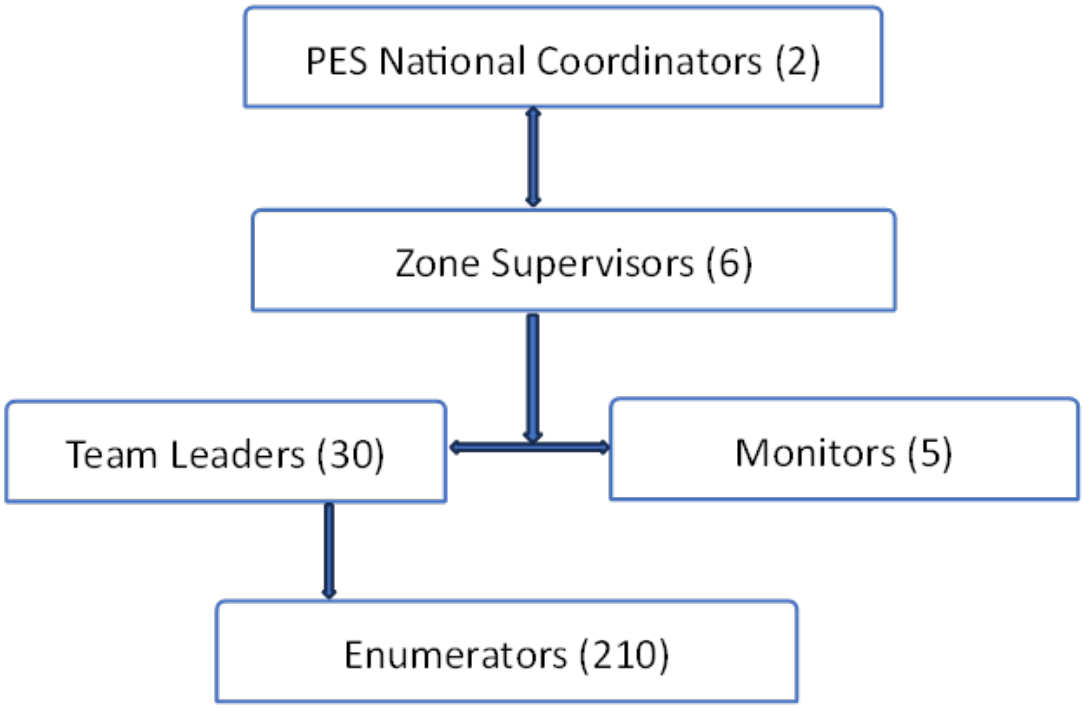
Maintaining daily communication with field staff was essential for ensuring the smooth operation of data collection efforts. The involvement of a PES coordination team to intervene in cases of troubleshooting or providing additional advice demonstrated a proactive approach to addressing challenges and ensuring that the data collection process stays on track. This coordination helped in maintaining consistency, resolving any issues promptly, and ensuring that the overall objectives of the census are met efficiently.

The face to face interview method was used in data collection. Enumerators canvassed all selected enumeration areas using GIS maps that have been also used for the census. Maps were showing all existing buildings, which were numbered. Interviewers and Supervisors had instruction manuals and report books, in which they were daily reporting on the progress of the survey.

Interviewers were organized in teams; District Team Leaders who were based on the field facilitating enumerators on day-to-day data collection endeavors, Data quality Monitors who were based at NISR premises checking on the daily basis the quality of data from the field using an interactive and in real time dashboard and reporting the challenges encountered for quick intervention, Supervisors who were monitoring data quality issues in their respective Zones, and PES Coordinators who were leading all PES endeavors at National Level.

The hierarchical structure of the field work organization is exhibited in the following diagram.

Figure 2.2: Filed work organization structure



2.7. Weighting and estimation methods

To obtain unbiased estimates from the PES data, it is necessary to apply appropriate weights to the sample data based on the probabilities of selection. It is also important to calculate measures of sampling variability for Census coverage and content estimates. The procedures for calculating the weights and variances are specified in this section. In order to avoid producing biased sample estimates, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household member is equal to the inverse of his/her probability of selection. As indicated before, since all households and household members are included in the PES Sample with certainty, the selection probability of a certain EA is exactly equivalent to the selection probability of a certain household and a household member within this EA. The selection probability of a certain EA is:

$$P_{\alpha_h} = \frac{\lambda_h M_{h_\alpha}}{\sum_{k=1}^{\alpha_h} M_{h_\alpha}}$$

Where,

P_{h_α} is the Probability of selection of the α^{th} EA in the sample of stratum h

M_{h_α} is the number of households in the α^{th} EA in the sample of stratum h

λ_h is the Number of EAs to be selected from stratum h

The estimates of Census coverage rate are calculated as ratios of two total estimates. Thus,

If we consider $w_h = \frac{1}{P_h}$ the district weight, and

If we let $\hat{R}$ denotes the estimate of Net Coverage Rate,

Then the ratio estimate $\hat{R}$ is defined as:

$\hat{R} = \frac{\hat{Y}}{\hat{X}}$, where

$\hat{Y} = \sum_{h=1}^{30} w_h (\sum_{i=0}^{6} y_{h_i})$ is estimated census population, where y_{h_i} is the number of population in the i[th] EA age-sex group of stratum h, and

$\hat{X} = \sum_{h=1}^{30} w_h (\sum_{i=0}^{6} x_{h_i})$ estimated True Population, where x_{h_i} is the True estimated Population in the i[th] EA age-sex group of stratum h.

2.8. Estimation of Coverage rates

The Following are key formulae that were used in the PES2022

1. **Matching Rate:** Is the matched population between the census and PES relative to PES population.

$$Matching\ rate = \frac{Matched\ pop}{PES_{pop}}$$

2. **Gross coverage error:** It is used as an indicator of the operational quality of the census enumerations.

$$Goss\ Coverage\ Error = Omissions\ +\ Erroneous\ inclusions$$

3. **Gross coverage error rate:**

$$Goss\ Coverage\ Error\ rate = \frac{Omissions\ +\ Erroneous\ inclusions}{CEN_{pop}}$$

4. **True Population:** This is the population estimated from the PES multiplied by the population from the census after correcting for erroneous inclusions and divided by matched population between the census and the PES.

$$TruePop = \frac{(PES_{pop})\ x\ (CEN_{pop} - ErronInclusion)}{Matched_{NonMovers} + Matched_{OutMovers}}$$

5. **Net coverage error:** universally known as the 'net omission rate' or the 'undercount' is the difference between what should have been counted (true population) and what was counted (census population).

$$Net\ Coverage\ Error = TruePop -\ CEN_{pop}$$

6. **Net coverage error rate** is the total net error relative to the dual-system estimate of the true population; that is, divided by TRUE_POP. This measure constitutes the single most important indicator of the quality of the census coverage.

$$Net\ coverage\ error\ rate = \frac{Net\ Coverage\ Error}{TruePop}$$

7. **Net Coverage rate** is the complement of the net coverage error rate.
   $Net\ coverage\ rate = 100\% - Net\ Coverage\ Error\ rate$

8. **Omissions:** This results from missing housing units, households, or persons during census enumeration. In the case of missing the whole housing unit, it implies that all households and persons living in that housing unit will also be missed during the census enumeration.

$$Omissions = True\ population - (Census\ population - Erroneous\ Inclusions)$$

2.9. Estimation of Content Errors

Content errors measure the discrepancy between records in the Census and the records in PES. These errors are only estimated for the matched persons and for some variables that are deemed to be more meaningful. These variables are mostly the ones that have been used in matching processes.

In the PES2022, the following Content Errors were computed and analyzed the national level for sex, age groups, marital status, and relationship to the Head of Households:

1) **Rate of Agreement:** The rate of agreement is the complement of the gross difference rate. A low rate of agreement indicates a high degree of variability, and vice-versa.

$$Rate\ of\ Agreement = \frac{\sum_i^c Y_{ii}}{n} \times 100$$

2) **Gross Difference Rate:** The gross difference rate (GDR) is calculated for the characteristic as a whole. It is the number of discrepancies between the census responses and the PES responses relative to the total number of persons matched.

$$GDR = \frac{(n - \sum_i^c Y_{ii})}{n} \times 100$$

3) **Index of Inconsistency (IoI):** The index of inconsistency is the relative number of cases for which the response varied between the census and the

PES. It is the ratio of the simple response variance to the total variance of the characteristic, including its variability in the population.

$$IoI = \frac{(y_{.i} - y_{i.} - 2y_{ii})}{\frac{1}{n}[y_{.i}(n - y_{i.}) + y_{i.}(n - y_{.i})]} \times 100$$

4) **Net Difference Rate (NDR):** NDR is the difference between the number of cases in the census and the number of cases in the PES that fall under each response category, relative to the total number of matched persons in all response categories.

$$NDR = \frac{(Y_{.i} - Y_{i.})}{n} \times 100$$

5) Aggregate index of Inconsistency: The Aggregate Index of Inconsistency (AII) is a summary measure of the index of Inconsistency (that is for all the response categories of the characteristic as a whole). The computation formula is as follows:

$$AII = \frac{(n - \sum_i^c Y_{ii})}{(n - \frac{1}{n}\sum_i^c Y_{i.}\, Y_{.i}} \times 100$$

2.10. Matching of PES and Census records

To assess the Census coverage, both PES and Census records were matched, a work that was done by a team of data scientists in the Unit of Data Revolution using data science techniques and tools.

Matching is a process of checking whether records from two different datasets relate to one person. In this work, both automatic and clerical matching methods were used:

- Automatic matching method: based on the predefined rules, the computer makes decision if a pair of records matches. This method reduces the number of records to match manually or clerically. This method is fast with low resources requirement;
- Clerical matching method: In this method, matching decisions are made by human judgement. The clerical matching will be done at household, Enumeration Area, district and countrywide. Clerical matching will help to estimate the false positive rate by manually reviewing a sample of pairs that were matched automatically and compute the precision. It will also estimate false negative rate by using clerical search on a sample of unmatched records and compute the

recall. In addition, this method will help resolve conflicting cases such as one PES record that is matching to multiple census records or vice versa.

The quality of the matching results is measured by ***precision*** and ***recall***, which are defined as follows:

a) Precision is a measure of the accuracy of the matches that have been made,

$$precision = \frac{number\ of\ true\ matches}{number\ of\ true\ matches + number\ of\ false\ matches}$$

b) Recall is a measure of the proportion of matches that have been made out of all the possible matches.

$$recall = \frac{number\ of\ true\ matches}{number\ of\ true\ matches + number\ of\ missed\ matches}$$

2.10.1. Matching steps

a) Preparation of the Matching Datasets:

Prior the matching, matching variables were selected and their combination was verified to uniquely identify a person that has to be matched. Moreover, the datasets were pre-processed by analyzing the values of variables, identifying the missing variables, removing non-alphanumeric characters from names, formatting the variables, deriving additional variables, removing persons marked as out of scope (children born after the census night (15 August), In-movers, confirmed duplicates in the PES and Census. Visitors were also removed from the Census data).

Additional variables that were derived include:

i. **The age tolerance limits**: This is one of the important parameters to match records. During the clerical matching work, there is an age tolerance that was suggested by UN agencies to increase individuals matching as. The following table shows the tolerances limits that were used in automatic matching.

Table 2.2: age tolerances used in automatic matching

Age	**Age Tolerance (in years)**
< 10 years	± 1
10-19 years	± 2
20-40 years	± 3

> 40 years	±4

ii. **Alpha-names:** is a derived variable that puts all characters of the names into alphabetical order.

iii. **Soundex**: is defined as the phonetic algorithm for indexing names by how they sound. It makes a code out of a name so that similar sounding names will the same code (for example the names JOSEPH and JOSEF are both J210. **RL-Soundex** is the Soundex function applied to the name with all 'R's switched with 'L's. In the context of Rwanda names, it was found that in many cases, "R" and "L" are used interchangeably. In such cases, the Soundex codes are not the same. For example, the Soundex codes of LIBERATTE and RIBERATA are L163 and R163 respectively, although these names are pronounced the same way in Kinyarwanda. To avoid this problem, the Soundex was applied to names with "Rs" and changed them into "Ls". Hence both LIBERATTE and RIBERATA get the same Soundex code of L163. Therefore, they could be matched.

2.10.2. Matching stages

The PES to CENSUS matching work was carried through the following stages:

Stage 1: Match within household: look for matches within the same household.

Stage 2: Match within enumeration area: look for matches within the same enumeration area.

Stage 3: Clerical resolution of unmatched cases at Enumeration Area level where the PES unmatched records will be matched to census records manually.

Stage 4: Probabilistic matching: where m and u probabilities were calculated for each variable used in matching.

The ***u-probability*** is defined as chance agreement between two records which are truly unmatched. On the other hand, **The "m probability"** is defined as chance agreement between two records which are truly matched[2].

Stage 5: Match within district: looking for matches within district.

Stage 6: Match within country: looking for matches within country.

Stage 7: Quality assurance. At this stage, the false positive and false negative match rates were estimated.

[2] https://link.springer.com/chapter/10.1007/978-3-319-78461-8_4

b) Development of Match-keys

After preparing the Matching Datasets**,** an algorithm was developed with detailed match-keys to match as many PES and CEN records as possible, hence minimize clerical matching work.

In that context, specific match-keys were developed according to each stage requirements in the following phases: the first phase was to develop a strict match-key that will result in obvious matches. The second phase was to create slightly relaxed match-keys aiming at discovering hard-to-find matches. These are for example a same person with two records whose names had errors or were swapped (first name in Census was recorded as last name in PES). Only relevant variables that can easily identify a person were selected in their respective categories as follows:

- Location variables: province, district, sector, cell, village, building number, household number, enumeration area (EA), person id
- Personal characteristics: names, sex, age, year of birth, month of birth and marital status,
- Other variables: relationship to household head.

Creation of a strict match key

A strict match key was created to find evident matches by joining together CENSUS and PES using variables such as names, year of birth, month of birth, sex and location identification. The location differs according to each stage of matching (Household, EA, district and Country levels). That match key did not allow any error or difference in the variables of the two records, they had to be similar.

Figure 2.3: Example of a strict match key

```
# Matchkey 1: Full Name + Year + Month + Household

matches_1 = pd.merge(left=CEN,
                     right=PES,
                     how="inner",
                     left_on =['first_name_cen','last_name_cen', 'year_birth_cen', 'birth_month_cen','sex_cen','hhid_cen'],
                     right_on=['first_name_pes','last_name_pes', 'year_birth_pes', 'birth_month_pes', 'sex_pes','hhid_pes'])
```

Creation of a relaxed match key

A relaxed match key is a key that identify hard to find matches by merging both CENSUS and PES and allowing errors in names and difference in other variables. Standardized Levenshtein Distance (Std_Lev) is a method used to measure the amount of difference in two names where the perfect agreement is one. If one variable is omitted in a match key, this means that the value of that variable can differ or be missing between two records.

Match keys had to be hierarchical, which meant that your first match key was to be your strictest, your second match key your second strictest and so on. As you add more and more match keys, you should allow for a gradual increase in the amount of errors allowed.

Figure 2.4: Example of a relaxed match key

```
# Matchkey 2: Standardized Levenshtein > 0.07 + Year + Month + Household
matches_2 = pd.merge(left=CEN,
                     right=PES,
                     how="inner",
                     left_on =['year_birth_cen', 'birth_month_cen', 'hhid_cen'],
                     right_on=['year_birth_pes', 'birth_month_pes', 'hhid_pes'])
matches_2['STD_EDIT1'] = matches_2[['first_name_cen','first_name_pes']].apply(lambda x: std_lev(str(x[0]), str(x[1])), axis=1)
matches_2['STD_EDIT2'] = matches_2[['last_name_cen','last_name_pes']].apply(lambda x: std_lev(str(x[0]), str(x[1])), axis=1)
matches_2 = matches_2[(matches_2.STD_EDIT1 > 0.70) & (matches_2.STD_EDIT2 > 0.70)]
```

Above key allows the variable "sex" to be different or missing in the two records while errors in names are allowed with a difference of one character. Match keys are created gradually to account possible errors spotted in the data for instance swapping of names.

Tuning period

The period refers to the period of PES data collection where some but not all of the PES data was available. The tuning aimed at testing and updating developed match-keys. The tuning period is a very important time because matching algorithms must be properly tuned using the real data to ensure that:

- The automatic matching captures as many correct matches as possible
- The automatic matching is as accurate as possible – ideally no false matches should be made automatically.
- The automatic matching algorithm is efficient and scalable to the real census and PES data.

Tuning the automatic algorithms leads to high precision and recall

During this time developed match keys (rules) were run on the available data and quality assurance of the results was done using clerical review. When the team was happy with the quality of the match keys, clerical review was used to estimate the False Positive (incorrect matches) rate for the automatic matching.

2.10.3. Matching Procedures

During Census to PES matching, the following procedures were done:

- **D*eterministic*** matching was used to automatically match persons that satisfied a series of rules called match-keys. The match-keys capture exact matches, but also allow for error in some of the variables. It was used at all matching stages.
- **P*robabilistic*** matching was also used to match person records. Pairs of Census and PES records that are awarded the agreement or disagreement weight for each of the variables used for matching (first name, last name, sex, birth year, marital status, building +household number) according to whether the variable agrees or disagrees for this pair. Partial scores are awarded when the variable partially agrees to allow for some error in names, year of birth and building+household number. The agreement and disagreement scores for the pair are then summed to give a final score for each pair and any record pairs scoring above this threshold are accepted as automatic matches. Probabilistic matching was only used at EA level.

  Agreement and disagreement weights were calculated from the m and u parameters as follows:

  - $Agreement\ weight\ = \log_2(\frac{m}{u})$
  - $Disagreement\ weight\ = \log_2(\frac{1-m}{1-u})$

- ***Associative clerical matching*** was used to find additional matches in households that had already been linked together by a person match. Relatively relaxed matchkeys were used for people where there is already one-person match in the same household.
- ***Clerical Resolution Online Widget (CROW):*** which is a computer based manual matching system which was used to solve conflicting matches (a single census record matching to multiple PES records and vice-versa) known also as *CROW Cluster*. A similar computer based system was also used to compare PES to Census records to confirm matching status (match or not match) known as *CROW Pairwise*
- ***Pairwise clerical matching*** was used to determine the true match status for pairs of census and PES records which were matched automatically. It was also used to estimate false positive[3].
- ***Clerical search*** was used to maximize matches at enumeration area which were not matched automatically.

[3] False Positive is a result that is classified as correct while it is incorrect. In this case, the result that shows that records are matching while not, is the one that is called a False Positive

Chapter 3 : Survey findings

3.1. Matching and Reconciliation

This chapter provides a summary of matching activities that were implemented after PES data collection was completed and presents findings from matching and reconciliation visits.

3.1.1. Matching activities

PES to Census matching was carried out in seven stages as defined in the methodology.

Stage 1: **Matching process within household:** In the first stage, PES and Census records passed through a set of automatic match keys (rules) ranging from strict to slightly relaxed ones which resulted in obvious matches. Conflicts (PES record matching to multiple Census records or vice versa) were resolved using clerical system (CROW Cluster). Any matches found here after resolving conflicts was considered to be correct and final. Automatic associative matching within household was conducted using relaxed match keys in search for possible matches from people where there is already at least one person matching between respective households. After household stage, matched PES and Census records were removed from the pot for next stages matching

Stage 2: **Matching process within Enumeration Area (EA)**: records which were not matched at the household level were searched within the same enumeration Area and conflicts were resolved (non-unique matches) using clerical system (CROW cluster). Any matches found here after resolving conflicts was considered to be correct and final. Automatic associative matching in households within the same Enumeration Area was conducted using relaxed match keys in search for possible matches from people where there is already at least one person matching between respective households.

Stage 3: **Clerical resolution of enumeration area unmatched:** Unmatched records in PES were matched to census records manually in sampled enumeration areas to find potential matches which may have not been matched automatically and require human judgement.

Stage 4: Probabilistic matching: Although the matching strategy was mainly based on the use of a deterministic match key method, Probabilistic method was explore to

find new matches at EA level using m and u probabilities which as required for the Fellegi-Sunter probabilistic method[4].

m probability: the probability that a variable agrees on the census and the PES given the pair are a true match and ***u probability:*** the probability that a variable agrees on the census and the PES, given the pair are not a true match.
EA_IDs were blocked on, meaning that every PES record was compared with every census record from the same EA.

The records being compared were awarded the agreement or disagreement weight for each of the variables used for matching (first name, last name, sex, birth year, marital status, building + household number) according to whether the variable agrees or disagrees for this pair. Partial scores were awarded when the variable partially agrees to allow for some error in names, year of birth and building + household number.

The agreement and disagreement scores for the pair were then summed to give a final score for each pair. Clerical review was used to set a score threshold of 23 because all scores above this one were showing matching records. Any record pairs scoring above this threshold were accepted as automatic matches.

The m and u values and agreement and disagreement weights calculated for the census to PES matching are given in the table below.

Table 3.1: m and u probabilities and agreement and disagreement weights used in the Fellegi - Sunter probabilistic matching algorithm in this work

Variable	m	u	Agreement weight	Disagreement weight
First name	0.933987	0.002935	8.31404	-3.91686
Last name	0.693792	0.003065	7.822281	-1.70299
Sex	0.992075	0.552997	0.843177	-5.81775
Birth year	0.837165	0.019545	5.420645	-2.59004
Marital status	0.930743	0.275495	1.756359	-3.38697
Building + household	0.875863	0.000889	9.944881	-3.00871

Stage 5: **Match within District**: Although most people matched at the correct location, some people will have moved between the census and PES data collection. We therefore looked for matches in the census data outside of the PES Enumeration Areas (EAs). To confirm that matches made outside of the PES EAs were correct, we

[4] https://www.census.gov/content/dam/Census/library/working-papers/1991/adrm/rr91-9.pdf

insisted that two or more persons were matching in the matched households. However, there were a great number of conflicts even with this confirmation step in place. As there was no way to tell which (if any) is the correct match in these circumstances, we decided that in the final matching run, it was only possible to accept matches outside of EAs where there was no conflict

Stage 6: Match within country: looked for people who moved to other districts or whose location could be wrongly recorded. ***Pairwise clerical matching*** was used to determine the true match status for pairs of census and PES records which were matched at this level.

Stage 7: Quality assurance: Precision is a measure of the accuracy of the matches that have been made. Recall is a measure of the proportion of matches that have been made out of all the possible matches.

For PES 2022 to Census matching, precision and recall set targets are 99.90% and 99.75% respectively for individual level matching (i.e. Matching between records of persons or people). These targets are extremely high compared to typical linkage quality targets because any linkage error directly impacts the quality of the census estimates. For this work, set targets were achieved with 99,94% accurate matches (precision) and 99,98 recall[5] .

3.1.2. Reconciliation of PES and Census unmatched records

After completing the matching, reconciliation processes were undertaken. The reconciliation aimed at helping resolve doubtful cases for potential matches and verifying the reasons for unmatched records in both Census and PES datasets. Specifically, the reconciliation was conducted to solve the following major issues:

- With regard to persons who appeared in the Census but not in PES, reconciliation visits and calls aimed at determining whether those persons were correctly or erroneously enumerated in the census;
- With regard to persons who appeared in the PES but not in the Census, the reconciliation visits and calls aimed at determining whether those persons were correctly or erroneously enumerated in the PES.
- Clarify cases with unclear or insufficient persons' presence.

The reconciliation of Census records not in PES records was done in two stages: The first stage was to examine missed members in matched households: a match key was developed to display unmatched persons from matched households. The resulting persons were visited in all districts of Kigali and called in other districts to confirm

[5] Recall is a measure of the proportion of matches that have been made out of all the possible matches.

whether they were usual residents or not. Findings confirmed that unmatched census persons were correctly enumerated (actual resident). The second stage was to compute the number of cases where PES had missed all members of the Household (where the whole household was missed in PES). Most of these cases were people in households that were moved or expropriated due to economic activities such as road constructions or other environmental protection reasons.

3.1.2.1. Reconciliation of unmatched PES records

In the case of persons appearing in the PES records but not in the census, reconciliation visits helped to confirm whether such persons were usual residents as of the census night (non-movers or out-movers in PES). On the other hand, whether they arrived or were born after the census night (in-movers). Below is the description of activities undertaken during reconciliation of PES unmatched:

- Unmatched PES persons who were matched to Census visitors: All PES unmatched were applied on Census dataset of visitors searching to see if there were persons in PES usual residents which were enumerated as visitors during Census. As a result, 480 matches were made and removed from PES population because they were out of scope of Census as visitors were not supposed to be enumerated as usual residents of the household.
- Review of Missed Members in already Matched households: In PES unmatched, there were situations where in the same household some records had matched in the Census dataset while other did not. Unmatched persons from those Census and PES households were displayed using CROW pairwise to search for potential matches. As a result, 547 matches were confirmed and were added to the matched data set. This mismatch had occurred due to the reason that most of the people had provided information in PES that is different from the ones provided in Census
- New found Matches outside household level: further relaxed match keys were developed allowing errors in names to search for further potential matches in both Census and PES unmatched because there were people who were enumerated from different household. Found matches were confirmed via telephone calls. 138 persons were confirmed matches and were added to the matched data set.
- Identified new (In-mover) households in PES: in unmatched PES, it was found that some households were moved to PES EAs after the Census. 90 persons from in-mover households were identified and were removed from the PES population since they were PES out of scope.

3.2. Matching results

The net coverage rate at national level is 98.7%. The highest net coverage rate is observed in Northern and Western Provinces with 99.0%, while the lowest is observed in the City of Kigali with 96.9%. Rural areas have a greater net coverage rate compared to urban areas with 99% and 97.7% respectively.

In age groups, the highest net coverage rate is observed in the group of 45 to 59 years with a rate of 99.1%, while the lowest NCR is observed in the group of 15 to 29 years with 97.9%.

Females have the greater net coverage rate compared to males with 98.9% and 98.4% respectively.

Table 3-2: Net Coverage Rate (%) by age group, province, Sex and area of residence

	Male	Female	Urban	Rural	Total
Rwanda	**98.4**	**98.9**	**97.7**	**99.0**	**98.7**
Age groups					
0-4	98.5	98.8	97.7	99.0	98.6
5-14	98.8	98.8	98.1	99.1	98.8
15-29	97.5	98.3	96.7	98.3	97.9
30-44	98.4	98.9	97.9	99.0	98.7
45-59	98.8	99.5	98.4	99.4	99.1
60+	98.7	99.3	97.6	99.5	99.0
Province					
Kigali City	95.9	97.8	96.8	97.6	96.9
Southern Province	98.2	98.9	98.2	98.6	98.5
Western Province	98.8	99.2	98.5	99.2	99.0
Northern Province	98.9	99.1	97.1	99.2	99.0
Eastern Province	99.2	99.1	98.2	99.3	99.1

Source: PES 2022

3.3. Coverages rates

3.3.1. Net Coverage Error Rate

The Net Coverage Error Rate (NCER) at national level is 1.31%. The highest NCER is observed in the City of Kigali with 3.14%, while the lowest is observed in Eastern

Province with 0.88%. By areas of residence, Urban and rural areas are having a notable difference in NCER of 1.8% and 0.7% respectively. [6]

In age groups, it was observed that the group of 15 to 19 years has the greatest NCER with 2.13%, while the smallest is in the group of 60 years and above. The table 3.2 summarize NCER of age groups and provinces by sex and residence areas.

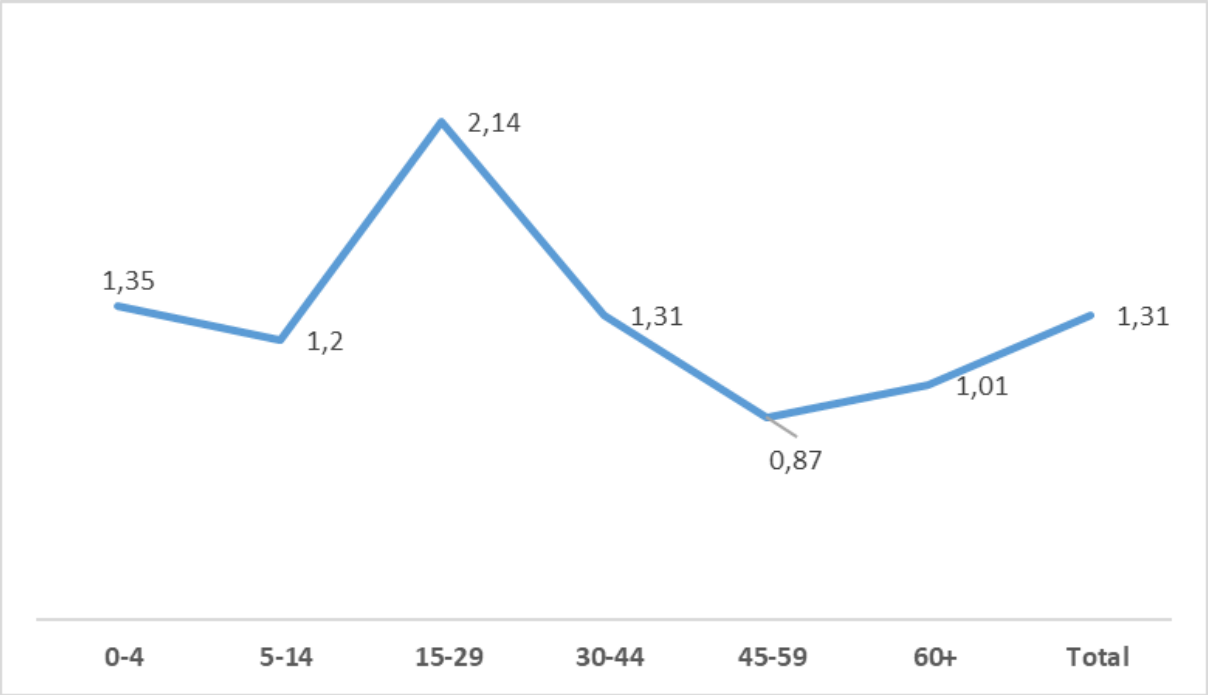
Table 3-3: Net Coverage error Rate (%) by age group, province, Sex and area of residence

	Urban			**Rural**			**Total**		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Rwanda	2.87	1.65	2.26	1.08	0.86	0.97	1.56	1.07	1.31
Age Groups									
0-4	2.56	2.12	2.34	1.09	0.9	0.99	1.48	1.22	1.35
5-14	1.88	1.94	1.91	0.98	0.91	0.94	1.22	1.18	1.2
15-29	3.88	2.63	3.26	2.05	1.42	1.74	2.54	1.75	2.14
30-44	2.64	1.54	2.09	1.16	0.88	1.02	1.55	1.06	1.31
45-59	2.79	0.4	1.6	0.64	0.57	0.61	1.22	0.52	0.87
60+	3.46	1.3	2.38	0.55	0.47	0.51	1.32	0.69	1.01
Total	2.87	1.65	2.26	1.08	0.86	0.97	1.56	1.07	1.31
Provinces									
City of Kigali	4.13	2.35	3.24	3.71	1.07	2.39	4.08	2.2	3.14
Southern Province	2.22	1.4	1.81	1.72	1.03	1.37	1.83	1.11	1.47
Western Province	1.71	1.2	1.46	1.01	0.62	0.81	1.21	0.78	1.
Northern Province	4.57	1.18	2.88	0.74	0.88	0.81	1.12	0.91	1.02
Eastern Province	2.17	1.42	1.8	0.55	0.85	0.7	0.82	0.95	0.88

Source: PES 2022

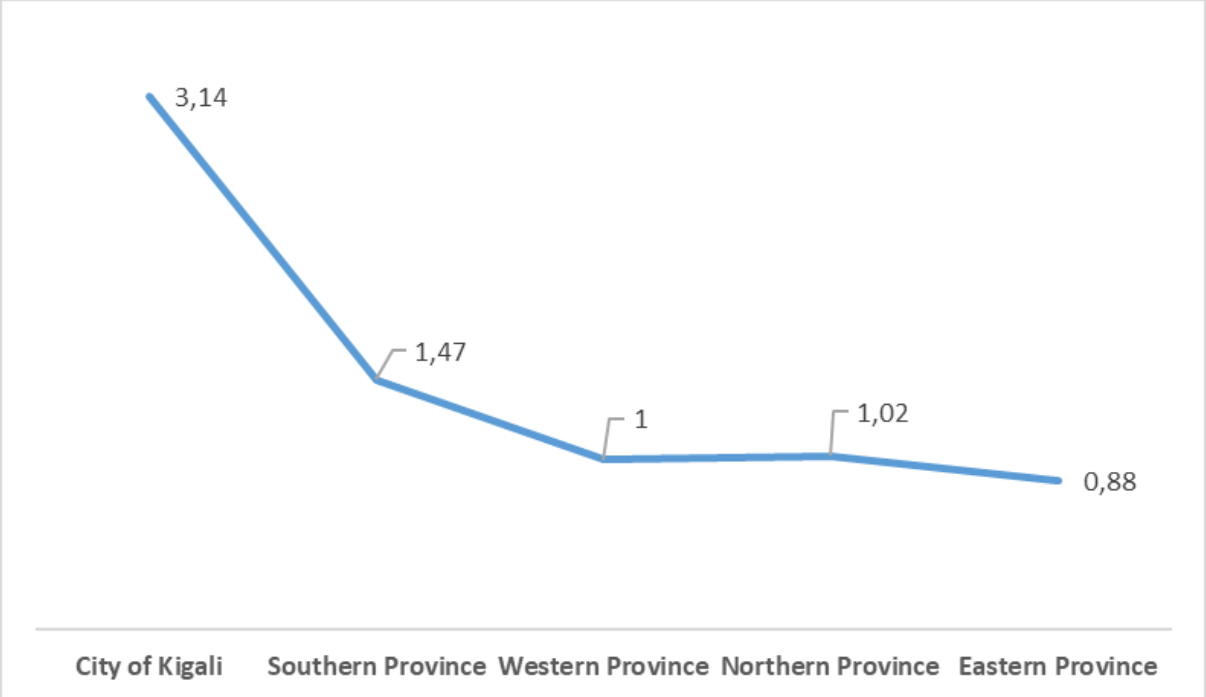
[6] During PES, the sampled EAs included only private Households; I.e, the estimated population in the sampled EAs was from private Households (Institutional Households excluded); while during the Census, the Population in both private and Institutional Households were interviewed.

Figure 3.1: Net Coverage Error Rate (%) by age groups



Source: PES 2022

Figure 3.2: Net Coverage Error Rate (%) by Provinces



Source: PES 2022

3.3.2. Census Omission Rates

It appears that at the national level, the COR is 1.51%. Furthermore, there's a distinction between urban and rural areas, with urban areas having a higher COR of 2.45% compared to rural areas, with 1.17%.

The difference of COR between Males and Females is 0.58 percentage points with 1.80% for males and 1.22 % for females.

There is also a variation in COR across age categories. Specifically, the highest COR is observed in the age group of 15 to 29 years, with a rate of 2.34%, while the lowest COR is found in the group of 45 to 59 years, with a rate of 1.12%.

By Provinces, the highest COR is observed in the City of Kigali with 3.45% while the lowest is observed in Northern Province with 1.09%. The table below summarize the COR for age groups and provinces by sex and residence areas.

Table 3-4: Census Omission Rates (%) by age group, province, Sex and area of residence

	Male	Female	Urban	Rural	Total
Rwanda	1.80	1.22	2.45	1.17	1.51
Age groups					
0-4	1.64	1.28	2.49	1.09	1.46
5-14	1.44	1.41	2.16	1.16	1.42
15-29	2.77	1.91	3.48	1.92	2.34
30-44	1.64	1.14	2.13	1.12	1.39
45-59	1.52	0.72	1.92	0.83	1.12
60+	1.81	0.84	2.51	0.90	1.33
Provinces					
City of Kigali	4.45	2.45	3.58	2.39	3.45
Southern Province	2.00	1.23	1.90	1.54	1.62
Western Province	1.35	0.92	1.58	0.95	1.13
Northern Province	1.23	0.95	2.99	0.88	1.09
Eastern Province	1.31	1.16	1.89	1.10	1.24

Source: PES 2022

3.3.3. Rate of Erroneous Inclusions

The overall Rate of Erroneous Inclusion (REI) at the national level is 0.19 %. The highest REI is observed in the City of Kigali with a rate of 0.31%. Northern Province has the lowest REI at 0.07%, indicating a lower rate of erroneous inclusions compared to both the national average and other provinces.

The age group of 30 to 44 years has the smallest REI at 0.08% whereas the highest REI is observed in the group of 60 years and above, with a rate of 0.27%. The difference in REI between males and females is 0.09 percentage points, indicating a slight variation in inclusion accuracy between genders. The difference in REI between urban and rural areas is not explicitly mentioned, suggesting that the REI is similar in both settings. The table below illustrates REIs of Provinces and age groups by sex and residence areas.

Table 3-5: Rate of Erroneous Inclusions (%) by age group, province, Sex and area of residence

	Male	Female	Urban	Rural	Total
Rwanda	0.23	0.14	0.19	0.19	0.19
Age groups					
0-4	0.16	0.05	0.14	0.09	0.11
5-14	0.22	0.22	0.25	0.21	0.22
15-29	0.23	0.17	0.23	0.19	0.20
30-44	0.08	0.08	0.04	0.10	0.08
45-59	0.29	0.19	0.32	0.21	0.24
60+	0.39	0.15	0.12	0.32	0.27
Provinces					
City of Kigali	0.37	0.25	0.34	0.00	0.31
Southern Province	0.17	0.12	0.09	0.16	0.15
Western Province	0.13	0.14	0.13	0.14	0.14
Northern Province	0.10	0.04	0.11	0.07	0.07
Eastern Province	0.42	0.21	0.09	0.36	0.31

Source: PES 2022

3.3.4. Gross Coverage Error Rate

The National Level Gross Coverage Error Rate (GCER) is as low as 1.78%. The level of gross coverage error rate is higher in urban areas (2.79%) compared to rural areas (1.41%). Also, the level of gross coverage error for males (2.15%) is above that of females 1.41%). With regard to age, gross coverage error rate is highest (2.60%) for the age group 15 to 29 and lowest for the group of 45 to 59 year (1.40%). The highest GCER is observed in the City of Kigali and lowest in Northern province with 4.09%, and 1.23% respectively. The table below summarize the Gross Coverage Error Rates in age groups and provinces by sex and residence areas.

Table 3-6: Gross Coverage Error Rate per unit enumeration (%) by age group, province, Sex and area of residence.

	Male	Female	Urban	Rural	Total
Rwanda	2.15	1.41	2.79	1.41	1.78
Age groups					
0-4	1.91	1.42	2.69	1.29	1.66
5-14	1.72	1.68	2.60	1.38	1.70
15-29	3.06	2.14	3.84	2.15	2.60
30-44	1.80	1.27	2.27	1.27	1.53
45-59	1.86	0.94	2.29	1.08	1.40
60+	2.54	1.03	3.04	1.33	1.78
Provinces					
City of Kigali	5.30	2.87	4.31	2.28	4.09
Southern Province	2.29	1.41	2.02	1.81	1.85
Western Province	1.53	1.08	1.72	1.14	1.31

Northern Province	1.38	1.08	3.19	1.01	1.23
Eastern Province	1.80	1.36	2.06	1.48	1.58

Source: PES 2022

3.4. Estimation of Content errors

All the PES/Census Content errors are computed basing on matched population for the following characteristics: Sex, Age Groups, Marital Status, Health Insurance, Relationship, Type of toilet Facility and so on. The following table guides in the interpretation of Contents Errors:

Table 3-7: Interpretation of content errors

Measure	**Level**		
	Low	**Medium**	**High**
Index of inconsistency	< 20	20 – 50	> 50
Aggregate index of inconsistency	< 0.01	0.01 0.05	> 0.05

3.4.1. Content Errors

a) **Net Difference Rate (NDR)**: The highest NDR is observed in the category of "Marital Status" with the characteristic "Married to one wife/husband officially," having a NDR of 1.03%. In addition, the category of "Relationship to the Head of Household" with the characteristic "Father in law/Mother in law", and the category of "Health Insurance" with the characteristic "NGOs", exhibits no difference in net difference rate with 0.00%. This indicates that there is no net difference in inclusion rates for individuals with these characteristics compared to the overall average inclusion rate.

b) **Index of Inconsistency**: The greatest index of inconsistency is observed among characteristics of Relationship to the head of households and Health Insurance where Unknown relationship (for Relationship to HH head) and NGO (for Health Insurance) has 100%, which means there's complete inconsistency for these categories. The lowest IOI is related to the "age groups", specifically for "60+ years," with an IOI of 3.17, which means there is relatively less inconsistency for this age group.

c) **Aggregate Index of Inconsistency:** Health insurance appears to have the highest level of inconsistency at 33.01%, whereas sex and age groups exhibit much lower levels of inconsistency, with 3.34% and 4.20% respectively. This might suggest that health insurance-related factors vary more widely compared to factors related to sex and age groups.

d) **Gross Difference Rate:** The highest GDR, which is 10.5%, is observed in the characteristic of Marital Status. On the other hand, the lowest GDR values are in the characteristics of Sex (1.67%) and Age groups (3.35%). This indicates relatively smaller differences in these characteristics compared to Marital Status.

e) Rate of agreement: The rate of agreements is more than 90% in all characteristics. The rate of agreements being more than 90% in all characteristics typically means that, after a survey or enumeration process has been conducted, the data collected through the survey shows a high level of agreement on various characteristics or variables being examined. For example, if a post-enumeration survey reveals that more than 90% of respondents agree on specific characteristics like Relationships to the Head of Households, Sex, Marital Status, Age groups, or any other variables of interest, it indicates that the data collected in the census was accurate and reliable. High levels of agreement in a post-enumeration survey is a positive sign, suggesting that the initial data collection process was effective and produced trustworthy results.

Table 3-8: Content Errors by characteristics7

	Net Difference Rate (NDR)	Lower NDR	Upper NDR	Index of Inconsistency	Lower Index of Inconsistency	Upper Index of Inconsistency	Aggregate Index of Inconsistency	Lower Aggregate Index of Inconsistency	Upper Aggregate Index of Inconsistency	Gross Difference Rate (GDR)	Rate of agreement
Relationships to the Head of Households											
Household head	0.04	-0.05	0.14	5.05	4.81	5.30	8.98	8.74	9.23	5.95	94
Spouse	0.02	-0.06	0.09	4.86	4.58	5.15	8.98	8.74	9.23	5.95	94
Household head son or daughter	0.28	0.17	0.40	6.03	5.80	6.26	8.98	8.74	9.23	5.95	94
Household adoptive child	-0.03	-0.08	0.02	78.01	71.78	84.77	8.98	8.74	9.23	5.95	94
Father / Mother	-0.01	-0.02	0.01	31.01	22.96	41.88	8.98	8.74	9.23	5.95	94
Father in law/Mother in law	0.00	-0.01	0.01	31.44	20.59	48.01	8.98	8.74	9.23	5.95	94
Bother in law / Sister in law	-0.05	-0.08	-0.02	59.15	50.88	68.77	8.98	8.74	9.23	5.95	94
Brother / Sister	-0.02	-0.07	0.03	40.92	37.38	44.79	8.98	8.74	9.23	5.95	94
Grand Child	-0.11	-0.19	-0.03	14.29	13.51	15.12	8.98	8.74	9.23	5.95	94
Son in Law / Daughter in law	0.01	-0.01	0.02	67.58	45.42	100.56	8.98	8.74	9.23	5.95	94
Other relation	0.13	0.06	0.20	53.59	50.39	56.99	8.98	8.74	9.23	5.95	94
Housemaid	-0.09	-0.14	-0.04	26.46	24.28	28.85	8.98	8.74	9.23	5.95	94
Not related	-0.09	-0.16	-0.03	71.25	66.69	76.12	8.98	8.74	9.23	5.95	94
Unknown relationship	-0.09	-0.11	-0.06	100.02	81.75	122.37	8.98	8.74	9.23	5.95	94
Sex											
Male	-0.10	-0.18	-0.01	3.34	3.18	3.52	3.34	3.18	3.52	1.67	98

[7] Confidence Intervals were computed 95% confidence level

	Net Difference Rate (NDR)	Lower NDR	Upper NDR	Index of Inconsistency	Lower Index of Inconsistency	Upper Index of Inconsistency	Aggregate Index of Inconsistency	Lower Aggregate Index of Inconsistency	Upper Aggregate Index of Inconsistency	Gross Difference Rate (GDR)	Rate of agreement
Female	0.10	0.01	0.18	3.34	3.18	3.52	3.34	3.18	3.52	1.67	98
Marital Status											
Married to one wife / husband officially	1.03	0.85	1.20	10.87	10.48	11.27	14.71	14.34	15.09	10.05	90
Married to one wife / husband not officially	-0.04	-0.23	0.15	20.15	19.47	20.84	14.71	14.34	15.09	10.05	90
Live in a polygamous union	0.25	0.15	0.35	74.93	70.22	79.96	14.71	14.34	15.09	10.05	90
Divorced	0.03	-0.01	0.06	46.25	38.54	55.50	14.71	14.34	15.09	10.05	90
Separated	-0.34	-0.46	-0.22	62.33	59.18	65.63	14.71	14.34	15.09	10.05	90
Never Married	-0.18	-0.32	-0.04	6.39	6.10	6.68	14.71	14.34	15.09	10.05	90
Widowed	-0.75	-0.88	-0.62	25.50	24.26	26.81	14.71	14.34	15.09	10.05	90
Health Insurance											
Mutuelle de Sante	-0.35	-0.50	-0.20	30.85	29.94	31.78	33.01	32.08	33.96	5.46	95
RSSB (Ex RAMA)	0.16	0.09	0.23	13.92	13.06	14.84	33.01	32.08	33.96	5.46	95
MMI	-0.10	-0.14	-0.06	19.91	17.91	22.13	33.01	32.08	33.96	5.46	95
Schools	-0.02	-0.03	0.00	75.41	56.72	100.25	33.01	32.08	33.96	5.46	95
Employer	-0.13	-0.15	-0.10	100.03	84.54	118.35	33.01	32.08	33.96	5.46	95
Private insurance	0.13	0.08	0.17	43.17	39.09	47.68	33.01	32.08	33.96	5.46	95
NGOs	0.00	-0.02	0.03	70.73	58.26	85.86	33.01	32.08	33.96	5.46	95
None	0.37	0.25	0.50	62.82	60.65	65.08	33.01	32.08	33.96	5.46	95
Do not know	-0.07	-0.10	-0.05	98.69	83.56	116.57	33.01	32.08	33.96	5.46	95
Age groups											
0-4	0.04	-0.02	0.10	3.55	3.29	3.82	4.20	4.05	4.35	3.35	97

	Net Difference Rate (NDR)	Lower NDR	Upper NDR	Index of Inconsistency	Lower Index of Inconsistency	Upper Index of Inconsistency	Aggregate Index of Inconsistency	Lower Aggregate Index of Inconsistency	Upper Aggregate Index of Inconsistency	Gross Difference Rate (GDR)	Rate of agreement
5-14	-0.12	-0.20	-0.03	4.53	4.31	4.77	4.20	4.05	4.35	3.35	97
15-29	-0.09	-0.17	0.00	4.28	4.07	4.50	4.20	4.05	4.35	3.35	97
30-44	0.08	0.01	0.16	3.95	3.72	4.19	4.20	4.05	4.35	3.35	97
45-59	0.03	-0.03	0.09	5.35	4.99	5.75	4.20	4.05	4.35	3.35	97
60+	0.05	0.01	0.09	3.17	2.85	3.52	4.20	4.05	4.35	3.35	97

3.5. Detailed Analysis Accounting for Sampling Errors by Urban/Rural Status, Gender, Age Groups, and Province

Table 3-10 provides estimates of the true population for different demographic categories in Rwanda, accompanied by their respective 95% confidence intervals and standard errors. Confidence intervals indicate the range within which the true population is likely to fall with 95% certainty. Standard errors reflect the precision of the estimates. For instance, the estimated true population of Rwanda falls within the interval of 13.31 million to 13.34 million, with a standard error of 8,310. Urban areas are projected to have around 3.69 to 3.74 million inhabitants, compared to rural regions, which are estimated to house approximately 9.60 to 9.64 million people. By gender, male population ranges from 6.41 to 6.45 million and females from 6.89 to 6.92 million. Age-wise, the largest population is in the 15-29 age group, followed by 5-14, indicating a youthful demographic profile. Province-wise, there are variations, with the City of Kigali having the smallest estimated population and the Eastern Province the largest. Smaller values of standard errors across all demographic categories indicate higher precision of the population. In practical terms, these confidence intervals and standard errors help policymakers, researchers, and analysts understand the reliability and variability of the population estimates, allowing them to make informed decisions and interpretations based on the data. These nuanced insights underscore the importance of tailored strategies to address the diverse needs of Rwanda's population across different regions and demographics.

Table 3-10: Estimated True Population, Confidence Intervals, and Standard Errors

	Estimated True Population	Confidence Interval 95%		Standard Error
		Lower	Upper	
Rwanda[8]	**13,321,977**	**13,305,690**	**13,338,264**	**8,310**
Sex				
Male	**6,427,562**	**6,409,059**	**6,446,066**	**9,441**
Female	**6,902,061**	**6,885,014**	**6,919,108**	**8,697**
Urban/Rural				
Urban	**3,711,141**	**3,685,286**	**3,736,995**	**13,191**
Rural	**9,617,268**	**9,597,008**	**9,637,527**	**10,337**
Age groups				
0-4	**1,732,362**	**1,725,776**	**1,738,948**	**3,360**
5-14	**3,294,910**	**3,286,300**	**3,303,520**	**4,393**
15-29	**3,723,378**	**3,710,732**	**3,736,024**	**6,452**

[8] *The estimated true population was derived from the census population excluding institutional population, as PES interviewed only households not institutional household.*

	Estimated True Population	Confidence Interval 95%		Standard Error
		Lower	Upper	
30-44	**2,539,383**	**2,529,141**	**2,549,625**	**5,226**
45-59	**1,182,663**	**1,178,648**	**1,186,678**	**2,048**
60+	**857,792**	**854,631**	**860,953**	**1,613**
Province				
City of Kigali	**1,788,503**	**1,764,087**	**1,812,920**	**12,457**
Southern Province	**3,015,877**	**3,003,438**	**3,028,317**	**6,347**
Western Province	**2,911,405**	**2,903,286**	**2,919,523**	**4,142**
Northern Province	**2,045,903**	**2,035,101**	**2,056,706**	**5,511**
Eastern Province	**3,568,259**	**3,556,490**	**3,580,029**	**6,005**

Source: PES 2022

Conclusions and Recommendations

The use of ICT and data science techniques and tools, brought numerous benefits. It reduced human potential errors, enhanced coordination efforts, and improved oversight mechanisms. Most importantly, it drastically shortened the timeline for obtaining matching results. While the PES2012 process, reliant on manual and paper-based tools, required 6 months to complete data linkage to its Census, PES2022 achieved the same outcomes in just three weeks. This marked improvement, underscoring the transformative impact of embracing technology and data science techniques and tools into statistical processes.

The PES2022 utilized technology-driven processes for its implementation, encompassing data collection, processing, matching operations between RPHC2022 data and PES2022 data, and comprehensive data analysis. The algorithms designed for data matching and analysis were developed by NISR's dedicated team of data analysts and data scientists using the Python programming language, with technical assistance from the UK Office for National Statistics (ONS). Additionally, the Python scripts developed during this exercise have been carefully documented and archived. These resources will be invaluable for future data linkage applications and similar projects, ensuring efficiency, consistency, and the continued application of cutting-edge methodologies in NISR's operations.

The RPHC2022 was successful in terms of quantity and quality. The Post Enumeration Survey found that Content errors are relatively small. The gross differences are small for all characteristics. This means a high level of consistency in the data from two sources. Therefore, the census results can be used for evidence-based decision-making and planning.

These results prove the effective planning, coordination and efforts of dedicated and well-trained teams. The use of modern technology played a vital role and hastened data collection, data transmission, and matching processes which used to take longer in previous Population and housing censuses and post-enumeration surveys. This is an achievement that has to be enhanced and where it may be possible extended to other surveys/censuses.

Accessibility of people living in urban areas (particularly in the city of Kigali) as well as young people is uneasy, and this requires extra efforts for interviewers to get them or sometimes getting the required information from them. It is in this regard that PES2022 recommends early engagement of Local Leaders living in major cities and

Youth leaders right from the outset of census preparations in order to play a leading role in mobilizing the people they lead. It also recommends more trainings of enumerators and always remind them to pay close attention while recording responses of household members, especially information of relationships to the head of household and housing characteristics because PES2022 results show that those characteristics have high index of inconsistency between Census and PES.

In the future, as administrative data systems become more advanced and robust, they may be utilized to assess the quality of census data.

References

1. RPHC4: Post Enumeration Survey, NISR, https://www.statistics.gov.rw/publication/rphc4-post-enumeration-survey.
2. Post Enumeration Surveys Operational guidelines, Technical Report, United Nations Secretariat Department of Economic and Social Affairs Statistics Division, https://unstats.un.org/unsd/demographic/standmeth/handbooks/Manual_PES en.pdf;
3. Quality Control and Quality Assurance Strategy for 2021 Census to CCS Person and Household Matching. Rachel Shipsey & David Edwards, May 2021, https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2021/11/EAP163-Quality-Control-and-Quality-Assurance-Strategy-for-2021-Census-to-CCS-Person-and-Household-Matching.pdf
4. Data-Driven Policy Impact Evaluation Nuno Crato Paolo Paruolo Editors How Access to Microdata is Transforming Policy Design, https://link.springer.com/content/pdf/10.1007/978-3-319-78461-8.pdf
5. RPHC4: Main Indicators Report https://www.statistics.gov.rw/publication/rphc4-main-indicators-report
6. Evaluation of Census Results with the Post Enumeration Survey https://ask.rks-gov.net/media/3251/pes_raporti_eng.pdf
7. http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_PES.pdf

Annex

Table A.0-1: Distribution of Matched persons according to their Age Groups as reported in the Census and the PES

		PES						
		0-4	**5-14**	**15-29**	**30-44**	**45-59**	**60+**	**Total**
Census	**0-4**	**11,536**	372	10	3	1	0	11,922
	5-14	337	**22,418**	390	2	2	0	23,149
	15-29	10	464	**23,729**	243	7	4	24,457
	30-44	1	0	399	**16,920**	191	5	17,516
	45-59	0	2	7	261	**7,842**	145	8,257
	60+	0	0	0	10	187	**5,754**	5,951
	Total	11,884	23,256	24,535	17,439	8,230	5,908	**91,252**

Source: PES 2022

Table A.0-2: Distribution of Matched persons according to their Sex as reported in the Census and the PES

	Male	**Female**	**Total**
Male	42,746	717	43,463
Female	805	46,984	47,789
Total	43,551	47,701	91,252

Source: PES 2022

Table A.0-3: Distribution of Matched persons according to their Relationships as reported in the Census and the PES

		PES														
		Household head	Spouse	Household head son or daughter	Household adoptive child	Father / Mother	Father in law / Mother in law	Bother in law / Sister in law	Brother / Sister	Grand Child	Son in Law / Daughter in law	Other relation	Housemaid	Not related	Unknown relationship	Total
Census	**Household head**	22,217	479	175	8	11	2	8	36	10	1	25	37	92	7	23,108
	Spouse	474	13,772	71	3	0	1	6	5	5	1	10	10	10	1	14,369
	Household head son or daughter	219	54	43,836	140	2	1	30	71	535	2	183	103	136	27	45,339
	Household adoptive child	1	1	140	83	0	0	1	6	15	0	39	23	46	4	359
	Father / Mother	9	1	0	0	49	4	0	1	0	0	2	0	1	1	68
	Father in law/Mother in law	3	0	0	0	8	24	0	0	0	0	1	0	1	0	37
	Bother in law/Sister in law	0	2	13	2	0	0	61	22	2	0	14	3	7	0	126
	Brother/Sister	29	5	67	5	0	0	37	357	7	0	54	9	20	1	591
	Grand Child	14	2	440	7	0	0	3	18	3,998	3	58	13	19	1	4,576
	Son in Law / Daughter in law	1	12	1	0	0	0	0	0	0	6	1	0	0	0	21
	Other relation	17	7	179	39	4	1	16	61	78	2	469	43	117	25	1,058
	Housemaid	27	8	74	18	0	0	4	9	6	0	30	759	49	3	987
	Not related	56	7	83	78	0	0	6	26	19	1	52	68	189	18	603
	Unknown relationship	0	1	4	2	0	0	0	0	0	0	0	1	2	0	10
	Total	23,067	14,351	45,083	385	74	33	172	612	4,675	16	938	1,069	689	88	91,252

Source: PES 2022

Table A.0-4: Distribution of Matched persons according to their Marital Status as reported in the Census and the PES

	Married to one wife/husband officially	Married to one wife/husband not officially	Live in a polygamous union	Divorced	Separated	Never Married	Widowed	Total
Married to one wife/husband officially	18,986	751	113	18	190	186	558	20,802
Married to one wife/husband not officially	571	8,495	194	7	241	453	243	10,204
Live in a polygamous union	139	303	165	0	46	23	42	718
Divorced	8	12	0	70	26	15	7	138
Separated	106	170	39	18	476	246	63	1,118
Never Married	135	352	28	4	276	25,396	120	26,311
Widowed	215	145	21	5	75	106	2,512	3,079
Total	20,160	10,228	560	122	1,330	26,425	3,545	62,370

Source: PES 2022

Table A.0-5: Distribution of Matched persons according to their Health Insurances as reported in the Census and the PES

	Mutuelle de Santé	RSSB (Ex RAMA)	MMI	Schools	Employer	Private insurance	NGOs	None	Do not know	Total
Mutuelle de Santé	80,389	356	127	17	17	54	24	1,414	74	82,472
RSSB (Ex RAMA)	411	3,198	82	7	9	51	4	1	0	3,763
MMI	88	29	727	0	6	2	5	4	0	861
Schools	13	0	0	8	0	2	0	2	0	25
Employer	2	0	0	0	0	11	0	0	0	13
Private insurance	81	23	13	7	94	269	16	16	11	530
NGOs	31	0	0	0	0	20	22	3	0	76
None	1,742	12	1	1	0	3	2	1,031	20	2,812
Do not know	30	1	0	0	2	1	1	4	1	40
Total	82,787	3,619	950	40	128	413	74	2,475	106	90,592

Source: PES 2022

Match-keys

All records were run through every match-key and conflicts were resolved clerically[9].

Table A.0-6: Within Household Match-keys.

Match Key	Description
1	First Name + Last Name + Year + Sex + Household
2	StdLEV (First Name)>0.79 + StdLEV (Last Name) >0.79 + Year + Month + Household
3	First Name + Last Name + Age_Difference + Sex +Household
4	First Name + Last Name + Relationship to Head of Household + Household
5	StdLEV (First Name) > 0.79 + StdLEV (Last Name) >0.79 + Sex + Household
6	Alphanames + Year + Sex +Household
7	StdLEV (Alphanames) > 0.79 + Sex + Year + Household
8	First Name + StdLEV (Last Name)>0.69 + Sex + Age + Household
9	First Name + (Last Name 1 Cen = Last Name 2 PES or Last Name 2 Cen = Last Name 1 PES + Age + Sex + Household
10	StdLEV (First Name) > 0.79 + RLSoundex[4] (Last Name) + Age_Difference + Sex + Relationship to Head of Household + Household

[1] The standardized Levenshtein Edit Distance that measures the number of insertions, deletions or substitutions required to transform one name into another calculated as a proportion of the length of the longest name.

[2] The age difference function that allows different tolerance in ages depending on the age (see paragraph 3.15

[3] Alphanames is a derived variable that puts all characters of the names into alphabetical order

[4] RLSoundex is the Soundex function applied to the name with all Rs replaced with Ls

[9] to confirm a right match in non-unique matches.

Match-keys – applied to persons in households where at least one person match has already been confirmed. These match-keys were run in a hierarchical order and conflicting matches were disregarded.

Table A.0-7: Within Household Associative

Match-Key	**Description**
1	StdLEV(Alphanames)>0.79 + Age
2	First Name + Last Name 1 + Age Difference +Sex
3	StdLEV(First Name)>0.79 + StdLEV((Last Name 1 CEN, Last Name 2 PES) or(Last Name 2 CEN, Last Name 1 PES)) >0.79 + Age +Sex
4	StdLEV(First Name)>0.79 + StdLEV(Last Name)>0.49 + Age + Sex
5	StdLEV(First Name)>0.79 + Last Name Initial + Age + Sex
6	StdLEV(Last Name)>0.79 + Age + Sex + Relationship to Head of Household
7	StdLEV(First Name)>049 +RLSoundex(Last Name) + Sex + Relationship to Head of Household
8	Age + Sex + Relationship to Head of Household + Marital Status

Table A8: Within EA Match-keys. All unmatched PES and Census records were run through every matchkey and conflicts were resolved clerically.

Table A.0-8: Within EA Match-keys

MatchKey	**Description**
1	First Name + Last Name + Year + Sex + EA
2	StdLEV(First Name)>0.79 + StdLEV(Last Name) >0.79 + Year + Month + EA
3	First Name + Last Name + Age_Difference + Sex + EA
4	First Name + Last Name + Relationship to Head of Household + EA
5	StdLEV(First Name)>0.79 + StdLEV(Last Name) >0.79 + Sex + Head of Household + Relationship to Head of Household + EA
6	StdLEV(First Name)>0.79 + StdLEV(Last Name) >0.79 + Sex + Age Difference + EA
7	Alphanames + Year + Sex + EA
8	StdLEV(Alphanames)>0.79 + Sex + Year + EA
9	First Name + StdLEV(Last Name)>0.69 + Sex + Age + EA
10	First Name + ((Last Name 1 Cen = Last Name 2 PES) or (Last Name 2 Cen = Last Name 1)) PES + Age + Sex + EA
11	StdLEV(First Name)>0.79 + RLSoundex(Last Name) + Age_Difference + Sex + Relationship to Head of Household + EA

Table A9: Within EA Associative Matchkeys – applied to persons in households where at least one-person match has already been confirmed. These matchkeys were run in a hierarchical order and conflicting matches were disregarded.

Table A.0-6: Within EA Associative

MatchKey	Description
1	StdLEV(Alphanames)>0.79 + Age
2	First Name + Last Name 1 + Age Difference +Sex
3	StdLEV(First Name)>0.79 + StdLEV((Last Name 1 CEN, Last Name 2 PES) or(Last Name 2 CEN, Last Name 1 PES)) >0.79 + Age +Sex
4	StdLEV(First Name)>0.79 + StdLEV(Last Name)>0.49 + Age + Sex
5	StdLEV(First Name)>0.79 + Last Name Initial + Age + Sex
6	StdLEV(Last Name)>0.79 + Age + Sex + Relationship to Head of Household
7	StdLEV(First Name)>049 +RLSoundex(Last Name) + Sex + Relationship to Head of Household
8	Age + Sex + Relationship to Head of Household + Marital Status

Table A10: Within District Matchkeys. All unmatched PES and Census records were run through every matchkey. However, matches were only accepted if there were two or more person matches in the same households and the match had no conflicts. In addition, matches made on the first matchkey only were accepted if they were unique. Clerical resolution was not used to resolve conflicts at this stage as there were too many conflicts even when the two matched persons rule was applied.

Table A.0-7: Within District Match-keys

Match Key	**Description**
1	First Name + Last Name + Year + Sex + Distict
2	StdLEV(First Name)>0.79 + StdLEV(Last Name) >0.79 + Year + Month + Sex + District
3	First Name + Last Name + Age_Difference + Sex + District
4	First Name + Last Name + Relationship to Head of Household + District
5	StdLEV(First Name)>0.79 + StdLEV(Last Name) >0.79 + Sex + Head of Household + Relationship to Head of Household + District
6	StdLEV(First Name)>0.79 + StdLEV(Last Name) >0.79 + Sex + Age + District
7	Alphanames + Year + Sex + District
8	StdLEV(Alphanames)>0.79 + Sex + Year + District
9	First Name + StdLEV(Last Name)>0.69 + Sex + Age + District
10	First Name + ((Last Name 1 Cen = Last Name 2 PES) or (Last Name 2 Cen = Last Name 1)) PES + Age + Sex + EA
11	StdLEV(First Name)>0.79 + RLSoundex(Last Name) + Age + Sex + Relationship to Head of Household + District

Table A11: Within Country Matchkeys. All unmatched PES and Census records were run through a single strict matchkey with no geography matching required. Only unique matches were accepted. Pairwise clerical matching was used at this stage to confirm true matches status.

Table A.0-8: Within Country Matchkeys

MatchKey	**Description**
1	First Name + Last Name + Year + Sex
2	(First Name + Last Name)>0.80 + Year + Sex

REPUBLIC OF RWANDA

MINISTRY OF FINANCE AND ECONOMIC PLANNING

NATIONAL INSTITUTE OF STATISTICS OF RWANDA

P.O. Box 6139 Kigali ***Tel : +250-788383103***

Hotline: 4321 ***E-mail: info@statistics.gov.rw***

POST ENUMERATION SURVEY

16 – 30 SEPTEMBER 2022

Legal Basis: **Law n° 45/2013 of 16/06/2013 on the organisation of statistical activities in Rwanda.**

2022 GENERAL POPULATION AND HOUSING CENSUS

POST ENUMERATION SURVEY (PES)

SECTION L: LOCALISATION AND IDENTIFICATION OF HOUSEHOLD		
L01.	PROVINCE/KIGALI CITY:	|__|
L02.	DISTRICT:	|__|
L03.	SECTOR:	|__|__|
L04.	CELL:	|__|__|
L05.	VILLAGE:	|__|__|
L06.	ENUMERATION AREA (N0 EA):	|__|__|
L07.	AREA OF RESIDENCE(1.URBAN 2.RURAL) :	|__|
L07A.	PES BUILDING NUMBER:	|__|__|__|
L07B.	PES HOUSEHOLD NUMBER:	|__|__|__|
L08.	CENSUS BUILDING NUMBER:	|__|__|__|
L09.	CENSUS HOUSEHOLD NUMBER:	|__|__|__|
L11.	GPS COORDINATES:	Latitude: |__|__|__|__|__|__|__|__| Longitude: |__|__|__|__|__|__|__|__|
L12.	DISTANCE:	|__|__|__|

My names are ....., I work for the National Institute of Statistics of Rwanda as the enumerator of the Post Enumeration Survey. The objective of the survey is evaluating the quality of the 5th Population and Housing Census. I wish to talk with the head of the household. In general, the interview will last 10 min. All provided answers will be kept confidential. I hope that you accept the interview, as your responses are very important.

L13.	CONSENT:	1. Interview accepted =**> L15** 2. Interview is not done |___|
L14.	THE REASON OF NO INTERVIEW:	1.Uninhabited dwelling =**>End** 2. Dwelling turned into business building=**> End interview** |___|

		3.Dwelling destroyed =**>End interview** 4.Refused 5.All residents are absents during the whole period of enumeration
L15.	HEAD OF HOUSEHOLD PHONE NUMBER	|___|

FOR ALL MEMBERS OF HOUSEHOLD

P01: Were your household usually living here on census night (15/08/2022)?

1.Yes => **P01C (Start by making a list of HH members)**

2.No

P01A: Prior to come here were your household residing in Rwanda or abroad?

1.In Rwanda =>**P01B**

2. New Household =>**P01C**

3.Abroad =>**End interview**

P01B: Where were your household residing previously?

(SELECT DISTRICT, SECTOR, CELL, VILLAGE) FROM THE LIST)

P01C: Serial Number of the person

P01D: Surname of .....

P01E: Other names of the .....

P01F: Is [NAME] usual resident or was a visitor last night?

1. Usual Resident
2. Visitor => **GO TO THE NEXT PERSON**

P02: What is [NAME]'s relationship to the Head of Household?

01.Household head
02.Spouse
03.Son or daughter
04.Adoptive child
05.Father/ Mother
06.Father-in-law/Mother-in-law
07.Bother-in-law/Sister-in-law
08. Brother/Sister
09. Grand Child
10. Son/Daughter-in-law
11. Other relative
12. House help
13. Non-relative
14. Unknown relationship

P03: What is [NAME]'s Sex? 1.Male 2.Female

P04: How old was [NAME] at his/her Last Birthday?

NOTE: RECORD AGE IN COMPLETED YEARS

P05AA: On which day was [NAME] born?

NOTE: RECORD 99, IF THE DAY OF BIRTH IS UNKNOWN

P05A: In which month was [NAME] born?

NOTE: RECORD 99, IF THE MONTH IS UNKNOWN

P05B: In which year was [NAME] born?

NOTE: RECORD 9999, IF THE YEAR IS UNKNOWN

P06: What is [NAME]'s marital status?

(ALL PERSONS AGED 12 YEARS AND ABOVE)

P01G: Did [NAME] sleep in this household last night?

1. Present Resident (PR)
2. Absent Resident (AR)

1.Married to one wife/husband officially

2.Married to one wife/husband not officially

3.Live in a polygamous union

4.Divorced

5. Separated

6. Never married

7. Widowed

RESIDENCE ON CENSUS DAY

P07R: Was (Name) a usual resident of this household at census night (15-16/08/2022)?

1. Yes **=>P14**

2. No

P09: In which District was [NAME] residing previously?

(SELECT DISTRICT, SECTOR, CELL, VILLAGE) FROM THE LIST)

P08: Why (Name) was not a usual resident of this household at census night?

1. Usual resident of another household in Rwanda **=>P09**
2. Coming from a long residence abroad (>6 month for non-student) **=>P14**
3. Coming from a long stay in prison etc. (> 6 month) **=>P14**
4. Born after census **=>P14**

P14: What is [NAME]'s Medical insurance?

1.Mutuelle

2.RSSB (former RAMA)

3.MMI

4.Schools

5. Employer

6. Private insurance companies

7. NGOs

8. None

9. Do not know

FOR THOSE PERSONS WHO HAVE MOVED OR DIED AFTER THE CENSUS (OUT MOVERS)

O01 **Are there any persons who were usually residents of the household at census time who are no longer usual residents of this household (including those who died after census)**

1. Yes
2. No => **End interview**

S/N	O02A: SURNAMES OF OUT-MOVERS (O-01)	O02B: OTHER NAMES OF OUT-MOVERS (O-01)	O03: What is [NAME]'s relationship to the Head of Household?	O04: What is [NAME]'s Sex?	O05: How old was [NAME] at his/her Last Birthday?	O06: In which month was [NAME] born? NOTE: RECORD 99, IF THE MONTH IS UNKNOWN	O06A: In which year was [NAME] born? RECORD 9999, IF THE YEAR IS UNKNOWN	O07: What is [NAME]'s marital status?	O08: What is the reason [NAME] is no longer a household member?
			01.Household head 02.Spouse 03.Son or daughter 04.Adoptive child 05.Father/ Mother 06.Father-in-law/Mother-in-law 07.Bother-in-law/Sister-in-law 08. Brother/Sister 09. Grand Child 10. Son/Daughter-in-law 11. Other relative 12. House help 13. Non-relative 14. Unknown relationship	01.Male 02.Female				1.Married to one wife/husband officially 2.Married to one wife/husband not officially 3.Live in a polygamous union 4.Divorced 5. Separated 6. Never married 7. Widowed	1. Died after Census 2. Move to another household 3. Move outside the country 4. Went in Institutional Household 5. Others/Specify
1	.....	.....							
2	.....	.....							

PERSONS WHO CONTRIBUTED TO THE POST ENUMERATION SURVEY

National Coordinators

MURANGWA Yusuf

MURENZI Ivan

National Technical Coordinators

NYIRIMANZI J. Claude

UWIMANA Therese

MUHOZA Didier

ONS Consultants for Data Matching and Analysis

Dr Rachel Shipsey

Anthony Edwards

Charlie Tomlin

PES Data Matching:

UWAMAHORO Sandrine

HAGENIMANA J. Damascene

UWIMBABAZI Denyse

IRANZI Orodha

NAHIMANA Samuel

Data Processing:

HARERIMANA Massoud

NIYONGIRA Patrick

NSHIMIYIMANA Clement

NDAYIRAGIJE Bosco

Supervisors:

NTAMBARA Juvenal

KAMBOGO Francois

MUGENZI Gilbert

NTAGENGERWA Bonus

AYINGENEYE Seraphine

GAGA RUKORERA Didier

Data Analysis :

NTAGENGERWA Bonus

MUHOZA Didier

GAGA RUKORERA Didier

monthsCartography:

BIZIMUNGU Clement

MUNDERERE Theophile