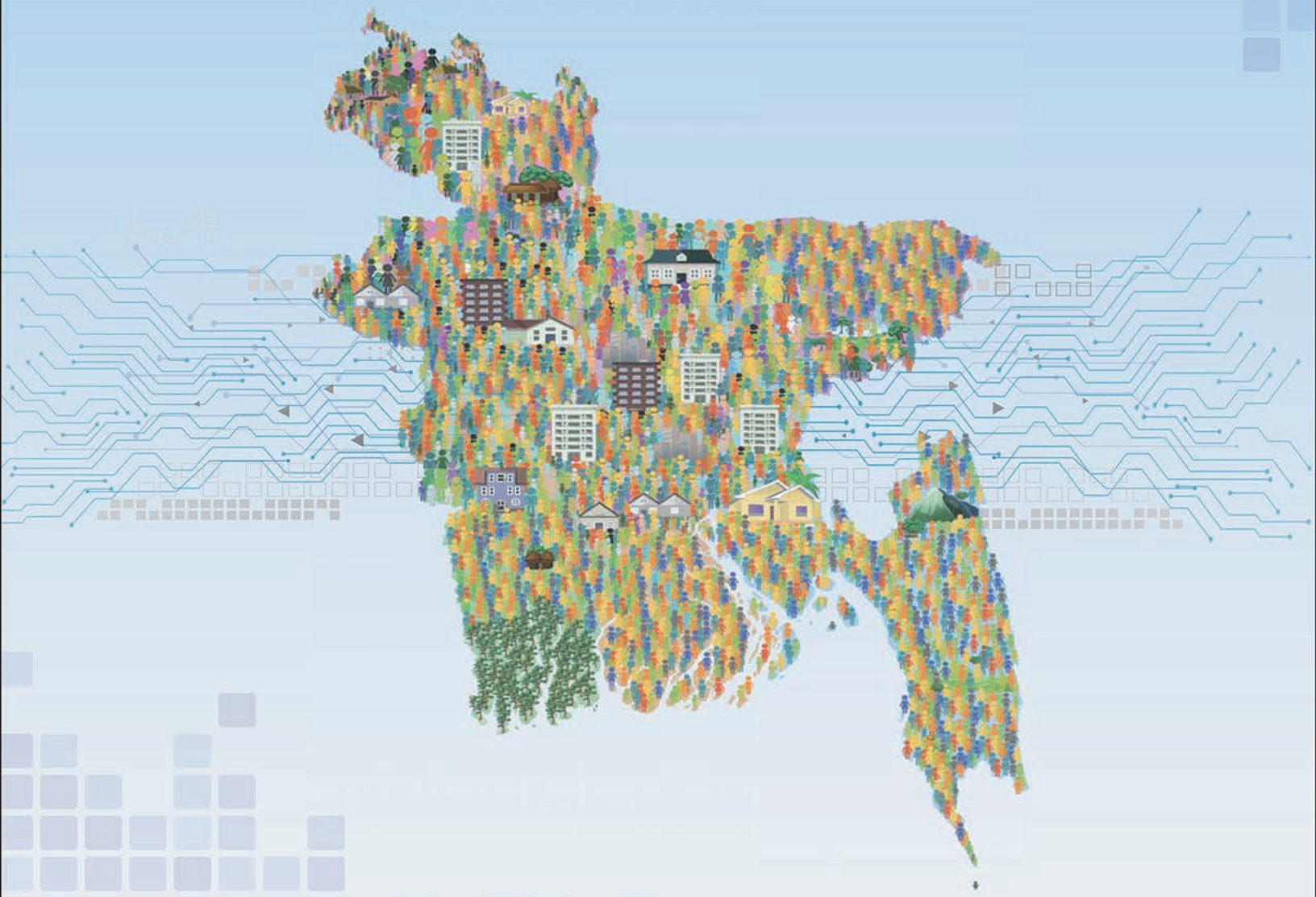
Report on
Post Enumeration Check of the Population and Housing Census, 2022



Bangladesh Institute of Development Studies (BIDS)
E-17, Agargaon, Sher-e-Bangla Nagar, Dhaka 1207

Report on

Post Enumeration Check of the Population and Housing Census, 2022

Binayak Sen
Mohammad Yunus
S M Zahedul Islam Chowdhury

January 2023

Bangladesh Institute of Development Studies
E-17 Agargaon, Sher-e-Bangla Nagar, Dhaka-1207

ISBN: 978-984-94917-3-6 COMPLIMENTARY

Director General

Bangladesh Institute of Development Studies

Foreword

The Bangladesh Institute of Development Studies (BIDS) was requested by the Bangladesh Bureau of Statistics (BBS) to conduct the Post Enumeration Check (PEC) of the Population and Housing Census 2022. Earlier, BIDS had carried out the PEC for the Population and Housing Census, 2011, and the Census of Undocumented Myanmar Nationals Staying in Bangladesh, 2016, on behalf of the BBS.

Following a G-2-G contract, BIDS conducted the PEC in October 2022 to provide information on the coverage and content errors of the Population and Housing Census 2022. The PEC was implemented independently of the BBS as all the enumerators were recruited independently and trained accordingly. This was important to maintain the operational and conceptual independence of BIDS as a neutral fact-checking entity. A stratified random sampling technique was used in selecting the enumeration areas (EAs) for re-interviewing the population, households, and dwelling units to provide quantitative information on census accuracy. The PEC survey was conducted in the sample EAs independently through directly recruited field enumerators and supervisors. Adequate emphasis was placed on extensive training of enumerators, supervisors, and concerned BBS field officials, proper supervision of fieldwork, and use of digital devices followed by effective matching operations to minimize non-sampling errors.

The results indicate a high coverage rate of the PHC 2022. It gives confidence in the census data. We hope that the results of the PEC will be useful to the government, data analysts, and other users of the census data.

On behalf of BIDS, I would like to extend my deep appreciation to BBS for making the necessary financial resources and digital technology available for undertaking the PEC. In particular, I extend my sincere appreciation to the Secretary of the Statistics and Informatics Division, the Director General of BBS, and the Project Director of Population and Housing Census 2022 for their wholehearted cooperation that made the PEC exercise successful. I would also like to thank the esteemed members of the advisory committee formed for this particular purpose, who guided us through this challenging exercise.

Finally, I would like to express my deep gratitude to Mr. M. A. Mannan MP, Hon'ble Minister, Ministry of Planning, Government of the People's Republic of Bangladesh, and Dr. Shamsul Alam, Hon'ble State Minister, Ministry of Planning, Government of the People's Republic of Bangladesh, for their constant encouragement in conducting the PEC.

Binayak Sen

Dhaka
January 2023

Director General

Bangladesh Bureau of Statistics

MESSAGE

The Bangladesh Bureau of Statistics (BBS), the national statistical organization, usually conducts three censuses, namely, Population and Housing Census, Economic Census, and Agriculture Census. The Population and Housing Census (PHC) is the most significant and gigantic activity among these undertakings. BBS completed the sixth and the first of its kind digital Population and Housing Census during 15-21 June 2022. The successful implementation of this digital census can be regarded as a landmark achievement of the BBS. The preliminary report on PHC 2022 was published on July 27, 2022, highlighting key statistics, including the country's total population, household, population growth rate, density, etc.

A Post Enumeration Check (PEC) was conducted after every census in Bangladesh to precisely present the census outcomes by following the UN guidelines and principles. The PEC involves a complete re-enumeration of a representative sample of a census population followed by matching each responding unit enumerated in the PEC with information obtained from the census enumeration for the dual system of estimation, independence between the PEC and the census is a prime requirement. In view of this, BBS decided to give the responsibility of conducting the PEC of the Population and Housing Census 2022 to an independent organization, Bangladesh Institute of Development Studies (BIDS), which is an autonomous professional body with experience in successfully conducting the PECs of the Population Census 2011, Economic Census 2013, and Census of the Undocumented Myanmar Nationals Staying in Bangladesh, 2016.

I am happy to note that the report on the PEC is going to publish by the BIDS. The results of the PEC are mainly used to measure coverage and content errors of the census data.

I express my heartfelt gratitude to Mr. M. A. Mannan MP, Hon'ble Minister, Ministry of Planning, and Dr. Shamsul Alam, Hon'ble State Minister, Ministry of Planning, for their guidance throughout the census operations and the PEC.

I express my profound gratitude to Dr. Shahnaz Arefin, *ndc*, Secretary, Statistics and Informatics Division, for her wholehearted cooperation and valuable inputs that made the PEC a success. I greatly appreciate the Director General, BIDS, and other participating researchers of BIDS for their sincere efforts in conducting the PEC and publishing this report. Last but not least, I extend my sincere thanks to all concerned officials of BBS who extended their cooperation in the smooth implementation of the PEC.

Md. Matiar Rahman

Dhaka
January 2023

CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The Bangladesh Institute of Development Studies (BIDS) conducted the Post Enumeration Check (PEC) in October 2022 to evaluate the coverage and content errors of the Population and Housing Census 2022. For the PEC, 326 enumeration areas (EAs) were drawn, taking 2 EAs from each of the 163 District Census Coordinator areas through stratified random sampling. Additionally, 28 EAs were selected. Of these, 14 EAs were drawn, taking 2 EAs from each of the seven haor districts; 12 EAs were drawn, taking one from each of the 12 city corporations; and 2 EAs were drawn from two coastal districts. Even though the PEC enumerators were tasked with collecting data from the 354 sample EAs, additional data were also collected from 24 adjacent EAs in those cases. Thus, data were collected from 378 EAs to ensure the credibility of the PEC estimates.

To ensure operational independence, 354 highly experienced enumerators and 36 supervisors were directly recruited, who underwent a six-day rigorous training conducted by the research team both on the concepts based on the printed questionnaire and the use of the digital device for collecting and uploading the data to central server of the Bangladesh Bureau of Statistics (BBS). Besides, all of the district-level officers of the BBS were trained and persuaded to coordinate the data collection activities in their respective districts through the upazila statistical officers.

The PEC fieldwork is split into three phases: publicity, listing, and enumeration. The supervisors and upazila statistical officers closely supervised the enumerators' work to minimize non-sampling errors. The core team members were dedicated to monitoring the field activities and visited several EAs along with officials from the Statistics and Information Division (SID) and BBS. Above all, the progress and quality of the data collection were digitally monitored at the BBS Head Office.

While most of the households in respective EAs were matched digitally by the enumerators, an additional digital matching exercise was undertaken after data collection to investigate whether the PEC persons/households were enumerated during the PHC. It was done by comparing the address listings in both the PHC and PEC to identify the corresponding households. Where this was inconclusive, the questionnaires were compared to see if a match could be found based on names and household structure. In this process, a large number of hitherto non-matched households were matched.

Taking the field-level household matching by the enumerators and computerized matching between the PHC and the PEC, the following steps were followed for each of the matched households: (i) Exact Matching: In this step, only a maximum of 20% of the variations in the characters of the names were allowed in both the PHC and PEC samples; (ii) Gender and Relationship with Household Head: Among the non-matched members in step (i) in both PHC and PEC, members were matched based on their gender and relationship with the household head; (iii) Gender and Age: Among the non-matched members in step (ii) in both PHC and the PEC, members were matched based on the gender that minimizes the age difference for each of them between PHC and the PEC

records; and (iv) Fuzzy Matching of Name and Age: Among the non-matched members in step; (iii) variation of names was minimized such that the age difference of matched observation from both the sources follows a specific age limit of 10 years. Non-matched records were then reconciled through field visits to identify erroneous inclusions.

Coverage Errors

Following the standard procedure, the dual system of estimation was used to evaluate the PHC coverage. The national net coverage error rate, weighted by population share, is estimated at 2.75%. The net coverage error rate is lower in rural areas (2.55%) than in urban areas (3.19%). Within urban areas, the net coverage error rates increase with the level and sophistication of urbanization. It is also noted that the standard errors in urban areas are higher than in rural areas. The rural coverage error rate is subject to a 16% relative error rate in contrast to 34% in the upazila sadar and growth center. Together with the higher standard error, the higher coverage error leads to a higher coefficient of variation.

The gender-specific coverage error rate analysis shows that the net undercount rate is higher at 2.81% for males compared to 2.69% for females. Besides, both the standard error and the coefficient of variation of the coverage error are lower for females. The coverage error rates by religion show that the extent of undercount appears to be higher for Muslims at 2.76% compared to the non-Muslims at 2.67%. However, the coverage error rate estimate for the non-Muslim population appears to be imprecise.

The analysis of demographic coverage error rates shows that the net undercount rates for age groups follow a U-shaped pattern. The successive rates secularly decrease until the minimum is reached at the 40-44 age group, and then, the trend is secularly upward. Both the lower and the higher age groups have higher rates; while the coverage error rate for the highest age group (75+ years) is estimated at 4.82%, the same is estimated at 4.23% for the 0-4 age group. In contrast, persons from the 30-34 age group to the 55-59 have coverage error rates of less than 2%. Despite the wide variations in the coverage error rates across age groups, many of the standard errors are 'high,' which tends to make the estimates imprecise.

It may be noted that the coverage error rates are lower in the western part of the country. Among the administrative divisions, the net coverage error rates in Khulna, Rajshahi, and Rangpur divisions are lower than the national rate. However, the overall rates across divisions mask the variations across rural and urban areas. Even in the western part of the country with lower overall coverage error rates, there are particular locations where net coverage error rates are higher than the national average rate.

The net coverage error rates provide a basis for adjustment of the census count of the population. Even though the corresponding adjustment factors are derived, it is up to

BBS to make any adjustment in consultation with the highest policy-making authority. If an adjustment is to be made, consideration needs to be given to what domains are important in the decision to adjust the PHC counts.

Content Errors

For measuring response variabilities between the PHC and the PEC, the net difference rate, index of inconsistency, aggregate index of inconsistency, rate of agreement, and gross difference rates were used. The variables under consideration were gender, relationship with the household head, marital status, religion, and age group.

The net difference rates (NDRs) range between 0.15% and -0.15% for females and males, respectively, and the corresponding ratios of |NDR|/P are zero for both females and males. The 'index of inconsistency' for both gender and the 'aggregated index of inconsistency' is only 3%. In harmony with this finding, the 'rate of agreement' is as high as 98.50%. The low values of content error measures suggest that the inconsistency of gender reporting between the PEC and the PEC is extremely low. These results are not surprising as gender is expected to be reported more or less reliably and consistently both in the PHC and the PEC.

The estimates of NDRs reveal mismatches but cannot point out the directions as the relationships are categorical variables. But the estimates of the |NDR|/P provide the severity of misreporting between the PHC and the PEC. While relationships termed the head of the household, spouse, and children appear to be reported correctly, there is a moderate degree of misreporting of parents, sons/daughters-in-law, siblings, and other relatives and non-relatives have a high degree of misreporting between the PHC and PEC. Given the preponderance of the household head, spouse, and children among the members of the households, the aggregate index of inconsistency appears to be low at 8.73%, which is corroborated by the rate of agreement at 93.51%. The low levels of misreporting are acceptable in that with the change in the respondent between the PHC and the PEC, the household head is likely to change, and with that change the members' relationships.

The estimates of the NDRs imply that the married groups of persons appear to be over-reported and the other groups under-reported in the PHC. A closer scrutiny through the |NDR|/P estimates reveals that the degrees of variation for the married groups are low but very high for the other groups. Given the high proportion of the married groups in the data, the aggregate index of inconsistency of 'marital status' shows a low level at 9.37%, with a rate of agreement at 94.96%. Thus, marital status is likely to be reported more or less reliably and consistently in the PHC and the PEC.

The estimates of the NDRs reveal that Muslims and 'other minority' religions are over-reported, while Hindu, Buddhist, and Christian are under-reported. However, a closer look through the |NDR|/P estimates reveals that the degrees of variations are low

for both Muslims and Hindus, moderate for both Buddhists and Christians and very high for the 'other minority.' However, the number of cases in the high category is very few. These findings are corroborated by the aggregate index of inconsistency, which shows a low level of inconsistency or variability (index < 20%) for religion. Thus, it can be concluded that religion is reported, more or less, reliably and consistently in both the PHC and the PEC.

The estimates of the NDR reveal that age has been under-estimated in 5-9 to 20-24 groups, and 45-49, 55-59, 65-69, and 75+ groups in the PHC. It appears to be over-estimated in the rest of the groups. Given these biases, the level of |NDR|/P is remarkably small for most of the age groups. View from indices of inconsistency, nine groups show a high degree (index > 50%), six groups show a moderate degree (20% < index < 50%), and only one group shows a low level of inconsistency or variability (index < 20%). These intergroup mismatches are epitomized in the estimates of a moderate level of the aggregate index of inconsistency (42.29%) and a low rate of agreement (60.10%). Consequently, 'age' may not be reported consistently between the PHC and the PEC**.**

In conclusion, the broad findings of the PEC point towards satisfactory correspondence between the estimate of the true population and the number of units enumerated in the PHC. The estimates of coverage error rates were looked at through different domains such as location, gender, religion, age group, and administrative division. It was found that these estimates were low and well within acceptable limits. For content errors, the findings indicate that the response variability is random in nature, and there is very little systematic bias associated with the reporting of either the group of the PHC enumerators or the group of PEC enumerators.

Lessons Learned

Several issues may be considered seriously for a successful PEC in the future PHC. Some of these include (i) conducting the PEC after extensive planning, testing questionnaire and methodologies, and procedures well in advance and in tandem with the PHC; (ii) completing the fieldwork of the PEC within a short period after the completion of the PHC to avoid a large distortion in the structure of the population enumerated; (iii) maintaining the independence of the PEC from the PHC should be ensured at all stages of PEC implementation; (iv) aligning the operational timelines with the PHC activities to allow for thorough and timely completion of PEC activities; (v) ensuring the in-house GIS capacity of BBS to delineate the boundaries of EAs and monitor both the PHC and the PEC enumerators so they do not transgress the EA boundaries while collecting data; and (vi) ensuring that the name and telephone number of the respondent, together with the GPS coordinates of the dwelling unit, are collected and shared with the organization or the unit that would conduct the PEC that reconciliation can be promptly made.

CHAPTER 1: INTRODUCTION

1.1. Background

Population and housing censuses are the main source of disaggregated statistical data down to the village level, the lowest geographical entity. These data are essential for planning, designing, and monitoring to ensure '*leave no one behind*.' Additionally, these data help develop short- and long-term development plans such as the Eighth Five Year Plan, Delta Plan 2100, Vision 2041, etc. The Bangladesh Bureau of Statistics (BBS) conducted the sixth Population and Housing Census (PHC) during June 15-21, 2022. One of the novelties of this PHC was application of digital technology by eschewing the paper-based data collection method used in the earlier censuses. The place of residence of each of the enumerated persons was assigned in the PHC by following the modified de facto method. Under this approach, besides counting household members in the place where they are at the census moment, members who are on journeys, in hospitals and hotels, or on duty at the census moment are counted in their usual residences. The BBS engaged as many as 365,695 local enumerators across the country to collect the PHC data using digital devices popularly known as *Tablets*.

The PHC yields a wealth of valuable information for analyzing changes in the socio-demographic profile of the population. It is also used for monitoring, planning, and decision-making by national and local governments, businesses, and other stakeholders. Further, it is integral to the derivation of reliable post-census population estimates and for charting future demographic trends at the national and local levels. Given the strategic significance of the PHC data and its diverse applications, the BBS made concerted efforts to ensure universal coverage of the PHC 2022.

1.2. An Overview of Population and Housing Census, 2022

Population and Housing Census 2022 is the maiden digital census of Bangladesh. The most significant characteristics of the first digital census include (i) clearly

identifying the enumeration area based on digital maps developed by integrating Geographic Information System (GIS) and Geo-code; (ii) collecting data with digital device tablets using Computer Assisted Personal Interviewing (CAPI) method; (iii) using the data center with Tier IV Security, in which a multilayer firewall was installed to prevent hacking by any means and to ensure utmost security and proper storage of data; (iv) transmitting data from the field to the server in completely encrypted form to protect the confidentiality of individual's information; (v) using a web-based Integrated Census Management System (ICMS) to manage all activities of the census; (vi) establishing a Network Operations Center (NOC) for real-time monitoring of the progress of data collection in the field and to follow the trends of data across various domains to ensuring data quality; (vii) establishing a Call Center to provide quick solutions to various problems arising at the field level during data collection and to ensure direct interaction between BBS and the people to be included in the enumeration by themselves; and (viii) campaign of the census through both the traditional approaches and online platform including social media.

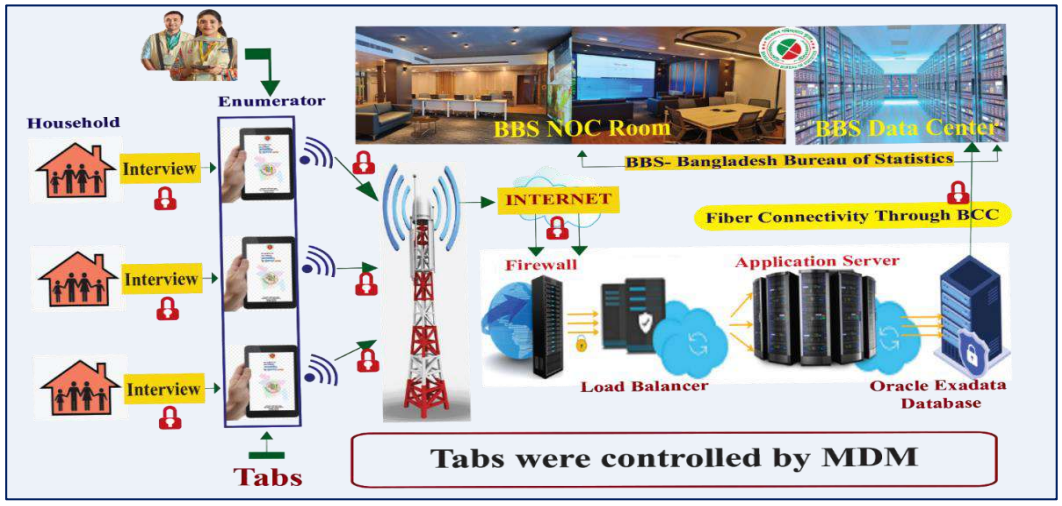
A total of three zonal operations were carried out as part of the census preparations. Under these zonal operations, the formation of enumeration areas, maps' development, updating of geocodes of all administrative units, selection of training centers, preparing the list of protected areas, and making the list of VIPs were done. Additionally, supervisors and enumerators were appointed with the help of the field administration. Integrating the information generated through these zonal operations into ICMS, map integration, etc., was also carried out during these operations.

The concerned officers/employees have been trained for four days in three phases to ensure the quality of census data. In the first phase, the training of Monitoring Officers and Master Trainers (Divisional and District Census Coordinators) was conducted at BBS Head Office. In the second phase, the training of Upazila Census Coordinators, Zonal Officers, and ICT Supervisors was conducted at the district level. In the third phase, enumerators and supervisors were trained at the zone level (Upazila/Municipality/Union/Ward). Training is imparted through audiovisual training modules along with in-person training. Separate interactive animated training Apps were also prepared for each question. The Apps have also been used for the self-learning of enumerators.

The modified de facto approach was followed to conduct the PHC 2022 in Bangladesh. The midnight of 14 June 2022 (the zero moments of June 15) was determined as 'the census reference point/time;' the following week, June 15-21, 2022, was regarded as 'the Census Week.' During this time, the PHC 2022 data were collected across the country using Computer Assisted Personal Interviewing (CAPI) method. However, due to the unpredicted sudden flash flood, the data collection period was extended up to June 28 in the districts of Sylhet, Sunamganj, Moulvibazar, and Netrokona, following the international protocol of census.

The data collected from households were directly sent to the Tier IV Data Center (Bangabandhu Hi-Tech City, Kaliakoir, Gazipur) of Bangladesh Data Center Company Limited (BDCCL) using a mobile network. Later, the data were transmitted to the application server passing through the load balancer and the multilayer firewall installed in the data center through the cloud. Subsequently, the data were transferred from the cloud to the data center of BBS through an optical fiber cable using Oracle Exadata installed in the Tier IV Data Center. Through this system, real-time visualization of data was made possible in the Network Operations Center (NOC) during the data collection at the field. The digital architecture used in the PHC 2022 is shown in Figure 1.1.

Figure 1.1: Flow of Digital PHC Data



Despite the best efforts by the BBS, it is usually accepted that the PHC is not perfect and that errors can and do occur at all stages of census operations.

Consequently, it might have missed some people, which may be non-trivial. It is almost impossible to conduct a perfect census since errors can occur at different points within data collection and processing operations. Therefore, it is important to assess the quality of the PHC, including coverage errors. Incomplete coverage may result from, for example, inadvertent omission of young children, difficulty in enumerating people on the move and those living in apartments in urban areas, as well as people not willing to cooperate with census enumerators. Omission also occurs due to misunderstanding of the question or concept or intentionally not reporting a person living in a household for any particular reason, etc., or the person is reported in the wrong place.

1.3. Objectives of the Post Enumeration Check

Several methods can be used in evaluating the quality of a census. These include demographic analysis, matching studies, non-matching studies, etc. These methods differ widely in terms of the level of sophistication, data requirements, and quality of results. While the demographic analysis requires a single data source (such as data from vital registration surveys), the matching and non-matching studies require multiple data sources. It may also be noted that the demographic analysis and the non-matching studies cannot evaluate the relative magnitude of coverage and content errors in total error (US Bureau of the Census, 1985).

Methods based upon a single source of data provide less insight into the magnitude and types of errors in census data than methods based on the comparison of two or more sources of data. In contrast, methods that depend on two or more data sources are more powerful in assessing the contributions of different types of errors (UN, 2010). Accordingly, a PEC is undertaken after the PHC to evaluate the completeness of census coverage. It involves an independent re-enumeration of a statistically representative sample of households and the persons within households covered by the PHC 2022. The basis of the methodology lies in comparing the re-interviewed persons with those enumerated during the PHC.

Omission or undercount includes the omission of individual persons in enumerated households and the omission of households and, consequently, persons

in those households. Duplication or overcount includes erroneous inclusion of persons in the enumerated households and erroneous inclusion of households and, consequently, persons in those households. Some of the major situations that lead to coverage errors are as follows:

1. There are unconventional dwellings, such as garages, mosques, temples, and school hostels, where people, though not expected to live, live. Census enumerators may consider these dwellings as non-residential, which mainly are residential buildings resulting in the omission of these residential dwellings and, consequently, the omission of the occupants of these residential dwellings.
2. One or more dwellings might be locked during the visit of the census enumerator in the respective EA. Hence, these residential dwellings and, consequently, the occupants dwelling there are omitted during the census and the PEC. These types of situations arise mostly in urban areas, where all the members of the households may be absent during the daytime when the enumerator visits due to various reasons such as work, school, etc.
3. Visitors and domestic servants, who may be living with the household during the entire enumeration period, may be omitted, as the respondent may not be aware that they are also to be enumerated.
4. People deliberately avoid the census by refusing to respond, fearing that the information given will be used against their interests when they mistrust the census confidentiality on personal data.
5. People may be reluctant to open their doors to strangers due to personal safety and security, especially in urban areas.

The PEC survey was conducted during October 10-16, 2022, to assess the coverage and quality of the PHC enumeration. The main goal of the PEC was to evaluate the quality of the PHC data by collecting closely monitored quality data after the completion of the census enumeration. The specific objectives of the PEC survey were as follows:

- To evaluate the accuracy of PHC data by providing quantitative information on coverage and content errors at specified domains of estimation;

- To provide stakeholders with quantitative information to enable the determination of the overall success of the census program; and
- To provide a statistical basis for adjustments of census data across domains if and when it becomes necessary.

The above specific objectives lead to the point that the PEC responds to the needs of a variety of stakeholders, including the unit record *data users, socio-economic planners,* and *decision-makers,* by helping them make judicious interpretations and use of census results given the PEC results of coverage error. Besides, measuring content error across a few critical domains helps *planners* of future censuses to improve the design and implementation of future censuses and large-scale surveys.

Against this backdrop, the PEC is intended to provide possible answers to the following two questions: (i) how accurately have the persons been enumerated (*coverage error*) in the PHC 2022? (ii) how precisely have certain characteristics of the individuals been recorded (content error) in the PHC 2022? To that end, the primary objective of the PEC is to estimate the magnitude of omissions (undercount) and duplications (overcount) of individuals in the PHC 2022, or in other words, to determine the *coverage error*. The coverage error investigated in the PEC consists of two components: (a) omission or duplication of persons due to omission or duplication of households and (b) omission or duplication of individuals in enumerated households.

The errors in response, or *content errors* in the recorded characteristics, were also assessed while conducting the PEC. Some of the individual-level characteristics include age, gender, religion, marital status, relationship with the household head, etc. The content error arises from how the enumerator explains the question(s), how the respondents understand the question(s), or both. Further, in mass operations like a census, it is not always that the concerned individual provides the data. Errors may also arise because the respondent may not know all the particulars of a person about whom the information is being reported in the census. For example, the head of the

household may not know her/his daughter-in-law's correct age or place of birth or some other person who stays with them at the time of the census. Errors can also occur because of the difficulty in understanding the concepts. For example, anyone may give their running age instead of the completed age, which is required.

1.4. Organization of the Report

This report describes and discusses the salient features of the PEC of PHC 2022, including its scope, methodology, the information gathered, and the results of the operations. The report is organized as follows. Following the introduction, Chapter 2 describes the issues related to data collection. Chapter 3 evaluates the dual system estimates of population and the consequent coverage errors, while Chapter 4 presents the estimates of content errors. Chapter 5 provides the summary and conclusions of the report.

CHAPTER 2: FRAMEWORK OF ANALYSIS

Large-scale data collection, such as through the PHC, is often prone to various types of non-sampling errors. Thus, an independent PEC is a standard procedure to estimate the nature and extent of the errors. This chapter describes the methodology and the assumptions used in conducting the independent PEC. Also, it provides a summary of the processes followed in designing the sample for the PEC and the assumptions used in developing PEC methodology documents and instruments.

2.1. Assumptions of the Dual System of Estimation

Methods that depend on two or more data sources are more powerful in assessing the contributions of different types of errors. The dual system estimation uses two independent sources or 'systems,' e.g., the PHC and the PEC, to estimate the *true population*. The dual system provides an estimate of the cases included in one source (PHC) and excluded from the other (PEC), and vice versa, as well as the count of those that were enumerated in both sources. It also allows for the computation of the number and the rate of persons or households missed by both the PHC and PEC using the principle of independence and probability methods. Thus, it is evident that both the PHC and PEC estimates contribute to the dual system estimate, which is more complete than either the PHC or the PEC estimate alone (UN, 2010). In the end, this *true population* is compared with the *population enumerated in the census,* and the difference is the net *undercount* **(***overcount)* when it is positive (negative). It may be noted that the accuracy of the dual system hinges on the following assumptions:

a) *Closed Population***:** Migration between the time of the PHC and the PEC is insignificant; hence, the composition of the population remains relatively unchanged.

b) *Independence between PHC and PEC*: The organization of the PHC and PEC, especially fieldwork operations, must be managed by a separate and independent organization.

c) *No Erroneous Inclusions*: The totals of the PHC population and the totals of the PEC population would be free from erroneous inclusions. If there are erroneous inclusions in either set, these persons need to be identified and removed from the totals.

d) *No Incomplete Match***:** The dual system estimation does not take into account cases in the PHC total and the PEC total that could never be found in the matched population.

Therefore, any failure to match PHC and PEC items should be due to actual omission and not the inability to match.

It may also be noted that the dual system estimates are subject to three types of biases: non-response bias (refusal, non-contacts), correlation bias (a higher probability of non-inclusion in PHC leads to a higher probability of non-inclusion in the PEC), and matching bias (erroneous matches and erroneous non-matches).

2.2. Sample Design of the PEC

The quality of the PHC data is very important not only for understanding the national statistical system and using the census frame for the subsequent national surveys for various purposes but also for building public trust in its accuracy. As a result, the PEC is needed to be conducted to assess the accuracy of coverage and selected contents. With these ends in view, the PEC attempts to estimate the total number of persons and households at the time of the census. The units of observation are the persons who spent the census night and/or the PEC night in their respective dwellings.

It may be recalled that the 64 administrative districts of the country were divided into 163 broad units for the PHC. Each of these units, consisting of several primary sampling units known as enumeration areas (EAs), was placed under a District Census Coordinator (DCC). In the initial draw, two EAs from each of the 163 DCCs were randomly selected, making the sample size 326 EAs. In addition, the urban areas were over-sampled to build on public trust and confidence in the PHC outcomes. Because of the high density of the population, entry barriers, especially to gated communities, confusion over the EAs, etc., it was decided to take the extra samples

from city corporations. Moreover, two additional samples were drawn from each of the flood-affected districts of the Meghna Basin. So, 12 additional EAs were drawn from the 12 city corporations and 14 EAs from the flood-affected Haor districts through stratified random sampling. Besides, two EAs were drawn from coastal districts with poor accessibility through simple random sampling after binning out the EAs selected in the initial draw. Eventually, 354 EAs were selected through these sampling techniques.

Even though the PEC enumerators were tasked with collecting data from the 354 sample EAs, data were collected from 378 EAs as some of the rural EAs had too few households. Accordingly, data were collected from 24 adjacent EAs in addition. This post-facto decision was taken to get higher accuracy in those EAs. Table 2.1 presents the distribution of the EAs by location across the administrative divisions, and Figure B.1 in Appendix B shows the spatial locations of the EAs across the country.

Table 2.1: Ex-Post Distribution of the Allocation of EAs by Stratification Factors

Division	Rural	Municipality	Upazila Sadar/ Growth Center	City Corporation	Total EAs (Count)	Share in Total EAs (%)
Barishal	17	3	3	1	24	6.34
Chattogram	40	7	15	8	70	18.52
Dhaka	55	10	2	40	107	28.30
Khulna	32	3	3	2	40	10.58
Mymensingh	23	3	-	2	28	7.40
Rajshahi	32	4	3	3	42	11.12
Rangpur	27	3	2	3	35	9.26
Sylhet	26	2	1	3	32	8.46
Bangladesh	252	35	29	62	378	100.00

Note: The computer-assisted stratified random sampling draw missed sample EA under upazila sadar and growth centers from Mymensingh Division. However, the sampling process missed the population that accounts for 5.2% of the total population in that category. Accordingly, this population was excluded when the weight was calculated and reassigned for each division-location in calculating of the net coverage errors.

The sample of EAs is large enough to provide estimates of census coverage at the national level and across other domains with reasonably low sampling errors. It needs to be emphasized here that the PEC can only generate reliable and accurate results if the sample is well-designed, its implementation is efficiently managed, matching operations are meticulously done, and the data analysis and estimations

are correctly executed. To minimize non-sampling errors, adequate emphasis was placed on extensive training of enumerators/supervisors, supervision of fieldwork, and the development of multi-step effective matching operations with different types of computer algorithms.

2.3. Questionnaire Development

The PEC questionnaire has been prepared from the PHC questionnaire to reflect the coverage and content errors of the census properly. Indeed, the PEC questionnaire is the same as the PHC questionnaire, with a few additional questions and conditionalities. The questions cover both the dwelling and individual modules with basic demographic information (age, gender, marital status, literacy, housing, etc.). The approach to questionnaire design focuses on re-capturing the main elements for measuring coverage and content.

The questionnaire was developed following *Procedure C*,[1] which is commonly used for evaluating census results (Dauphin & Canamucio, 1993; UN, 2010). Accordingly, the PEC questionnaire has the provision for the classification of each listed person in a household as non-mover, out-mover, in-mover, or out of scope, i.e., died before PHC, born after PHC, etc. These population parameters have been defined as follows:

> *Non-mover* refers to a person who resided in the dwelling on the reference date of the PHC (14 June) and still living at the reference date of the PEC (9 October). In other words, the person was present on both census and PEC nights.
>
> *Out-mover* refers to a person who lived in the dwelling on the PHC date but did not live in the dwelling on the PEC date. Such a person left the dwelling before the PEC night.

[1] In accordance with *Procedure C*, the data collection identified all persons in the household at the time of the PHC as well as those at the time of the PEC. Under the *modified de facto* enumeration, this means the persons who spent the reference nights in the household. Therefore, every person, young or old, including babies, the elderly, visitors and non-citizens, who were present in the household on either or both reference night should be enumerated.

In-mover refers to a person who lived in the dwelling on the PEC date but did not live in the dwelling on the PHC date. Such a person arrived in the dwelling after the PHC night.

Out-of-scope refers to a person who does not belong to the target population of the PHC date. For example, a child born after the census date and/or a de facto census person who lived outside the country on the census date but lived in the dwelling on the PEC date.

2.4. Fieldwork

As the PEC mimics the PHC in the sample EAs, most of the methodologies and procedures for data collection are based on PHC methodologies and procedures. Extra measures are put in place to ensure that the PEC is conducted as a more complete audit of the PHC. For example, extensive probing is conducted to correctly identify and classify all structures and dwelling units, including the number of dwelling units and the number of persons per dwelling unit. For this purpose, the PEC fieldwork is split into three phases: publicity, listing, and enumeration. Publicity and listing per EA are conducted at the same time. Publicity focused on informing through well-articulated leaflets and educating respondents and relevant stakeholders about the purpose of the PEC to ensure successful coverage of all dwelling units in selected EAs. Listing involved recording all structures (including all dwelling units, number of households in dwelling units, and number of persons in households) in the sample EAs. Enumeration involved interviewing respondents and recording digital responses in the fields provided in the PEC questionnaire.

2.4.1 Planning for and Execution of Data Collection

A well-thought-off and realistic work plan is essential for ensuring the timely completion of fieldwork and getting accurate data. In this regard, therefore, the priority is to prepare a work plan before the actual fieldwork is started. While preparing the work plan, the following major considerations are kept in mind: (i) identifying the respective EAs through printed EA maps together with the help from the BBS field staff, (ii) establishing rapport in the respective EA; (iii) timely and smooth completion of the fieldwork; and (iv) close supervision and monitoring of field works.

BIDS recruited as many as 354 enumerators and 36 experienced supervisors from a large pool of applicants to digitally collect household- and individual-level data and monitor the fieldwork. These enumerators and supervisors underwent a six-day rigorous training by the research team on the paper-based questionnaire and the use of the questionnaire on the *Tablet*. Besides, all of the district-level officers of the BBS were trained by the research team and persuaded to coordinate the data collection activities in their respective districts through the respective upazila statistical officers. Specifically, the field officers of the BBS were tasked with facilitating the delineation of the respective EAs to ensure the complete overlap between the PHC and PEC enumeration activities.

2.4.2 Adherence to the Principle of Independence

The principle of independence between the evaluated domain and the evaluation tools is a basic requirement in any evaluation process. For the PEC to achieve its objectives, its processes need to be independent of the census. The following measures were taken to maintain the operational independence of the PEC:

- Made independent drawings of EAs for the PEC sample;
- Used separate/independent field staff in the PEC;
- Conducted the PEC after completion of PHC field work to avoid contact between the PHC enumerators and PEC interviewers;
- The 2022 PEC used more tightly controlled data collection procedures; but
- The same definitions and classifications were used in the PEC as in the PHC.

2.4.3 Supervision and Monitoring of Data Collection

The supervision and monitoring of data collection involved a multi-layer approach. The supervisors kept close contact with the enumerators to continuously monitor the data collection activities explained any operational confusion of the PEC questions, correctly determined the EA, maintained liaison with the upazila statistical officers, and provided daily updates on the respective WhatsApp group opened for the purpose. The supervisors and upazila statistical officers closely supervised the

work of the enumerators and performed the following duties to ensure the quality of data: (i) on-the-spot verification of the interviewing technique to ensure that respondents are interviewed properly and correctly; (ii) discuss problems with the enumerators at the end of each day through specifically created WhatsApp platforms; and (iii) review the interviewers' daily progress reports.

2.4.4 Field Visits by Research Staff and Officials

Strict supervision of the field works was made to ensure quality data collection. To that end, the core team members were dedicated to monitoring the field activities and visited several EAs along with SID and BBS officials. Above all, the progress and quality of the data collection were digitally monitored at the NOC of the BBS Head Office. The BIDS research team and the officials of the BBS Head Office provided prompt resolutions to relevant issues in data collection. The BBS also entrusted high officials with visiting the fields to take necessary steps immediately. Apart from supervisory checks, these high-level officials gave on-the-spot solutions to the problems. These visits also helped boost the morale of the field teams to carry out strenuous jobs and maintain the desired quality of work.

2.5. Matching Operations

2.5.1 Matching Process

After the data were collected, consistency and quality checks were performed, and the entire dataset was cleaned duly before the analysis. The basic process of matching involves comparing names, addresses, and demographic characteristics between PHC and PEC. In general, it is an arduous process whereby households/dwelling units and persons enumerated during PHC and PEC operations are compared for similarities. A two-way matching process is normally used to identify omissions and erroneous inclusions in the PHC.

In the PEC, 155,463 persons were enumerated in the 378 EAs, and a match status had to be determined for all of them. The digital data-receiving procedure was the same as it was in the case of the PHC (see Figure 1.1). Once the PEC data, sent by the enumerators, were uploaded to the BBS server, the next stage required PEC data to

be matched against the corresponding PHC data to check the completeness of the enumeration. Therefore, a set of matching rules was developed for persons engaged in the matching operations.

Matching involved several stages: matching of EAs, matching of households, and matching of individuals. Matching was essentially done in two phases. For each of the EAs, all households were matched through a computer algorithm using the PHC and PEC data. During the first phase, strict matching rules were followed, resulting in obvious 'match' and 'possible match'. This matching process had two main purposes: (i) to record whether the household was matched and, if not, whether it was missed or unresolved; and (ii) to record the identifying information for the PHC questionnaire against which the PEC questionnaire was matched. After completing this matching process, a group of 'reviewers' was assigned to examine the matching process and compile the final match status.

2.5.2 Matching EAs

The first stage of matching an EA involved locating the PHC EA data corresponding to the PEC EA data. Insofar as the EAs were based on the areas as determined before the PHC and any changes made subsequently were not included in the original listings, the ex-post delineations of EAs did not always correspond with the ex-ante delineations. For example, it was found that during the PHC 2022, some EAs were split, combined, or had boundaries altered. Thus, the total number of households and/or persons for an EA differed between the PHC and PEC but were required for matching to that EA in the PEC.

Attempts were made in the BBS Headquarters to resolve these problems by importing household data from the adjacent EA(s), which eventually increased the number of EAs by 24; hence, the total number of EAs increased to 378. Sometimes the staff needed to look at the GPS coordinates in the maps to find the corresponding EA or EAs. In some cases, it was even necessary to talk to the relevant district/regional statistical officers to try to determine what had happened. Thus, it took an unusually long time for some of the EAs to locate the corresponding PHC EAs. Even though this affected a number of the EAs of the PEC, ultimately, matching could be done as required.

The problems with the correspondence of boundaries of the EAs between PHC and the PEC no doubt made matching difficult; however, these problems did not have any major impact on the calculated undercount. The methodology was based on the match status of each household and person. If a household and its occupants were enumerated as part of another EA in the census, thorough searching procedures adopted during the computer-based matching operation ensured that they were not erroneously treated as undercounts. If there were any doubts about whether they were enumerated, every effort was made to resolve such cases satisfactorily. Sometimes, even when corresponding EAs were found, other problems meant that visits to the EAs were necessary. Accordingly, the issues were resolved by re-visiting the corresponding EAs.

2.5.3 Matching Households

Once the corresponding EAs were located, the next task was to match at the household level. It was done by comparing the address listings in both the PHC and the PEC data to identify the corresponding households. Where this was inconclusive, the questionnaires were compared to see if a match could be found based on names and household structure. In doing so, a large number of households were matched during the PEC enumeration.

Each household was classified as *matched, missed,* or *unresolved* in the matching process. Where a household was classed as missed or unresolved, the reason for this was recorded if known. Matching was a straightforward process in areas with accurate addresses, where it was not difficult to identify a corresponding household or to confirm whether a household was missed in the PHC. However, matching was more complex in areas with confusing formal addresses, especially in urban areas. Difficulties arose when the names of household members did not match or were not unique or when the composition of households had changed. Sometimes it was impossible to confirm whether a particular household and its members were enumerated. In such cases, the household was classed as unresolved, and the decision was taken based on the results of the follow-up visits. Sometimes a household was found, but the household present at the time of the PEC was completely different

from the one in the PHC. It arose as the original household had left, and a new household moved in between the PHC and the PEC. In that process, a sizeable number of the hitherto non-matched households were matched.

2.5.4 Matching Persons

Once the corresponding household was identified, a match status was allocated to each person in the PEC data. In most cases, it was possible to identify a match based on the name. However, this was not easy in cases where, for example, a different name or initials appeared to have been used. In such cases, a judgment on whether or not a person was matched was made based on age, marital status, gender, and relationship to the head of the household. These variables did not have to be exactly the same for a match to be made as often; particularly for age, the responses differed slightly.

Usually, when a corresponding PHC household was identified, all members of the PEC household were classified as enumerated or missed. However, there were still some situations where it was necessary to allocate a code of unresolved for a person in the PEC data, for example, where some characteristics of a person in the PHC data were similar, others were still different, and the household structure did not help indicate whether the persons matched. In such cases, the unresolved category was used.

Similarly, if a person was in the PHC data but not in the PEC data, normally such persons were considered potential overcount if the household was the same in the PHC and the PEC, and there was no out-mover. However, it was not possible to draw this conclusion easily for all cases because of potential undercounts by the PEC. The undercount/missing persons, either by the PHC or the PEC, were duly coded during the matching operations. Specifically, the following steps were involved in completing the matching of persons:

1. Taking the field-level household matching by the enumerators and computerized matching between the PHC and the PEC, the following steps were followed for each of the matched households:
    i. *Exact Matching:* In this step, a maximum of 20% of the variations in names (variation of only a few characters (letters) between the sources) were allowed in both the PHC and PEC samples.

ii. *Gender and Relationship with Household Head*: Among the non-matched members in step (i) in both PHC and PEC, members were matched based on their gender and relationship with the household head.

iii. *Gender and Age:* Among the non-matched members in step (ii) in both PHC and the PEC, members were matched based on gender, which minimizes the age difference for each of them between PHC and the PEC records.

iv. *Fuzzy Matching of Name and Age:* Among the non-matched members in step (iii), the variation of names was minimized such that the age difference of matched observation from both sources follows a specific age limit of 10 years.

2. Among the field-level non-matched households, the matched households from the PHC and the PEC data files were first identified at the BBS Head Office, and then the four steps mentioned above were applied.

The matching operations for each household stop soon after the maximum number of possible matches for the household is reached both in (1) and (2).

2.5.5 *Follow-Up Phone Calls to the BBS Field Staff and Field Visits*

The purpose of the reconciliation calls to field staff of the BBS and field visits by the research team is to collect relevant information to determine the final match status of unresolved cases identified during the computerized matching specifically: (i) resolve the final match status for ''possible match'' cases; (ii) determine whether households and/or persons enumerated in the PHC but not in the PEC is correctly or erroneously enumerated in the PHC; (iii) determine whether households and/or persons enumerated in the PEC but not in the PHC is correctly or erroneously enumerated in the PEC; and (iv) clarify doubtful cases or cases with insufficient or unclear information. It may be noted that the steps suggested in Mule (2012) were followed in determining the correctly enumerated persons through these processes.

Final matching involves using the results obtained from the reconciliation phone calls, field visits, and initial matching phases to assign a definite match status to each case. Table 2.2 illustrates the expected outcomes from the final matching.

Table 2.2: PEC Matching Status of Household Members

1.	Matched
In the PEC but not in the PHC:	
2.	Missed in the PHC
3.	PEC erroneous inclusion-- cases that are outside the EA boundaries or otherwise erroneously included in the PEC
4.	PEC insufficient information-- cases for which a final match status cannot be assigned due to insufficient information
5.	In-mover
6.	Born after the PHC
In PHC but not in the PEC:	
7.	Correctly enumerated in the PHC but missed in the PEC
8.	PHC erroneous inclusion
9.	PHC insufficient information-- cases for which a final match status cannot be assigned due to insufficient information.

2.6. Estimation Procedure

Since the PHC and the PEC are conducted by different enumerators, and the two operations are independent, the dual system of estimation procedure has been used to estimate the persons omitted by both the PHC and the PEC. The population eligible for enumeration in the PHC can be categorized as those enumerated both in the PHC and the PEC, those enumerated only in the PEC or the PHC, and those not enumerated in either the PEC or the PHC.

The dual system of estimation is used in the PEC to estimate gross and net coverage errors. In the dual system, data from the PEC is matched with those obtained in the PHC to arrive at the true population. The resulting tally can be represented in a 2×2 contingency table. It symbolically shows inputs into the dual system estimation of the *true population* (Table 2.3).

Table 2.3: Tally of a 2×2 Contingency Table

	In PHC	Out of PHC	Total
In PEC	N_{11}	N_{12}	N_{1+}
Out of PEC	N_{21}	N_{22}	N_{2+}
Total	N_{+1}	N_{+2}	N_{++}

where,

N_{11} is the estimate of the number of persons counted in both the PHC and the PEC;

N_{12} is the estimate of the number of persons counted only in the PEC;

N_{21} is the estimate of the number of persons counted only in the PHC;

N_{22} is the estimate of the number of persons missed by both the PHC and the PEC;

N_{1+} is the estimate of the total number of persons counted in the PEC;

N_{+1} is the total number of persons counted correctly in the PHC after factoring out the erroneous inclusions; and

N_{++} is the estimate of the total population.

Three of the components of the total population, N_{11}, N_{21}, and N_{12}, are available from the PHC and the PEC data. The number of persons omitted by both the PEC and the HCC is not available. Under the assumption of independence of the PEC and the PHC, the unknown number N_{22} can be estimated by following Chandrasekaran and Deming (1949) as:

$$N_{22} = (N_{12} X N_{21})/ N_{11} \quad (2.1)$$

This formula is conceptualized as a stochastic Bernoulli event, which implies that N_{11}, N_{21}, and N_{12} are random variables (the sums of Bernoulli outcomes based on the strong assumptions of perfect matching), and the samples are independent (US Bureau of the Census, 1985).

The nicety of the dual system estimate is that it raises the corrected census total (where erroneous enumerations are subtracted from the census population) by the total estimate of the number of persons in the PEC divided by the estimate of the PEC number that matched the PHC. The net omission is obtained by subtracting the number of persons that have been duplicated from the number of persons omitted.

CHAPTER 3: EVALUATION OF COVERAGE

Following the US Bureau of the Census (1979), Dauphin and Canamucio (1993), and the operational manual of the United Nations (UN, 2010), the coverage measures were constructed only for cases belonging to the universe of interest. In other words, the system excluded the erroneous inclusions. The objective is to identify all the elements that are essential in deriving the dual system estimates. The estimation procedure is outlined below.

3.1. Sample Weights

The sample allocation is described in Table 2.1 in Chapter 2. The EAs were stratified by rural and urban, and urban EAs, in turn, were stratified by upazila sadar and growth center, municipality, and city corporations. Within each stratum, units were sorted geographically across the administrative divisions. The weight of a sample EA was equal to the inverse of the population concerned. Within each EA, the weight for each household and each person was equal to the EA sampling weight since their probability of selection, given the selection of the EA, was equal to one. It should be noted that during the data collection process, there were no substitutions of selected EA due to problems related to field operations or frame; hence, no adjustment was made to the weights. Thus, the inverse selection probabilities were applied as calculated from the computer program without any further adjustments.

3.2. Basic Tenets of Coverage Estimation

The estimates are calculated based on the PEC sample for the following parameters in the initial tabulation. These estimates consist of the sum of the sample values either from the so-called P sample or the E sample.

a. Total number of non-movers in the universe (P sample);
b. Total number of out-movers in the universe (P sample);
c. Total number of in-movers in the universe (P sample);
d. Total number of matched non-movers in the universe (P sample);
e. Total number of matched out-movers in the universe (P sample);

f. Estimated total number of matched in-movers in the universe (P sample);

[**Note:** *since a matching of the in-movers is not attempted, the number of matched in-movers cannot be calculated directly. However, the assumption of a closed population implies that the 'out-mover' and the 'in-mover' constitute the same group in the universe: the 'mover'. Accordingly, one can assume that in the universe, the match rate for in-movers would be the same as that for out-movers. This match rate can be estimated by e/b. Hence, the total number of matched in-movers in the universe is estimated indirectly by [(e/b)*c].*[2]

g. Total number of census erroneous inclusions in the population (E sample);

h. Total number of cases correctly enumerated in the census but missed in the PEC (E sample);

i. Total number of census persons with insufficient information (E sample); and

j. Total number of PEC erroneous inclusions and PEC insufficient information cases (E sample).

For the operational purpose, the dual system of estimation of the PHC and the PEC starts by assigning symbols to various estimates in Table 3.1 to facilitate the development of compact standard formulas as illustrated in Table 3.2:

Table 3.1: Basic Elements of Dual System Estimates

Symbol	Parameter	Derivation from the Questionnaire
I1	Number of non-movers	q17a (PEC) == 1 & q17b (PHC) == 1
I2	Number of in-movers	q17a (PEC) == 1 & q17b (PHC) == 2
I3	Number of out-movers	q17a (PEC) == 2 & q17b (PHC) == 1
I4	Number of matched non-movers	q17a (PEC) == 1 & q17b (PHC) == 1 & match status == "Yes"
I5	Estimated rate of matched out-movers	q17a (PEC) == 2 & q17b (PHC) == 1 & match status == "Yes"
I6	Estimated number of matched in-movers	[(I5/I2) * I3] following Dauphin and Canamucio (1993)
I7	Number of PHC erroneous inclusions	Present in PHC data & q34 ≠ 1
I8	Number of correctly enumerated PHC persons missed in the PEC	Present in PHC & match status == "No" & q34 == 1
I9	Number of PHC persons with insufficient information	Total number of PHC persons with insufficient information to be confirmed that they were correctly enumerated during the census (PHC sample)

Note: The question numbers in the derivation column refer to the question numbers on the PEC Questionnaire in Appendix A.

[2] For details, see Dauphin and Canamucio (1993).

Following Dauphin and Canamucio (1993) and UN (2010), the operational definitions below are used in estimating critical parameters based on Table 3.1.

(a) The 'matched' population is given by the total number of matched non-movers plus the estimated total number of matched in-movers in the universe.

$$\text{Matched Population} = \text{Matched Non\_movers} + \text{Estimated Matched In\_movers} \quad (3.1)$$

(b) The estimate of the population enumerated in the PHC [Uncorrected PHC Population] is the sum of the matched population, the population erroneously included in the PHC, the population correctly enumerated in the PHC but missed in the PEC, and the PHC cases with insufficient information.

$$\text{Uncorrected PHC Population} = \text{Matched Population} + \text{Correctly enumerated in PHC but missed in PEC} + \text{PHC Erroneous Inclusion} + \text{PHC Insufficient Information} \quad (3.2)$$

(c) The corrected PHC population is calculated without adding the erroneous inclusions and the PHC persons with insufficient information

$$\text{Corrected PHC Population} = \text{Matched Population} + \text{Correctly enumerated in PHC but missed in the PEC} \quad (3.3)$$

(d) The PEC-sample estimate of the total population [PEC Population] is the sum of the non-movers and in-movers in the population.

$$\text{PEC Population} = \text{Non\_movers} + \text{In\_movers} \quad (3.4)$$

(e) The PEC-enumerated population missed in the census is calculated by subtracting the matched population from the PEC estimate of the total population to obtain:

$$\text{PEC Population missed in PHC} = \text{PEC Population} - \text{Matched Population} \quad (3.5)$$

(f) The rate of PEC population missed in the PHC is the missed population above relative to the PEC estimate of the total population. The estimated total number of PHC erroneous inclusions is calculated by summing over fabrications, duplications, geographic misallocations, aliens, etc. The main purpose of this construct is to provide an estimate to permit a correction in

the dual system estimate of the true population. The PHC erroneous inclusion rate is equal to the total number of persons erroneously included in the PHC relative to the estimate of the PHC population.

(g) The preliminary dual system estimate of the true population is the population estimated from the PEC multiplied by the population estimated from the PHC (after correcting for erroneous inclusions and insufficient information) and divided by the matched population.

(1) $$\text{True Population} = \frac{\text{PEC Population} * \text{Corrected PHC Population}}{\text{Matched Population}} \quad (3.6)$$

(i) The net coverage error – also known as the 'net omission rate' or the 'net undercount rate' – is the difference between what should have been counted (true population) and what was counted (census population). The net coverage error rate is the total net error relative to the dual system estimate of the true population. This measure constitutes the single most important indicator of the quality of census coverage.

$$\text{Net Undercount Rate} = \frac{(\text{True Population} - \text{Uncorrected PHC Population})}{\text{True Population}} \quad (3.7)$$

(j) The gross coverage error – the 'gross omission' – is, as defined in this context, what the census truly missed without taking into account the overcount. It is the gross omission relative to the true population, as opposed to the net omission, without being offset by the erroneous inclusions.

$$\begin{aligned} &\text{Gross Coverage Error} = \text{Population counted in PEC but missed in PHC} - \\ &\text{Population missed in both PHC and PEC} = \text{PEC Population missed in PHC} + \\ &\frac{(\text{True Population} - \text{Corrected PHC Population}) * (\text{True Population} - \text{PEC Population})}{\text{True Population}} \end{aligned} \quad (3.8)$$

$$\text{Gross Coverage Error Rate} = \frac{\text{Gross Coverage Error}}{\text{True Population}} \quad (3.9)$$

Equivalently,

$$\text{Gross Coverage Error Rate} = \left(1 - \frac{\text{Matched Population}}{\text{PES Population}}\right) = \text{Rate of PEC persons missed in PHC,}$$

which means the total gross error can be calculated as follows:

$$\text{Total Gross Error} = \text{Rate of PEC persons missed in PHC} \times \text{True Population} \quad (3.10)$$

(k) The final dual system estimate of the true population, which corresponds to the 'adjusted population', is obtained through a ratio estimator of the total, which is superior in accuracy to the preliminary estimate as it reduces both variance and bias.

$$\text{Final Dual System of True Population} = \left[\frac{\text{Preliminary True Population}}{\text{Uncorrected PHC Population}}\right] * \text{Actual PHC Count} \qquad (3.11)$$

where the ratio inside the bracket represents the 'adjustment factor' for the census count.

(l) The relation between the undercount rate and the adjustment factor is the following:

$$\text{Adjustment Factor} = \frac{1}{1-\text{Undercount Rate}} \qquad (3.12)$$

In other words, the adjustment factor is the reciprocal of the complement of the undercount rate. For example, an undercount rate of 2% implies an adjustment factor of 1.02, subject to rounding errors. Likewise, an undercount rate of 10% implies an adjustment factor of 1.11, and so forth. Another way of viewing the adjustment factor is the following:

$$\text{Adjustment Factor} = \frac{\text{PES Population} * \text{Corrected Census Population}}{\text{Matched Population} * \text{Uncorrected Census Population}}$$

If we consider $\frac{\text{Matched Population}}{\text{PEC Population}}$ as the 'coverage rate', then:

$$\text{Adjustment Factor} = \left(\frac{1}{\text{Coverage Rate}}\right) * \left(\frac{\text{Corrected PHC Population}}{\text{Uncorrected PHC Population}}\right)$$

While the first term inside the parentheses is a correction for under-enumeration, the second term – the proportion of the PHC population that was correctly enumerated, i.e., not erroneously included – serves as a correction for over-enumeration. Hence, the final adjusted population is, in effect, calculated as follows:

$$\text{Final Adjustment True Population} = \text{Underenumeration correction factor} * \text{Overenumeration correction factor} * \text{PHC count} \qquad (3.13)$$

Also, note that the under-enumeration correction factor is always ≥ 1, and the over-enumeration correction factor is always ≤ 1. The overall factor can theoretically

fall on either side of 1, depending on which of the two factors is stronger. The outcome of the above formulae is summarized in Table 3.2, and the probabilities of inclusion and omission of a person are presented in Table 3.3.

Table 3.2: Schematic Derivations of the Dual System Estimators

	Parameter	Derivation
A1	(I4 + I5) + I7 + I8 + I9	PHC Population
A2	I1 + I3	PEC Population
A3	I4 + I6	Matched Population
A4	A2–(I4 + I6)	PEC Persons Missed in PHC – Total
A5	(A4/A2)*100	PEC Persons Missed in PHC – Rate (%)
A6	100-A5	PHC Coverage Rate (%)
A7	I8	Correctly Enumerated in PHC but Missed in PEC
A8	(A7/A1)*100	Correctly Enumerated in PHC but Missed in PEC – Rate (%)
A9	I7	PHC Erroneous Inclusions – Total
A10	I9	Estimated Number of PHC Persons with Insufficient Information
A11	(A9/A1)*100	PHC Erroneous Inclusions – Rate (%)
A12	(A1*A2)/A3	Preliminary Dual System Estimate of True Population
A13	A12–A1	Net Error (Net Undercount) – Total
A14	(A13/A12)*100	Net Error (Net Undercount) – Rate (%)
A15	A4+A9	Gross Coverage Error – Total
A16	(A15/A12)*100	Gross Coverage Error Rate Relative to True Population (%)
A17	A6/A1	'Adjustment Factor' for the PHC
A18	A9*Actual PHC Count	Final Dual System Estimate of True Population

Note: The parameters are based on groupings in Table 3.1.

Table 3.3: Schematic Derivation of Probabilities of Inclusions and Omissions

P (included in Census)	Census Population Corrected/True Population
P (included in PEC)	PEC Population/True Population
P (included in both Census and PEC)	P (included in Census) * P (included in PEC), based on the independence assumption
P (included in Census, but missed in PEC)	P (included in Census) * [1- P (included in PEC)]
P (included in PEC, but missed in Census)	P (included in PEC) *[1- P (included in Census)]
P (missed in both Census and PEC)	[1-P (included in Census)] *–1 - P (included in PEC)]

The distribution of the *true population* based on the preliminary dual system estimate after removing the cases of erroneous inclusions and insufficient information in the PHC is presented in Table 3.4.

Table 3.4: Schematic Distribution of True Population

		Census Enumeration		
		Included	*Omitted*	Total
PEC Population	*Included*	Matched population	In PEC but missed in PHC	PEC population
	Omitted	In the PHC but missed in the PEC	Missed in both	
	Total	PHC population corrected	Gross omission	True population

The elements in the above matrix are calculated using the formula listed in the following box.

PHC population corrected for erroneous inclusion and insufficient information = P (included in PHC) × Dual system estimate of the population

PEC population (excludes erroneous inclusion and insufficient information) = P (included in PEC) × Dual system estimate of the population

Population included in both PHC and PEC = P (included in both PHC and PEC) × Dual system estimate of the population

Population included in the PHC but missed in the PEC = P (included in PHC but missed in PEC) × Dual system estimate of the population

Population included in the PEC but missed in the PHC = P (included in PEC but missed in PHC) × Dual system estimate of the population

Population missed in both the PHC and the PEC = P (missed in both the PHC and the PEC) × Dual system estimate of the population

3.3. Empirical Elements of the Coverage Evaluation

The estimate of the true population is derived by exploiting two independent sources of information: the PHC and the PEC. Based on an exhaustive enumeration, the first step in measuring the true population starts with the population enumerated in the PHC. The second attempt yields the PEC estimate of the total population based on sampling techniques. Both of these are used to derive a third composite estimate of the true population--called the dual system estimate. It may be stressed that neither the PHC nor the PEC is considered superior to the other as both are subject to the same types of non-sampling errors, e.g., non-responses. However, the dual system

estimate is more comprehensive than the PHC or the PEC estimate alone as it draws on the information from both the PHC and the PEC. Hence, the true population is compared with the population enumerated in the census to arrive at the estimate of the net undercount rate, both nationally and across other domains in which the PEC sample is representative, e.g., locations, personal attributes, administrative area, etc.

3.3.1 Estimate of True Population

Although the counts of the population by rural and urban locations have been reported, only the national estimates are explained in detail below for brevity. It may be noted that four components together make up the dual system estimate of the true population in Bangladesh. Table 3.5 provides a breakdown of the dual system estimate of the true population of 158,352 persons. Following the cell composition of Table 2.3, components N_{11}, N_{12}, and N_{21} are based on direct observation and obtained through a matching process. In contrast, component N_{22} is obtained through a mathematical derivation. It is based on an assumption of independence between the PHC enumeration and the PEC sample following Chandrasekaran and Deming (1949). Note that component N_{11} – the population included in both the PHC and the PEC – is estimated at 142,509 persons. Component N_{12} – the population included in the PEC but missed in the PHC – is estimated at 4,110 persons. Component N_{21} – the population included in the PHC but missed in the PEC – is estimated at 11,403 persons. Finally, component N_{22} – the population missed in both the PHC and the PEC – is estimated at 329 persons.[3]

Table 3.5: Empirical Distribution of True Population

		PHC		
		Included	**Omitted**	**Total**
PEC	Included	142,509	4,110	146,619
	Omitted	11,403	329	11,732
	Total	153,912	4,439	158,352

Note: Estimates are subject to rounding errors.

[3] It may be noted that there were 6 and 11 cases of transgender in the PHC and the PEC, respectively. As these numbers were small, these cases were dropped from both the PHC and the PEC. The inclusion of these cases unduly influences the results of the coverage and content analyses, especially the gender domain. These omissions may introduce biases in the coverage and reflect different consistency patterns, and thus the content error measures might also be somewhat biased. However, the directions of these biases cannot be determined *a priori*.

Table 3.6 presents a categorical breakdown of the dual system estimates. Of the 153,912 persons for the in-scope sub-universe of the population, 142,766 persons are non-movers, 8,051 are out-movers, and 3,853 are in-movers. As many as 138,831 persons could be matched between the PHC and the PEC. Similarly, 7,686 persons could be matched as out-movers. On the assumption of a closed population, following the US Bureau of the Census (1979) and Dauphin and Canamucio (1993), the matched in-movers are estimated at 3,678 persons. Only 181 persons, or approximately 0.11% of the total PHC persons, were erroneously included in the PHC. This small group constitutes only foreign nationals; cases of duplications and fabrications were not found. Finally, 7,214 persons are estimated to be correctly enumerated in the PHC that were missed in the PEC.

Table 3.6: Elements of the Dual Estimates of Population

Symbol	Description	Persons
I1	Non-movers	142,766
I2	Out-movers	8,051
I3	In-movers	3,853
I4	Matched non-movers	138,831
I5	Matched out-movers	7,686
I6	Matched in-movers [(I5/I2)*I3]	3,678
I7	Erroneous inclusions	181
I8	Correctly estimated in PHC but missed in PEC	7,214

Note: Estimates are subject to rounding errors.

Following the dual system of estimation suggested by Dauphin and Canamucio (1993) and UN (2010), the PHC and the PEC persons were estimated at 153,912 and 146,619, respectively (Table 3.7). From these two independent sources of enumeration, 142,509 persons could be matched. As many as 4,110 persons, or 2.8% of the PHC persons, were missed in the PEC. Similarly, 7,214 persons, or 4.69% of the PEC persons, were missed in the PHC. The overall coverage rate of the PHC was 97.20%. The rate of erroneous inclusions in the PHC is estimated at 0.12%. Note that this estimate of total omission does not take into account the erroneous inclusions. Once it is offset by the 181 erroneous inclusions, the above coverage results would change.

Table 3.7: Empirical Derivations of the Dual System Estimates

Symbol	Derivation	Description	Value
A1	(I4 + I5) + I7 + I8	PHC Population	153,912
A2	I1 + I3	PEC Population	146,619
A3	I4 + I6	Matched Population	142,509
A4	A2–(I4 + I6)	PEC persons missed in PHC – Total	4,110
A5	(A4/A2)*100	PEC persons missed in PHC – Rate (%)	2.803
A6	100-A5	PHC Coverage Rate (%)	97.197
A7	I8	Correctly enumerated in PHC but missed in PEC	7,214
A8	(A7/A1)*100	Correctly enumerated in PHC but missed in PEC – Rate (%)	4.687
A9	I7	PHC erroneous Inclusions – Total	181
A10	(A9/A1)*100	PHC erroneous Inclusions – Rate (%)	0.118
A11	(A1*A2)/A3	Preliminary Dual System Estimate of True Population	158,351
A12	A11–A1	Net Coverage Error (Net Undercount) – Total	4,439
A13	(A12/A11)*100	Net Coverage Error (Net Undercount) – Rate (%)	2.803
A14	A4+A9	Gross Coverage Error – Total	4,291
A15	(A14/A11)*100	Gross Coverage Error Rate Relative to True Population (%)	2.927
A16	A11/A1	'Adjustment Factor' for the PHC	1.029
A17	A16*Actual PHC Population	Final Dual System Estimate of True Population	158,352

Notes: (1) Estimates are subject to rounding errors. (2) The term 'True Population' should be interpreted as the actual population of the EAs counted and not the 'True Population' counted in the country as a whole. (3) The actual PHC count in the Table should not be confused with the total PHC count of the country as a whole. (4) Row A17 is used only to show the correspondence between Table 3.5 and Table 3.7.

Table 3.8 shows the overall empirical probabilities of inclusion and omission of a person in the PHC or the PEC, which are obtained following the details on derivations presented in the previous section. The results imply that a person of the in-scope sub-universe had approximately a 97.20% chance of being enumerated in the PHC, a 92.59% chance of being enumerated in the PEC, and a 90.00% chance of being enumerated in both. Conversely, a person had approximately a 7.20% chance of being included in the PHCs but missed in the PEC, a 2.60% chance of being included in the PEC but missed in the PHC, and a 0.21% chance of being missed in both. As mentioned before, the probability of the population not being included in either the PHC or PEC is very low. It implies that the enumeration procedure has captured the target population well.

Table 3.8: Probabilities of Inclusions and Omissions of Persons

Symbol	Derivation	Value
P (included in the PHC)	A1/A11	0.9720
P (included in the PEC)	A2/A11	0.9259
P (included in both the PHC and the PEC)	(A1/A11)*(A2/A11)	0.9000
P (included in the PHC but missed in the PEC)	(A1/A11)*(1 - A2/A11)	0.0720
P (included in the PEC but missed in the PHC)	(1 - A1/A11)*(A2/A11)	0.0260
P (missed in both the PHC and the PEC)	(1 - A1/A11)*(1 - A2/A11)	0.0021

Note: Estimates are subject to rounding errors.

3.4. Net Coverage Error Rates

Following Dauphin and Canamucio (1993) and UN (2010), the net coverage error or undercount is defined as the difference between the estimated true population (dual system estimate) and the estimated PHC population. The corresponding rate is expressed as a percentage of the estimated true population. The net coverage error rates, the corresponding standard errors, and the coefficient of variations are reported in Table 3.9. The standard errors are estimated following Cochran (1977) and Endlich et al. (1988) for the case of weighted mean,[4] while the coefficient of variation is estimated following the US Bureau of the Census (1979).

Table 3.9: Net Coverage Errors by Location

Domain	Rate (%)	SE (%)	95% CI	CV (%)	Pop. Weights
Bangladesh	2.75	0.59	[1.60, 3.90]	21.39	
Rural	2.55	0.41	[1.74, 3.34]	16.09	68.49
Urban	3.19	0.48	[2.25, 4.13]	15.10	31.51
Upazila Sadar/Growth Center	2.62	0.89	[0.87, 4.37]	34.04	7.79
Municipality	2.99	0.78	[1.46, 4.53]	26.11	10.84
City Corporation	3.71	0.90	[1.94, 5.47]	24.27	12.89

Note: Estimates are subject to rounding errors.

The national net coverage error rate is estimated at 2.75%, which is a population-weighted average of the constituting estimates in rural and urban areas. Unlike the earlier estimates, the current estimates of the national and other domain-specific coverage error rates were weighted by the population share of the districts. The net

[4] Bootstrapping results by Gatz and Smith (1995) show that the variance formula suggested by Cochran (1977) and Endlich et al. (1988) is a reasonable estimator for the square of the standard error of the mean.

coverage error rate is lower in rural areas (2.55%) than in urban areas (3.19%). It may be noted that the net coverage error rates increase with the level and sophistication of urbanization. It is also noted that the standard errors in urban areas are higher than in rural areas. The rural coverage error rate is subject to a 21.39% relative error rate in contrast to 34.04% in the upazila sadar and growth center. The higher the coverage error and the standard error, the higher the coefficient of variation.

The erroneous inclusion rate provides an idea about the extent of misrepresentations of PHC cases due to fabrications, duplications, aliens, and geographic misallocations. As reported in Table 3.7, the erroneous inclusion rate is only 0.12%. So, this error rate has little influence on the estimated net coverage error rates.

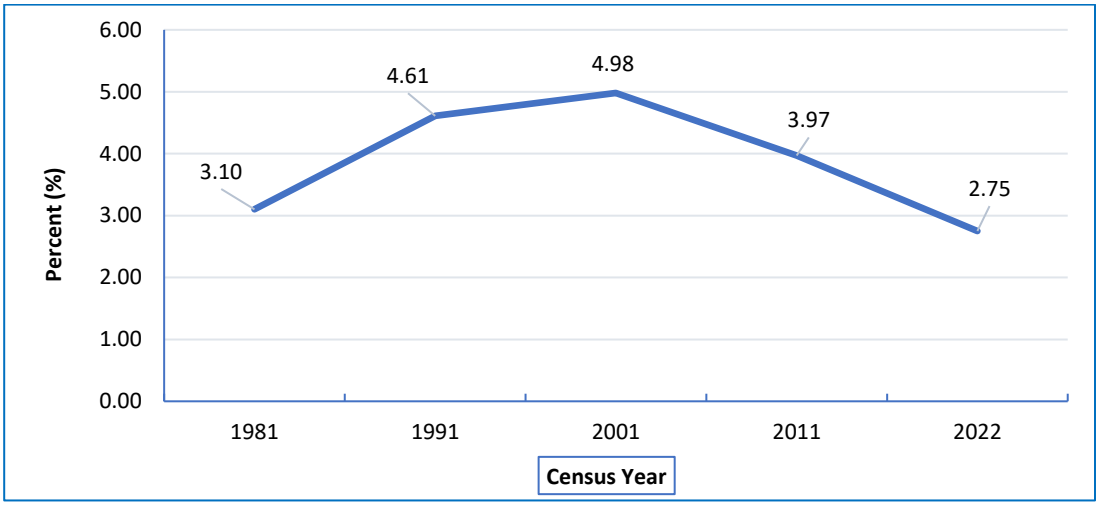
The estimates of the net coverage error rates for PHC 2022 may be compared with the historical rates (Figure 3.1). A comparison of decennial census coverage error estimates reveals that the national estimate for PHC 2022 is 1.23 percentage points lower than that found in PHC 2011 (Mannan, Sohail, & Bhattacharjee, 2012) and 2.23 percentage points lower than that found in PHC 2001 (BBS, 2003). Moreover, the national estimate for PHC 2022 is 1.86 percentage points lower than that found in the PHC 1991 (BBS, 1994) and 0.35 percentage points lower than that found in the PHC 1981 (BBS, 1984).[5] It may be noted that the national coverage error rate was not estimated for the PHC 1974. Instead, the coverage error rates were estimated for the four major cities (Chattogram, Dhaka, Narayanganj, and Khulna) at 19.30% and 6.50% for the rest of the country. However, due to the small number of blocks matched, the results of the PEC of the PHC 1974 remain of limited value (BBS, 1977). Be that as it may, it would be interesting to assess why the census coverage error rate follows an inverted U-shaped trajectory over the course of four decades between 1981 and 2022.

The above national coverage error rates are based on the implicit assumption that the undercount is homogenously distributed across locations, administrative

[5] It may be noted the coverage error rates reported in the PEC of the PHC 2011 were not weighted by the population shares but by the sample shares and hence are not strictly comparable. Whether or not the coverage rates reported for the PHC 2001 and earlier were weighted could be ascertained.

divisions, and attributes of persons wherein the matching process is conducted. However, the homogeneity assumption is strong and thus needs to be empirically verified. To that end, the coverage error rates are estimated across a few attributes of persons and administrative divisions of the country.

Figure 3.1: Trend of the Historical National Coverage Error Rates



Sources: BBS (1984; 1994; 2003); Mannan, Sohail, and Bhattacharjee (2012); authors' estimates.

Table 3.10 presents coverage error rates by gender, religion, and age group. The estimates reveal that the net undercount rate is higher at 2.81% for males compared to 2.69% for females. Besides, both the standard error and the coefficient of variation of the coverage error rate are also lower for females. One of the plausible reasons for the higher coverage error rate for males could be mobility. There is a higher likelihood that a person on the move away from the usual dwelling would be missed in the PHC. However, this is an empirically verifiable proposition. The coverage error rates by religion show that the extent of undercount appears to be higher for Muslims at 2.76% compared to the non-Muslims at 2.67%. The coverage error rates for the non-Muslim population by broad religion could not be estimated due to data limitations. Be that as it may, the estimate of the coverage error rate for the non-Muslim population is imprecise as its 95% confidence interval includes 0 in it.

The coverage error rates for age groups follow a U-shaped pattern. The successive rates secularly decrease until the minimum is reached at the 40-44 age group; beyond that point, the rates follow a secularly upward trend until the maximum age group is reached. Both the lower and the higher age groups have higher rates; while the

coverage error rate for the highest age group (75+ years) is estimated at 4.82%, the same is estimated at 4.23% for the 0-4 age group. These estimates are rather surprising as there is a high likelihood that persons in both of these age groups tend to stay at home and, therefore, have a lower chance of missing the PHC enumeration. In contrast, persons between the 30-34 and 55-59 age groups have coverage error rates of less than 2%. Despite the wide variations in the coverage error rates across age groups, many of the standard errors are 'high,' which tends to make the estimates imprecise, which is evident from the inclusion of 0 in the 95% confidence internals from the age group 35-39 and above. This caveat needs to be kept in mind while using the coverage error rates of these age groups.

Table 3.10: Net Coverage Errors by Gender, Religion, and Age Group

Domain	Rate (%)	SE (%)	95% CI	CV (%)	Pop. Weights
Gender					
Female	2.69	0.72	[1.28, 4.10]	26.63	50.50
Male	2.81	0.78	[1.28, 4.34]	27.86	49.50
Religion					
Muslim	2.76	0.74	[1.31, 4.21]	26.76	91.00
Non-Muslim	2.67	1.49	[-0.25, 5.59]	55.87	9.00
Age Group					
0-4	4.23	1.33	[1.62, 6.83]	31.47	9.44
5-9	3.06	1.31	[0.49, 5.62]	42.86	9.28
10-14	3.09	1.32	[0.50, 5.67]	42.77	9.89
15-19	3.52	1.11	[1.34, 5.69]	31.57	10.03
20-24	3.80	0.97	[1.90, 5.70]	25.55	9.08
25-29	2.58	1.11	[0.40, 4.75]	43.08	8.71
30-34	1.95	0.89	[0.20, 3.69]	45.73	7.34
35-39	1.66	0.98	[-0.26, 3.58]	59.17	7.70
40-44	1.17	0.61	[-0.03, 2.36]	52.31	6.08
45-49	1.44	1.01	[-0.54, 3.42]	70.32	5.01
50-54	1.64	1.14	[-0.60, 3.87]	69.67	4.72
55-59	1.87	1.20	[-0.49, 4.22]	64.30	3.47
60-64	2.22	1.32	[-0.37, 4.80]	59.56	3.39
65-69	2.54	2.15	[-1.68, 6.75]	84.77	2.38
70-74	3.25	2.99	[-2.61, 9.11]	92.10	1.71
75+	4.82	2.74	[-0.55, 10.19]	56.89	1.79

Note: Estimates are subject to rounding errors.

Table 3.11 presents the net coverage error rates across the administrative divisions. The results indicate that the coverage error rates are lower in the western part of the country. Among the administrative divisions, the overall net coverage error rates in Khulna, Rajshahi, and Rangpur divisions are lower than the national average rate. The corresponding standard errors are also lower, lowering the

coefficients of variation as a consequence. However, the rates across divisions mask the variations across rural and urban areas (not shown in the table). Even in the western part of the country with lower overall coverage error rates, there are particular locations where net coverage error rates are higher than the national average rate.

Table 3.11: Net Coverage Errors by Administrative Division

Division	Rate (%)	SE (%)	95% CI	CV (%)	Pop. Weights
Barishal	2.42	0.61	[1.22, 3.63]	25.30	5.51
Chattogram	2.86	0.47	[1.95, 3.78]	16.30	20.10
Dhaka	3.13	0.73	[1.70, 4.56]	23.27	26.77
Khulna	2.23	0.57	[1.13, 3.34]	25.31	10.55
Mymensingh	3.26	0.66	[1.96, 4.56]	20.36	7.40
Rajshahi	2.12	0.12	[1.88, 2.36]	5.79	12.32
Rangpur	2.27	0.51	[1.27, 3.27]	22.43	10.66
Sylhet	3.33	0.23	[2.87, 3.79]	7.05	6.68

Note: Estimates are subject to rounding errors.

3.5. Adjustment Factors for Population Counts

Census results may be adjusted if coverage errors are substantial and the validity of census results is questionable. Insofar as the decision to adjust census figures is sensitive, it is bound to be decided at the highest levels of the government bureaucracy or the cabinet level (UN, 2010). "In any case, it will not likely be the analysts, statisticians, or technicians alone who ultimately decide whether an adjustment will be done; it is their responsibility, based on their technical expertise, to make appropriate recommendations to the decision-makers" (US Bureau of the Census, 1985).

The net coverage error rates provide a basis for adjustment of the census count of the population. However, it is up to the BBS to adjust the PHC counts at the direction of the highest policy-making authority. If an adjustment is to be made, consideration needs to be given to what domains are important in the decision to adjust the PHC counts. Any adjustment to the PHC will have far-reaching consequences. If coverage for a certain population group is not up to the standards set by the BBS, the adjustment may be of great benefit to the population group(s) disproportionately under-represented and the economy as a whole.

At least two issues arise in the adjustment of census counts of the population. *First*, the PHC population could be adjusted either at the aggregate level or at the

disaggregate levels by domains. The overall coverage estimates, when broken down by domains (such as location, gender, religion, age group, or administrative division), could be skewed as persons and households are not evenly missed over the sub-domains of the respective domains. Accordingly, these complexities call for adjustment factors by domain.

Table 3.12: Adjustment Factors for the PHC 2022 by Domain

Domain	Adjustment Factor
Bangladesh	**1.028277635**
Location	
Rural	1.026139810
Urban	1.032923504
Upazila Sadar/Growth Center	1.026879160
Municipality	1.030795718
City Corporation	1.038503403
Gender	
Female	1.027666362
Male	1.028935226
Religion	
Muslim	1.028380429
Non-Muslim	1.027397837
Age Group	
0-4	1.044131372
5-9	1.031528165
10-14	1.031847522
15-19	1.036446799
20-24	1.039463773
25-29	1.026445152
30-34	1.019849311
35-39	1.016841448
40-44	1.011799494
45-49	1.014571511
50-54	1.016634675
55-59	1.019017732
60-64	1.022665598
65-69	1.026023754
70-74	1.033553988
75+	1.050604394
Administrative Division	
Barishal	1.024803678
Chattogram	1.029403211
Dhaka	1.032314928
Khulna	1.022812201
Mymensingh	1.033702185
Rajshahi	1.021662724
Rangpur	1.023230823
Sylhet	1.034450632

Notes: (1) The adjustment factors are derived from the net coverage error rates reported in Table 3.9, Table 3.10, and Table 3.11, respectively, following eq. (3.12) discussed in the text. (2) The factors are fixed at nine decimal points to minimize the discrepancy between the estimates derived from the national rate and the weighted sum of the sub-domain rates. Despite these efforts, the estimates are subject to rounding errors.

Second, a decision has to be made on whether the national adjusted population should be the separate ratio estimate of the total (summing up the adjusted population across adjustment domains) or the combined ratio estimate of the total using the national adjustment factor. The combined-ratio estimate has a higher variance but its bias is lower than that of the separate-ratio estimates due to the consistency property of ratio estimators (which makes the bias diminish as n gets larger). Even though the separate-ratio estimates produce a lower variance with a higher bias than the combined-ratio estimates, it should be used as an alternative only when each sub-domain has a large number of observations. Besides, the 'PEC universe' may be limited to a subset of a larger population from which net coverage error rates across domains and sub-domains can be estimated.

Given that there was a sufficient number of observations across the sub-domains chosen, Table 3.12 presents the adjustment factors by different domains along with the corresponding sub-domains. The adjusted population can be obtained by multiplying the appropriate adjustment factor by the actual census count in the corresponding census adjustment sub-domain and then summing across sub-domains. In practice, it is equivalent to using a standard weighting procedure where the 'weight' corresponds to the adjustment factor.

If the coverage error rates, and hence the adjustment factors, of a segment of the population cannot be estimated, the total universe for the PHC needs to be partitioned into two sets: 'population within in-scope sub-universe' and 'balance of population'. Each person needs to be determined to be in or out of the target population based on the domains. In other words,

Unadjusted census population = Unadjusted 'population within in-scope sub-universe'

+ Unadjusted 'balance of population' (3.14)

It may be noted that the PEC was limited to a subset of the population as the socio-demographic information of a non-trivial number of persons could not be collected. Because the coverage rates in the 'balance of population' are unknown, no adjustment

can be made for these persons. Only eligible cases, i.e., cases in the in-scope sub-universe, should receive the designated adjustment factors. Non-eligible cases, i.e., balance-of-population cases, should receive an adjustment factor of 1. The eligible person should be assigned, on an individual level, the adjustment factor corresponding to the adjustment domain s/he belonged to, such as location, gender, religion, age group, and administrative division. Census counts, both unadjusted and adjusted, should then be calculated separately for the two population subsets and sum the two subsets to get the adjusted population:

Adjusted census population = Adjusted 'Population within in-scope sub-universe'

+ Unadjusted 'Balance of population' (3.15)

It is worth noting that PEC adjustment factors were based on persons' original geographic and demographic domains. Therefore, to maintain compatibility between the distribution of PEC cases and the PHC cases, the original domains should be used to decide which factor a person would receive.

It may be noted that information on 85,957 persons from 17,507 households was partially available in the PHC 2022 (BBS, 2022). While the information on 12,629 transgender people was available in the PHC 2022, only 11 of them were found in the PEC EAs. These transgender people were dropped from the analysis of PEC results for reasons cited before (see footnote 3). It may also be noted that the 'balance of population' varies between 0 and 98,586 by domain. As these persons are available by location and administrative division despite the above issues, the 'balance of population' is 0. In contrast, 98,586 (=85,957+12,629) persons constitute the balance of population in the gender domain but reduce to 85,957 in the domains of religion and age groups. So, the adjustment factors estimated in Table 3.12 should be used along with eq. (3.15) to ensure comparability across domains.

Besides the above procedure to adjust the census counts, there is an alternative, albeit less used, procedure. Under this procedure, the actual adjustment can be made by creating homogeneous adjustment classes with similar coverage rates following

the Chi-square Automatic Interaction Detection (CHAID) technique *à la* Kass (1980) and calculating a common population, undercount rate, and adjustment factor, for each class separately. The national adjusted population can be obtained by summing the adjusted classes. In this process also, only the population within the scope of the PES received adjustment factors. The totals for the 'balance of population' receive an adjustment factor of 1. However, adjustment through the CHIAD technique is beyond the scope of this report.

A minimum domain level is usually required in either of these procedures for the estimates to be statistically reliable. For instance, if an adjustment is made at a lower level of sub-domain, such as location, the technique chosen will not necessarily provide estimates at all lower levels that would sum to adjusted counts (US Bureau of the Census, 1985). Further, population adjustment alone may do more harm than good if other critical variables, e.g., household, remain unadjusted (US Bureau of the Census, 1985). In this case, adjusting population figures alone, leaving households unadjusted, will create a downward bias in household size.

CHAPTER 4: EVALUATION OF CONTENT

4.1. Foundations of Content Error

Content error, also known as response error, is defined as the deviation of the obtained value from the true value for a given characteristic of the unit (person or household). Response error can be divided into response bias (systematic error) and response variance (variable error) depending on whether essential or transient conditions are involved. The content error arises from how the enumerator canvasses the question, how the respondents understand the question or both. Further, in a mass operation like the PHC, the concerned individual does not always provide the data. Errors may also arise because the respondent may not know all the particulars of a person about whom the information is being reported in the PHC. So, the content error arises mainly because (i) either the instructions are not strictly followed or properly understood by the enumerators, (ii) the respondents do not understand the questions asked by the enumerators, or (iii) sometimes the respondents do not provide factual positions intentionally because of some reasons or the other. Errors can also occur due to the difficulty in understanding the concepts and definitions followed in both PHC and PEC by enumerators and respondents. Therefore, some of the sources of non-sampling errors include the sampling frame, non-interview, processing operations, interviewing, respondents, weighting, questionnaire, and reporting (US Bureau of the Census, 1985). Evaluation of census content error involves the estimation of variance and bias components of total error in a census statistic. To that end, consider a characteristic, *j*, say, age or gender, measured both in the PHC and the PEC, and the response to item *j* of any unit in the population. The total error can be modeled as follows:

$$y_{ij} = \mu_{ij} + e_{ij} \qquad \forall\ i = 1, 2, 3, \ldots, n;\ j = 1,2,3, \ldots, C \tag{4.1}$$

where μ_{ij} is the true value for the characteristic i, and e_{ij} is the error committed by enumerator i. Let us further assume that each of the *N* observations in the population responded to the questions related to the characteristic. The structure of e_{ij} is

assumed to be $e_{ij} = b_i + \epsilon_{ij}$ where b_i is the systematic error committed by enumerator i and ϵ_{ij} is the variable error associated with the observation (i, j). Thus, eq. (4.1) can be rewritten as $y_{ij} = \mu_{ij} + b_i + \varepsilon_{ij}$. Insofar as the sample mean is the unbiased estimator of the population mean, one can write $E(\bar{y}) = \mu + B_b + B_\epsilon$, where μ is the true population mean, and B_b and $B_\epsilon$are biases due to the interviewers and the respondents. The corresponding variance for continuous data can be written as:

$$V(\bar{y}) = \left(1 - \frac{n-1}{N-1}\right)\frac{\sigma_\mu^2}{n} + \frac{\sigma_b^2}{k} + \frac{\sigma_\epsilon^2}{n} \quad (4.2)$$

The variance for the qualitative data, such as gender, religion, etc., can be analogously defined: the first term in eq. (4.2) is the population variance or systematic variance; the middle term is the enumerator variance or correlated component variance, and the last term is the random error variance. The variance due to correlated errors does not decrease as the sample size increase. Hence, as n $\rightarrow \infty$, $V(\bar{y})$ is dominated by $\frac{\sigma_b^2}{k}$ (US Bureau of the Census, 1985).

As the PEC is a replication, an independent re-interview of a sample from the PHC facilitates estimating the variable error, not the bias. As a corollary of the dual system estimation, the content error analysis measures consistency, not which answers are right or wrong, i.e., it measures how differently answers are reported between the PHC and the PEC. One of the objectives of the PEC is to conduct an assessment of the quality of the particulars recorded in the PHC for the individuals who were enumerated. To ensure comparability between the PEC and the PHC, the same wording, response categories, sequences, and concept definitions were maintained in the PEC.

4.2. Elements of Content Evaluation

Unlike the coverage errors, which are based on both matched and non-matched cases, the content errors are estimated only for the subset of matched cases. Specifically, the following issues must be noted regarding the use of the PEC for the measurement of content error: (i) it is limited to matched cases; (ii) it is limited to the in-scope sub-universe, consisting of dwelling units within in-scope EAs; (iii) the PEC

is not assumed to provide the 'truth'; therefore, only the response variance but not the response bias is measured; and (iv) comparison is of unedited the PEC and the PHC socio-demographic responses.

It may also be noted that the estimated total persons shown in the content analysis do not coincide with the final PHC totals for each of the characteristics considered because (i) they are based on the sample of PHC records in the PEC and are, therefore, subject to sampling variability; (ii) they include only matched cases, not the full sample; (iii) they are unedited while the PHC characteristics are edited; and (iv) they include only the in-scope sub-universe while the final PHC totals include the full universe.

Content variability between the PHC and the PEC is usually measured through four different indicators: the net difference rate, the index of inconsistency (simple and aggregate), the rate of agreement, and the gross difference rate. These measures are presented for the selected person characteristics that more or less remain unchanged between the PHC and the PEC. Wherever applicable, the empirical analysis of these statistics, along with the associated 95% confidence interval, is discussed. While the net difference rate and the simple index of inconsistency are calculated for each of the categories, the aggregate index of inconsistency, the rate of agreement, and the gross difference rate are estimated as a whole.

4.2.1 Net Difference Rate

The net difference rate (NDR) is the difference between the number of cases in the census and the PEC that fall under each response category relative to the total number of matched persons in all response categories. The NDR formula for the i^{th} category is:

$$NDR = \frac{Y_{.i} - Y_{i.}}{n} \times 100 \qquad \forall\ i = 1, 2,3, \ldots, C \qquad (4.3)$$

where $Y_{.i}$ = unweighted number of PHC cases in the i^{th} category, $Y_{i.}$ = unweighted number of PEC cases in the i^{th} category, n = unweighted number of matched cases, and C = total number of response categories for characteristic Y.

If the sign of this index is positive, it would mean that a larger number of persons have been classified in the i^{th} category in the PHC than is the case. If the sign is negative, it would mean that there has been an underestimation of the i^{th} category in the PHC. It may be noted that NDR is a measure of bias only when the reinterview (PEC interview) is considered more accurate than the original response. Following the US Bureau of the Census (1985), the 95% confidence interval (CI) of the net difference rate for category i and $(Y_{.i} - Y_{i.}) \neq 0$ is given as:

$$\frac{(Y_{.i}-Y_{i.}) \pm 2\sqrt{Y_{.i}+Y_{i.}-2Y_{ii}+1}}{n} \times 100 \qquad \forall \text{ i = 1, 2, 3, …, C} \tag{4.4}$$

Sometimes the absolute value of the NDR is normalized by the mean of the population proportion of the respective category to assess the severity of the content error (US Bureau of the Census, 1985).

4.2.2 Index of Inconsistency

The index of inconsistency (II_C) is the relative number of cases for which the response varied between the PHC and the PEC. It is the ratio of the simple response variance to the total variance of the characteristic, including its variability in the population. It is calculated for each response category i according to the following formula:

$$II_C = \frac{(Y_{.i}+Y_{i.}-2Y_{ii})}{\frac{1}{n}[Y_{.i}(n-Y_{i.})+Y_{i.}(n-Y_{.i})]} \times 100 \qquad \forall \text{ i = 1, 2, …, C} \tag{4.5}$$

where Y_{ii} = the number of cases where category i was given as a response in both the PHC and the PEC. The 95% confidence interval of the index of inconsistency for category i is contingent on the magnitude of $\left(\frac{Y_{.i}+Y_{i.}-2Y_{ii}}{n}\right)$. If $\left(\frac{Y_{.i}+Y_{i.}-2Y_{ii}}{n}\right) \leq 0.10$, then the 95% CI of II_C is

$$\frac{(Y_{.i}+Y_{i.}-2Y_{ii}+2) \pm 2\sqrt{Y_{.i}+Y_{i.}-2Y_{ii}+1}}{Y_{.i}\left(1-\frac{Y_{i.}}{n}\right)+Y_{i.}\left(1-\frac{Y_{.i}}{n}\right)} \times 100 \qquad \forall \text{ i = 1, 2, 3, …, C} \tag{4.6}$$

And if $\left(\frac{Y_{.i}+Y_{i.}-2Y_{ii}}{n}\right) > 0.10$, then the 95% CI is

$$\frac{(Y_{.i}+Y_{i.}-2Y_{ii}+2) \pm 2\sqrt{\frac{1}{n}(Y_{.i}+Y_{i.}-2Y_{ii})(n-Y_{.i}-Y_{i.}+2Y_{ii})}}{Y_{.i}\left(1-\frac{Y_{i.}}{n}\right)+Y_{i.}\left(1-\frac{Y_{.i}}{n}\right)} \times 100 \quad \forall \text{ i = 1, 2, 3, …, C} \tag{4.7}$$

4.2.3 *Aggregate Index of Inconsistency*

Aggregate Index of Inconsistency (II_{AG}) is the relative number of cases, the categories of the characteristic taken as a whole, for which the response varied between the PHC and the PEC. The following formula is used to calculate the II_{AG} for all the response categories of the characteristics as a whole:[6]

$$II_{AG} = \frac{(n-\sum_i^c Y_{ii})}{(n-\frac{1}{n}\sum_i^c Y_{.i}Y_{i.})} \times 100 \qquad \forall\ i = 1, 2, 3, \ldots, C \tag{4.8}$$

Similar to the case of the II, the 95% CI for the II_{AG} is contingent on the magnitude of $\left[\frac{n-\sum_{i=1}^c Y_{ii}}{n}\right]$. If $\left[\frac{n-\sum_{i=1}^c Y_{ii}}{n}\right] \leq 0.10$, then the 95% CI of the II_{AG} is given by

$$\frac{(n-\sum_{i=1}^c Y_{ii}+2)\pm 2\sqrt{n-\sum_{i=1}^c Y_{ii}+1}}{(n-\frac{1}{n}\sum_{i=1}^c Y_{.i}Y_{i.})} \times 100 \qquad \forall\ i = 1, 2, 3, \ldots, C \tag{4.9}$$

And if $\left[\frac{n-\sum_{i=1}^c Y_{ii}}{n}\right] > 0.10$, then the 95% CI is given by

$$\frac{(n-\sum_{i=1}^c Y_{ii}+2)\pm 2\sqrt{\frac{1}{n}(n-\sum_{i=1}^c Y_{ii})(\sum_{i=1}^c Y_{ii})}}{(n-\frac{1}{n}\sum_{i=1}^c Y_{.i}Y_{i.})} \times 10 \qquad \forall\ i = 1, 2, 3, \ldots, C \tag{4.10}$$

4.2.4 *Rate of Agreement*

The rate of agreement (RA) is the complement of the gross difference rate. A low rate of agreement indicates a high degree of variability and vice-versa.

$$RA = \frac{\sum_i^c Y_{ii}}{n} \times 100 \qquad \forall\ i = 1, 2, 3, \ldots, C \tag{4.11}$$

4.2.5 *Gross Difference Rate*

The gross difference rate (GDR) is calculated for the characteristics as a whole. It is the number of discrepancies between the PHC and the PEC responses relative to the total number of persons matched. It is equivalent to the sum of off-diagonal cells for all categories or the complement of the sum of the diagonal cells.

$$GDR = \frac{(n-\sum_i^c Y_{ii})}{n} \times 100 \equiv 100 - RA \qquad \forall\ i = 1, 2, 3, \ldots, C \tag{4.12}$$

[6] The US Bureau of the Census (1985) suggested an alternative expression for the II_{AG} in terms of variances as $II_{AG} = \frac{\sigma_b^2+\sigma_\varepsilon^2}{\sigma_\mu^2+\sigma_b^2+\sigma_\varepsilon^2} \equiv \frac{SRV}{SV+SRV}$, where SRV is the simple response variance, and SV is the sample variance with the individual variances, as defined in the text. However, the empirical estimates are based on eq. (4.8) in the text.

The consequent estimate of content errors needs to be compared against some guidelines to conclude regarding the severity of the errors. The UN manual for the PEC does not provide any guidelines to assess the severity of content errors. However, courtesy of the US Bureau of the Census (1985), the excerpt in Table 4.1 provides standards for the interpretation of different content error measures based on the experience in the censuses of the USA. Despite their shortcomings, these ranges have been used as guiding principles to comment severity of the estimates of content errors.

Table 4.1: Standards for the Interpretation of the Different Content Error Measures

Measure	Low	Moderate	High
Index of Inconsistency	<20	20–50	>50
Aggregate Index of Inconsistency	<20	20–50	>50
Absolute value of NDR relative to the population mean of the proportion	<0.01	0.01–0.05	>0.05

Source: US Bureau of the Census (1985).

4.3. Estimates of Content Errors of Persons

Content analysis can be done at two levels: persons and households. Even though household-level characteristics are usually time-invariant, the analysis is confined to only a few person-level attributes that are supposed to remain unchanged between the time of the PHC and the PEC. The person-level analyses include gender, age, relationship to the head of household, marital status, and religion. Some other attributes, such as employment status, literacy, education, and mobile phone use, are also interesting, and, thus, demand analysis of content errors. However, these attributes are available for the segmented population for which net coverage error rates need to be estimated. Therefore, the attributes that involve the segmented population are not analyzed.

4.3.1 Gender

Even though gender is evident from the names of the members, it is sometimes not so apparent. Therefore, both the PHC and the PEC questionnaires asked the gender of each member of the household to assess the accuracy of the reporting of the gender. Table 4.2 presents the measures of the content variations of the gender of the persons between the PHC and the PEC based on the count variations shown in Table C.1 in Appendix C.

Table 4.2: Response Variability Measures for Gender

Age group	Total in PHC	Total in PEC	Total Consistent Cases	Net Difference Rate			Index of Inconsistency	
				Rate (%)	95% CI	|NDR|/P	Rate (%)	95% CI
Female	74,659	74,446	73,454	0.15	[0.08,0.21]	0.00	3.00	[2.87,3.13]
Male	71,859	72,072	70,867	-0.14	[-0.21, -0.08]	0.00	3.00	[2.87,3.13]
Total	146,518	146,518	144,321					

Aggregated Index of Inconsistency (%) = 3.00 with 95% CI: [2.87, 3.13]
Gross Difference Rate (%) = 1.50
Rate of Agreement (%) = 98.50

Note: Estimates are subject to rounding errors.

It is not surprising that content error is trivial in the case of gender. The NDRs range between 0.15% and -0.15% for females and males, and the corresponding ratios of the absolute value of NDR to the mean of the population proportion (p), |NDR|/P, are zero for both females and males. The index of inconsistency for both gender and the 'aggregated index of inconsistency' is only 3%. In harmony with this finding, the rate of agreement is as high as 98.50%. In the PEC of the PHC 2011, the index of aggregate inconsistency was 3.96% (Mannan, Sohail, & Bhattacharjee, 2012), which increased from zero in 1981 (US Bureau of the Census, 1985), i.e., the index appears to have an inverted U-shaped trajectory over time. Be that as it may, the difference between the current estimate and that from 1981 or 2011 is significant as the 95% confidence interval of the current estimate does not overlap with the indices estimated in the PECs for the PHCs of 1981 and 2011, respectively. In any case, the low values of content error measures suggest that the inconsistency of gender reporting between the PEC and the PEC is extremely low. These results are not surprising as gender is expected to be reported more or less reliably and consistently both in the PHC and the PEC.

4.3.2 Relationship with the Household Head

Both the PHC and the PEC questionnaire recorded detailed data on the relationship of the other household member with the household head. For analytical convenience, the housemaid was merged with the other non-relatives. Even though the relationship with the household head is apparently a time-invariant variable, the relationship changes with the respondent.

Table 4.3: Response Variability Measures for Relationship with the Household Head

Relationship	Total in PHC	Total in PEC	Total Consistent Cases	Net Difference Rate			Index of Inconsistency	
				Rate (%)	95% CI	|NDR|/P	Rate (%)	95% CI
Head of Household	37,029	37,190	35,577	-0.11	[-0.19, -0.03]	0.00	5.53	[5.33,5.73]
Spouse	29,725	29,710	28,279	0.01	[-0.06,0.08]	0.00	6.07	[5.85,6.30]
Children	55,714	55,626	53,689	0.06	[-0.03,0.15]	0.00	5.74	[5.56,5.92]
Parents	4,612	4,517	3,910	0.06	[0.02,0.11]	0.02	14.80	[14.00,15.64]
Siblings	1,993	1,758	1,316	0.16	[0.11,0.21]	0.11	30.22	[28.46,32.08]
Son/Daughter in Law	5,017	5,087	4,411	-0.05	[-0.10,0.00]	0.01	13.14	[12.43,13.90]
Other Relatives	9,437	10,083	7,858	-0.44	[-0.53, -0.36]	0.06	20.88	[20.21,21.56]
Other Non-Relatives	2,991	2,547	1,968	0.30	[0.25,0.36]	0.12	29.49	[28.04,30.99]
Total	146,518	146,518	137,008					

Aggregated Index of Inconsistency (%) = 8.73 with 95% CI: [8.56, 8.91]
Gross Difference Rate (%) = 6.49
Rate of Agreement (%) = 93.51

Note: Estimates are subject to rounding errors.

Table 4.3 presents the measures of the content variations of the members' relationships with the household head between the PHC and the PEC based on the count variations shown in Table C.2 in Appendix C. The NDR estimates reveal mismatches but cannot point out the directions as the relationships are not continuous or ordinal variables but are categorical variables. But the estimates of the |NDR|/P provide the severity of misreporting between the PHC and the PEC. While relationships termed the head of the household, spouse, and children appear to be reported correctly, there is a moderate degree of misreporting of parents, sons/daughters-in-law, siblings, and other relatives and non-relatives have a high degree of misreporting between the PHC and PEC. The indices of inconsistency lead to a similar conclusion. Given the preponderance of the head of the household, spouse, and children among the members of the households, the aggregate index of inconsistency appears to be low at 8.73%, which is corroborated by the 'rate of agreement' at 93.51%. So, the relationships appear to be reported more or less reliably and consistently between the PHC and the PEC. The low levels of misreporting are acceptable in that with the change in the respondent between the PHC and the PEC, the household head is likely to change, and with that change the members' relationships.

4.3.3 Marital Status

Both the PHC and the PEC questionnaires recorded detailed data on the marital status of each of the members in the household aged 10 years and above. The categories of this variable include: never married, currently married, widow/ widower, divorced, and separated.

Table 4.4: Response Variability Measures for Marital Status

Marital Status	Total in PHC	Total in PEC	Total Consistent Cases	Net Difference Rate			Index of Inconsistency	
				Rate (%)	95% CI	|NDR|/P	Rate (%)	95% CI
Never Married	60,657	60,270	58,115	0.26	[0.17,0.36]	0.01	6.61	[6.42,6.81]
Currently Married	79,067	78,691	75,611	0.26	[0.15,0.37]	0.00	8.97	[8.76,9.20]
Widowed/Widower	5,985	6,559	5,171	-0.39	[-0.46, -0.33]	0.09	18.34	[17.57,19.14]
Divorced	439	531	154	-0.06	[-0.1, -0.03]	0.19	68.48	[63.35,74.01]
Separated	370	467	84	-0.07	[-0.1, -0.03]	0.26	80.16	[74.19,86.6]
Total	146,518	146,518	139,135					
Aggregated Index of Inconsistency (%) = 9.37 with 95% CI: [9.15, 9.59]								
Gross Difference Rate (%) = 5.04								
Rate of Agreement (%) = 94.96								

Note: Estimates are subject to rounding errors.

Table 4.4 presents the measures of content variations of the persons' marital statuses between the PHC and the PEC based on the count variations shown in Table C.3 in Appendix C. Viewed from the NDRs, the married groups of persons appear to be over-reported and the other groups under-reported in the PHC. Further, inspection through the |NDR|/P estimates reveals that the degrees of variation for the married groups are low and very high for the other groups. At the level of each response category, three categories show a low level of inconsistency (index < 20%), and two categories show a high degree (index > 50%) of inconsistency.

Given the preponderance of the married groups in the data, the 'aggregate index of inconsistency' of 'marital status' shows a low level at 9.37%, with a rate of agreement at 94.96%. Thus, marital status is likely to be reported more or less reliably and consistently in the PHC and the PEC. The estimate of the aggregate index of inconsistency from this PHC may be contrasted with that found in the PEC of the Population and Housing Census 2011 and 1981 at 5.64% and 5.70%, respectively (Mannan, Sohail, & Bhattacharjee, 2012; US Bureau of the Census, 1985). Given the increasing trend in the aggregate index of inconsistency, it is recommended to put more emphasis on the definitions of the statuses of divorced and separated when collecting information on marital status in the future PHC and PEC.

4.3.4 *Religion*

Both the PHC and the PEC questionnaire recorded the reported religion of each of the members of a household**.** Table 4.5 presents the measures of content variations of the religion of the persons between the PHC and the PEC based on the count variations shown in Table C.4 in Appendix C. The estimates of the NDRs reveal that Muslims and 'other minority religions are over-reported, while Hindus, Buddhists, and Christians are under-reported. However, a closer look through the |NDR|/P estimates reveals that the degrees of variations are low for both Muslims and Hindus, moderate for Buddhists and Christians, and very high for the 'other minority.' At the level of each response category, four categories show a low degree (index<20%) of inconsistency, and only one category shows a high degree (index>50%) of inconsistency. However, the number of cases in the high category is very few. These findings are corroborated by the aggregate index of inconsistency, which shows a low level of inconsistency or variability (index<20%) for religion. Apart from the case of the 'other religion' category, it can be argued that religion is reported more or less reliably and consistently in both the PHC and the PEC.

Table 4.5: Response Variability Measures for Religion

Religion	Total in Census	Total in PEC	Total Consistent Cases	Net Difference Rate			Index of Inconsistency (%)	
				Rate (%)	95% CI	|NDR|/P	Rate (%)	95% CI
Muslim	131,701	131,659	131,405	0.03	[0.00,0.06]	0.00	2.06	[1.89,2.25]
Hindu	12,460	12,514	12,226	-0.04	[-0.07, -0.01]	0.00	2.28	[2.09,2.49]
Buddhist	1,479	1,496	1,456	-0.01	[-0.02,0.00]	0.01	2.14	[1.66,2.75]
Christian	778	791	732	-0.01	[-0.02,0.01]	0.02	6.73	[5.54,8.18]
Others	100	58	2	0.03	[0.01,0.05]	0.44	97.52	[83.02,114.55]
Total	146,518	146,518	145,821					

Aggregated Index of Inconsistency (%) = 2.57 with 95% CI: [2.39, 2.78]
Gross Difference Rate (%) = 0.48
Rate of Agreement (%) = 99.52

Note: Estimates are subject to rounding errors.

4.3.5 *Age Groups*

In both the PHC and the PEC questionnaires, data on age were reported by the respondent even though age derived from the date of birth of each of the members is preferred over the reported age. However, the latter option is out of reach when the

official birth certificates/NID cards are unavailable, or the respondent refuses to cooperate. Besides, age is reported in full years. These issues have serious implications for the analysis of content errors. For the brevity of analysis, the age data are grouped into five-year intervals, and the last few groups are aggregated due to the paucity of data.

Table 4.6 presents the measures of content variations of the age group of the persons between the PHC and the PEC based on the count variations shown in Table C.5 in Appendix C. The estimates of the NDRs reveal that age has been under-reported in 5-9 to 20-24 age groups, 45-49, 55-59, 65-69, and 75+ groups in the PHC. It appears to be over-estimated in the rest of the groups. Given these biases, the level of |NDR|/P is remarkably small for most of the age groups categories; for 3 of the age groups, the levels of this indicator are distantly below 0.01, and 11 have levels between 0.01 and 0.05, and the level is above 0.05 for the remaining two age groups.

Table 4.6: Response Variability Measures for Age Group

Age group	Total in PHC	Total in PEC	Total Consistent Cases	Net Difference Rate			Index of Inconsistency	
				Rate (%)	95% CI	|NDR|/ P	Rate (%)	95% CI
0-4	13,356	12,882	11,613	0.32	[0.25,0.40]	0.03	12.61	[12.16,13.08]
5-9	13,658	13,765	11,138	-0.07	[-0.17,0.02]	0.01	20.71	[20.14,21.29]
10-14	14,298	14,331	11,389	-0.02	[-0.13,0.08]	0.00	22.65	[22.07,23.25]
15-19	14,495	14,530	10,795	-0.02	[-0.14,0.09]	0.00	28.43	[27.78,29.10]
20-24	12,652	12,737	8,035	-0.06	[-0.19,0.07]	0.01	40.19	[39.36,41.03]
25-29	12,245	12,183	6,895	0.04	[-0.10,0.18]	0.00	47.51	[46.60,48.44]
30-34	10,913	10,830	5,380	0.06	[-0.09,0.20]	0.01	54.56	[53.53,55.61]
35-39	11,724	11,592	5,816	0.09	[-0.06,0.24]	0.01	54.44	[53.45,55.46]
40-44	9,292	9,158	4,076	0.09	[-0.05,0.23]	0.01	59.57	[58.40,60.75]
45-49	7,630	7,941	3,194	-0.21	[-0.34, -0.08]	0.04	62.29	[61.00,63.60]
50-54	7,311	7,070	3,094	0.16	[0.04,0.29]	0.03	59.91	[58.60,61.25]
55-59	5,292	5,639	2,024	-0.24	[-0.35, -0.12]	0.07	65.41	[63.85,67.00]
60-64	5,157	5,126	2,022	0.02	[-0.09,0.13]	0.01	62.88	[61.31,64.49]
65-69	3,461	3,773	1,324	-0.21	[-0.31, -0.12]	0.09	65.00	[63.11,66.94]
70-74	2,511	2,429	916	0.06	[-0.02,0.13]	0.03	63.99	[61.74,66.33]
75+	2,523	2,532	1,524	-0.01	[-0.07,0.06]	0.01	40.40	[38.64,42.24]
Total	146,518	146,518	89,235					

Aggregated Index of Inconsistency (%) = 42.29 with 95% CI: [42.01, 42.56]
Gross Difference Rate (%) = 39.10
Rate of Agreement (%) = 60.90

Note: Estimates are subject to rounding errors.

View from indices of inconsistency, nine groups show a high degree (index>50%), six groups show a moderate degree (20%<index<50%), and only one group shows a

low level of inconsistency or variability (index<20%). These intergroup mismatches are epitomized in the estimates of a moderate level of the aggregate index of inconsistency (42.29%) and a low rate of agreement (60.10%). Consequently, 'age' between the PHC and the PEC may not be reported consistently**.** Even though the aggregate index of inconsistency increased by about six percentage points compared to the results in the PHC 2011 (Mannan, Sohail, & Bhattacharjee, 2012), it is not far from where the full-fledged census in the country started in 1981 (US Bureau of the Census, 1985). The three-month lag between the PHC and the PEC and the use of full years in the age reporting might have contributed to differences in reporting the age of the members.

4.4. Summary of the Content Error Analysis

To measure the variability of responses between the PHC and the PEC, the net difference rate, both in absolute form and normalized by the mean of the population proportion, category-wise and aggregate indices of inconsistency, and the rate of agreement and its complement gross difference rate are used. These indicators are applied to six characteristics of persons, viz. gender, relationship with the household head, marital status, religion, and age groups.

The estimates reveal that the characteristics such as 'gender,' 'relationship with the household head,' 'marital status,' and 'religion' show low levels of variability between the PHC and the PEC. Hence, these characteristics are likely to be measured reliably and consistently between the PHC and the PEC. In contrast, 'age' shows a moderate variability level, which might indicate a need for in-depth probing, for example, using historical landmarks of years.

The above evidence implies that marginal totals for each of the enumerators in the PHC and the PEC are very close for most of the items, indicating that the response variability is random in nature and there is very little systematic bias associated with the reporting of either the group of PHC enumerators or the group of PEC enumerators. This absence of a systematic bias associated with the enumerators is an indication that the PHC procedures in data collection, including training and supervision, were of acceptable quality.

CHAPTER 5: CONCLUDING REMARKS

5.1. Conclusions

The major objectives of the PEC exercise are to assess coverage and content errors and to give credibility to the PHC 2022 results. As the PEC is an integral part of the PHC 2022, policymakers were determined to make it a meaningfully independent assessment.

The broad findings of the PEC suggest a good correspondence between the estimate of the true population and the number of units enumerated in the PHC. The estimates of coverage error rates were looked at through different domains such as location, gender, religion, age group, and administrative division. It was found that these estimates were low and well within acceptable limits. Two factors that seem to have contributed to the low coverage error rates were (i) the listing of all households well in advance of the PHC and (ii) the use of the digital data collection process along with electronic monitoring of data collection. Reconciliation of non-matched PHC persons and households by the research team was greatly eased owing to strenuous jobs with due diligence by the field officials of the BBS. As a result, the incidence of reconciliation was significantly high for non-matched PHC and PEC persons, which helped reduce the extent of coverage error rates.

In the case of content errors, the levels of response variance for different time-invariant characteristics covered in the PHC reveal that enumerator agreements are pretty close across characteristics if one looks into the estimates of net difference rates, category-wise and aggregate indices of inconsistency and rate of agreement. These findings indicate that the response variability is random in nature, and there is very little systematic bias associated with the reporting of either the group of the PHC enumerators or the group of PEC enumerators. The absence of systematic bias associated with the enumerators is an indication that the PHC procedures, including training and supervision, were of acceptable quality.

5.2. Lessons Learned

Several lessons were learned from the current PEC exercise that has been conducted independently for the first time in the country's history of the census. These lessons, however, may not be the panacea but will nevertheless be helpful for better PEC results in the future if the issues are considered seriously.

- As the PEC is an integral part of the PHC, it needs to be ensured that the PEC is conducted after extensive planning, testing of questionnaire and methodology, and procedures well in advance and in tandem with the PHC. If the PEC activities are not planned together with the PHC, it becomes difficult to conduct the survey as soon as the PHC is completed.
- The fieldwork of the PEC needs to be completed within a short period (within recommended three months) after the completion of the PHC to avoid a large distortion in the structure of the population enumerated due to birth, death, migration, mobility, etc.
- The operational timelines must be aligned with census activities to allow for the thorough and timely completion of PEC activities. It would cause delays in implementing different stages of PEC if the contractual agreement with an independent organization, budget, and staffing for different stages of PEC are not planned, secured, and strictly followed.
- The institutional or operational independence of the PEC from the PHC should be maintained at all stages of PEC operations, such as fieldwork, data matching, data management, and field reconciliation. It can be achieved if an organization outside of BBS can be contracted to plan and implement the PEC from starting to the end. Alternatively, the PEC unit in BBS is manned by staff that do not participate in census activities.
- BBS should invest in in-house GIS activities, inter alia, to delineate and update the boundaries of EAs. If necessary, the final demarcation should be reviewed to minimize changes within the EA. Both the PHC and PEC enumerators should be closely monitored so they do not transgress the EA boundaries while collecting data.

- Both the PHC and the PEC should contain information on the name and telephone number of the respondent and the GPS of the respondent's dwelling unit so that matching and reconciliation, if necessary, can be promptly made. The independent organization or the unit that would conduct the PEC should be given detail of household and member-level data of the designated EAs so that it can access the full address of the dwelling unit, the household(s) in it, and the names of the member recorded in the census.
- A pilot test should be conducted to test various steps of the PEC. If the entire operation could not be tested due to a lack of resources, at least the PEC questionnaire, reconciliation, and matching operations should be tested before finalizing operations for the PEC.
- Lastly, the current nomenclature of the 'The Post Enumeration Check' should be changed to 'The Post Enumeration Survey' in the future. This is particularly important as the current nomenclature gives the impression that some mistakes have been made deliberately in the PHC, and it is incumbent on the PEC to unearth them. The current nomenclature negates the premise that neither the PEC nor the PHC data are treated with superiority in the dual system estimation. Besides, the UN and most countries use the nomenclature as the 'Post Enumeration Survey' and not Post Enumeration Check (Dauphin & Canamucio, 1993; UN, 2010).

REFERENCES

BBS. (1977). *Population Census of Bangladesh 1974 - National Volume.* Dhaka: Bangladesh Bureau of Statistics.

BBS. (1984). *Bangladesh Population Census 1981 - Analytical Findings and National Tables.* Dhaka: Bangladesh Bureau of Statistics.

BBS. (1994). *Bangladesh Population Census 1991 - Volume 1.* Dhaka: Bangladesh Bureau of Statistics.

BBS. (2003). *Population Census 2001 - National Report (Provisional).* Dhaka: Bangladesh Bureau of Statistics.

BBS. (2022). *Population and Housing Census 2022: Preliminary Report.* Dhaka: Bangladesh Bureau of Statistics.

Chandrasekaran, C., & Deming, W. E. (1949). On a Method of Estimating Birth and Death Rates and the Extent of Registration. *Journal of American Statistical Association, 44*(245), 101-115.

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). Nashville, TN: John Wiley & Sons.

Dauphin, M., & Canamucio, A. (1993). *Designing and Implementation of Post-Enumeration Survey: Developing Country Example.* Washington DC: US Bureau of Census.

Endlich, R. M., Eynon, B., Ferek, R., Valdes, A., & Maxwell, C. (1988). Statistical Analysis of Precipitation Chemistry Measurements over the Eastern United States. Part I: Seasonal and Regional Patterns and Correlations. *Journal of Applied Meteorology, 27*(1322), 1322-1333.

Gatz, D. F., & Smith, L. (1995). The Standard Error of a Weighted Mean Concentration—I. Bootstrapping vs Other Methods. *Atmospheric Environment, 29*(11), 1185–1193.

Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics, 29*(2), 119-127.

Mannan, M. A., Sohail, M., & Bhattacharjee, D. P. (2012). *Report of the Post Enumeration Check (PEC) of the Population and Housing Census, 2011.* Dhaka: Bangladesh Institute of Development Studies.

Mule, T. (2012). *Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States.* Washington DC: United States Department of Commerce.

UN. (2010). *Post Enumeration Surveys: Operational Guidelines.* Department of Economic and Social Affairs. New York: United Nations.

US Bureau of the Census. (1979). *Popstan: A Case Study for the 1980 Censuses of Population and Housing.* Washington DC: US Bureau of the Census.

US Bureau of the Census. (1985). *Evaluating Censuses of Population and Housing.* Washington DC: US Bureau of the Census.

APPENDIX A: QUESTIONNAIRE OF THE POST ENUMERATION CHECK

শুমারি পরবর্তী যাচাই জরীপ প্রশ্নপত্র
বাংলাদেশ উন্নয়ন গবেষণা প্রতিষ্ঠান (বিআইডিএস)
ই-১৭, আগারগাঁও, শেরে বাংলা নগর, ঢাকা-১২০৭
পরিকল্পনা মন্ত্রণালয়

গোপনীয়

বিভাগ☐☐ জেলা☐☐ ডিসিসি☐☐ ইউসিসি☐☐ উপজেলা/থানা☐☐ জোন☐☐ ইউনিয়ন/ওয়ার্ড☐☐☐ মৌজা☐☐☐ গ্রাম/মহল্লা ☐☐☐ আরএমও☐ সুপারভাইজার☐☐ তথ্যসংগ্রহকারী☐☐☐ গণনা এলাকা☐☐☐

খানা মডিউল

1(a). পিইসি খানার ক্রমিক নম্বর	1(b). শুমারি খানার ক্রমিক নম্বর	2. খানার ঠিকানা: ফ্ল্যাট নম্বর/ হোল্ডিং/ বাড়ির নম্বর/ প্রতিষ্ঠানের নাম: (ফ্ল্যাট/হোল্ডিং/বাড়ির নম্বর না থাকলে, বাড়ির নাম লিখুন। প্রতিষ্ঠান হলে, প্রতিষ্ঠানের নাম অবশ্যই লিখুন)	3a. বসবাসের ধরন ① বস্তি [➪Q4] ② ভাসমান/ছিন্নমূল ③ বস্তি নয় [➪Q4] 3b. আপনার অন্য কোথাও বসতঘর আছে কি? ① হ্যাঁ ② না [➪Q11] * ভাসমান পিইসি গণনাযোগ্য	4. খানার প্রকার ① সাধারণ ② প্রাতিষ্ঠানিক ③ অন্যান্য প্রাতিষ্ঠানিক ও অন্যান্য হলে, Q21 থেকে শুরু	5a. খানার প্রধান বসতঘরের মেঝের উপকরণ ① সিমেন্ট/কংক্রিট/ ইট/পোড়ামাটি ② টাইলস/মোজাইক ③ কাঠ/বাঁশ/চটাই/তালগাছ/ সুপারিগাছের গুড়ি ④ মাটি/বালি/কাদা ⑤ অন্যান্য উপকরণ	5b. খানার প্রধান বসতঘরের দেয়ালের উপকরণ ① সিমেন্ট/কংক্রিট/ইট/পোড়ামাটি ② সিরামিক টাইলস ③ ধাতব টিন/সিআই শিট/ঢেউটিন ④ কাঠ/বাঁশ/চটাই/তালগাছ/ সুপারি গাছের গুড়ি ⑤ খড়/বন/গোলপাতা/ তালপাতা/ছিটানি/পলিথিন ⑥ মাটি/কাদা মাটি ⑦ অন্যান্য উপকরণ ⑧ দেয়াল নেই	5c. খানার প্রধান বসতঘরের ছাদের/ছাউনির উপকরণ ① সিমেন্ট/কংক্রিট/টালি ② ধাতব টিন/সিআই শিট/ ঢেউটিন ③ কাঠ/বাঁশ/চটাই/খড়/বন/ ছিটানি/ গোলপাতা/ তালপাতা/ পলিথিন ④ হাতে তৈরী টাইলস/টালি ⑤ অন্যান্য উপকরণ ⑥ ছাদ/ছাউনি নেই	6a. খানায় বাসগৃহের সংখ্যা [একই বাসগৃহে একাধিক খানা থাকলে, সকল খানার ক্ষেত্রে 0 লিখুন, শেষ খানার ক্ষেত্রে 99 লিখুন; বাসগৃহ ভাগাভাগি করে খানা খানার পৃথক বাসগৃহ থাকলে, তার সংখ্যা 0 ও 99-এর সাথে যোগ করে লিখুন] 6b. ভবনের মোট তলার সংখ্যা গৃহ-১ গৃহ-২ গৃহ-৩ গৃহ-... গৃহ-৯	7. বাসগৃহের মালিকানা ① নিজস্ব ② ভাড়া- কিন্তু, অন্য স্থানে নিজস্ব বাসগৃহ আছে ③ ভাড়া এবং অন্য কোথাও নিজস্ব বাসগৃহ নেই ④ বিনা ভাড়া- কিন্তু অন্য স্থানে নিজস্ব বাসগৃহ আছে ⑤ বিনা ভাড়া এবং অন্য কোথাও নিজস্ব বাসগৃহ নেই	8. খানায় খাবার পানির প্রধান উৎস ① ট্যাপ/পাইপ (সাপ্লাই) ② টিউবওয়েল (গভীর/অগভীর) ③ বোতলজাত পানি/পানির জার ④ কূপ/কুয়া ⑤ পুকুর/নদী/ খাল/লেক ⑥ ঝরনা/ছড়া ⑦ বৃষ্টির পানি ⑧ অন্যান্য	9a. টয়লেট সুবিধা ① ফ্লাশ করে/পানি ঢেলে নিরাপদ নিষ্কাশন ② ফ্লাশ করে/পানি ঢেলে অনিরাপদ নিষ্কাশন ③ স্লাবসহ পিট ল্যাট্রিন/ ভেন্টিলেটেড ইমপ্রুভড ল্যাট্রিন/ কম্পোস্টিং ল্যাট্রিন ④ স্লাব ছাড়া পিট ল্যাট্রিন/উন্মুক্ত পিট ⑤ কাঁচা/খোলা/ঝুলন্ত ল্যাট্রিন (স্থায়ী/অস্থায়ী) ⑥ খোলা জায়গা/ ল্যাট্রিন নেই [➪Q9c]	9b. টয়লেট ব্যবহারের ধরন ① একক ② যৌথ 9c. সাবান ও পানিসহ হাতধোয়ার পৃথক ব্যবস্থা (টয়লেট ব্যতীত) কী? ① সাবান-পানি ② শুধু সাবান ③ শুধু পানি ④ কোনোটিই নেই/ পৃথক ব্যবস্থা নেই	10a. খানায় বিদ্যুৎ সুবিধার প্রধান উৎস ① জাতীয় গ্রিড ② সৌরবিদ্যুৎ ③ অন্যান্য ④ বিদ্যুৎ সুবিধা নেই 10b. রান্নার জ্বালানির প্রধান উৎস ① খড়/পাতা/তুষ/ঘুঁটি ② কাঠ/খড়ি/লাকড়ি ③ কাঠ-কয়লা/ চারকোল/ শুকনো গোবর ④ কেরোসিন/প্যারাফিন ⑤ পেট্রোল/ডিজেল ⑥ বিদ্যুৎ ⑦ সাপ্লাই গ্যাস ⑧ এলপিজি গ্যাস ⑨ বায়োগ্যাস ⑩ অন্যান্য	11. খানায় কোনো ক্ষুদ্র নৃগোষ্ঠীর সদস্য আছে কি? ① হ্যাঁ [Q33 পূরণ হবে] ② না [Q33 পূরণ হবে না]	12. খানায় কোনো বিদেশি নাগরিক (Foreigner) আছে কি? ① হ্যাঁ [Q34 পূরণ হবে] ② না [Q34 পূরণ হবে না]	13. খানায় কী ধরনের অর্থনৈতিক কর্মকাণ্ড পরিচালিত হয়? ① কৃষি ② অকৃষি ③ কৃষি-অকৃষি উভয়ই ④ কোনোটিই হয় না 14. গত ২ (দুই) বছরে খানায় কোনো বৈদেশিক রেমিট্যান্স (অর্থ/পণ্য) গ্রহণ করা হয়েছে কি? ① হ্যাঁ ② না
☐☐☐		☐	☐	☐	☐	☐	☐	☐ ☐	☐	☐	☐	☐	☐ ☐	☐	☐	☐ ☐

15	a. পিইসি রাতে খানায় অবস্থানকারী সদস্য (আত্মীয়/অনাত্মীয়সহ) সংখ্যা b. পিইসি রাতে খানায় অনুপস্থিত (ভ্রমণরত/ডিউটিরত) সদস্য সংখ্যা b1. পিইসি রাতে অনুপস্থিত কিন্তু শুমারি রাত্রিতে গণনাকৃত সংখ্যা	e. খানায় অন্তর্ভুক্ত মোট সদস্য সংখ্যা (a+b+b1) ☐	d. খানার কতজন 'বাংলাদেশি নাগরিক' সদস্য বর্তমানে বিদেশে থাকেন? ☐ e. গত ২ (দুই) বছরে খানার কতজন সদস্য বিদেশ হতে স্থায়ীভাবে ফেরত এসেছেন? ☐

ব্যক্তি মডিউল

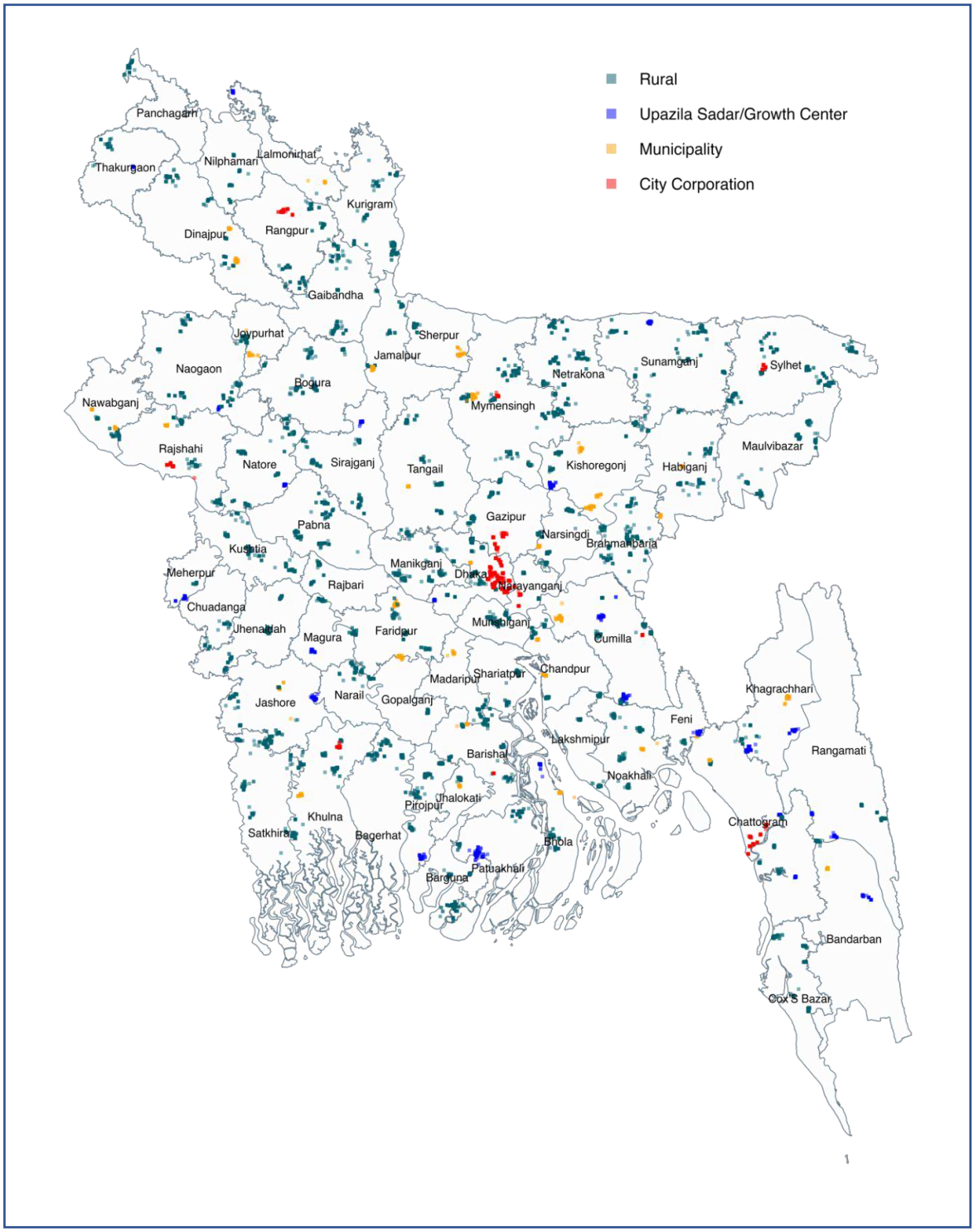
16. সদস্যের ক্রমিক নম্বর	17. [খানা প্রধান থেকে শুরু করুন এবং ইংরেজি CAPITAL LETTER-এ পূর্ণনাম লিখুন]	17 (a) পিইসি রাত্রিতে কোথায় ছিলেন 1. এই খানায় 2. অন্যত্র 3. জন্ম নিয়েছে	17 (b) শুমারি রাত্রিতে কোথায় ছিলেন 1. এই খানায় 2. অন্যত্র 3. জন্ম হয়নি	17 (c) শুমারিতে গণনাকৃত কিনা? 1. হ্যাঁ 2. না 3. জানিনা	17 (d) কোথায় গণনাকৃত 1. এই খানায় 2. অন্যত্র	18. বয়স (পূর্ণ বছরে)	19. লিঙ্গ ① পুরুষ ② মহিলা ③ হিজড়া	20. খানাপ্রধানের সাথে সম্পর্ক ① খানাপ্রধান ② স্ত্রী/স্বামী ③ সন্তান ④ মা/বাবা ⑤ ভাই/বোন ⑥ পুত্রবধূ/জামাতা ⑦ অন্যান্য আত্মীয় ⑧ গৃহকর্মী ⑨ অন্যান্য অনাত্মীয় * লিঙ্গের (Q19) ভিত্তিতে সম্পর্ক প্রদর্শন নির্ধারিত হবে	21. বৈবাহিক অবস্থা ① অবিবাহিত ② বর্তমানে বিবাহিত ③ বিধবা/বিপত্নীক ④ তালাকপ্রাপ্ত ⑤ দাম্পত্য বিচ্ছিন্ন	22. ধর্ম ① মুসলমান ② হিন্দু ③ খ্রিষ্টান ④ বৌদ্ধ ⑨ অন্যান্য	23. প্রতিবন্ধী হলে, তার ধরন? [একাধিক উত্তর হতে পারে] ☐ 1-অটিজম ☐ 2-শারীরিক ☐ 3-মানসিক ☐ 4-দৃষ্টি ☐ 5-বাক ☐ 6-বুদ্ধি ☐ 7-শ্রবণ ☐ 8-সেরিব্রাল পালসি ☐ 9-ডাউন সিনড্রোম ☐ 10-অন্যান্য ⓪ প্রতিবন্ধী নয়
☐☐☐	☐	☐	☐	☐		☐		☐	☐	☐	☐☐☐☐☐☐

24. পড়তে ও লিখতে পারেন কি? ① পড়তে ও লিখতে পারেন ② শুধু পড়তে পারেন ③ কোনোটিই পারেন না	25. বর্তমানে শিক্ষার্থী কি? ① হ্যাঁ ② না	26. সর্বোচ্চ শ্রেণি পাশ (তালিকা থেকে নির্বাচন করুন)	27. পাশের ক্ষেত্র ① সাধারণ ② কারিগরি ③ ধর্মীয় ④ অন্যান্য	28. কাজের মর্যাদা ① কাজে নিয়োজিত ② গৃহস্থালি কাজ/ গৃহকর্ম [➪Q30] ③ কাজ খুঁজছে [➪Q30] ④ কাজ করে না [➪Q30]	29a. কর্মরত হলে, কাজের ধরন ① বেতন/মজুরির বিনিময়ে ② মুনাফার উদ্দেশ্যে ③ খানার ব্যবহার ভোগের জন্য ④ শিক্ষানবিস 29b. কর্মরত হলে, কাজের ক্ষেত্র ① কৃষি ② শিল্প ③ সেবা	30. বর্তমানে কোনো প্রশিক্ষণে নিয়োজিত আছেন কি? ① হ্যাঁ ② না	31a. নিজস্ব ব্যবহারের মোবাইল ফোন আছে কি? ① হ্যাঁ ② না 31b. গত ০৩ মাসে একবারও ইন্টারনেট (ফেসবুক, ইউটিউব, হোয়াটসঅ্যাপ, ভাইবার, ইমু, জুম, অনলাইন পত্রিকা পড়া, ইমেইল, ইত্যাদি) ব্যবহার করেছেন কি? ① হ্যাঁ ② না	32a. ব্যাংক/বিমা/ ক্ষুদ্রঋণ/ ডাকঘর সঞ্চয় অ্যাকাউন্ট আছে কি? ① হ্যাঁ ② না 32b. মোবাইল ব্যাংকিং অ্যাকাউন্ট (বিকাশ/ নগদ/ রকেট/ইত্যাদি) আছে কি? ① হ্যাঁ ② না	33. ক্ষুদ্র নৃগোষ্ঠীর নাম নির্বাচন করুন [Q11 হ্যাঁ হলে, তালিকা হতে নির্বাচন করুন]	34. জাতীয়তা কী? ① বাংলাদেশি [➪Q35a] ② বিদেশি [➪Q35b]	35a. বাংলাদেশি হলে, নিজ জেলা [তালিকা হতে নির্বাচন করুন] 35b. বিদেশি হলে, দেশের নাম [তালিকা হতে নির্বাচন করুন]
০৫ বছর ও তদূর্ধ্ব বয়সের সদস্যদের জন্য								১৫ বছর ও তদূর্ধ্ব ব্যক্তিদের জন্য			
☐	☐	☐	☐	☐	☐ ☐	☐	☐	☐	☐	☐	☐

* 17(b) এর উত্তর 3 হলে 18 তে যাবে, 17(c) এর উত্তর 2 বা 3 হলে 18 তে যাবে

APPENDIX B : LOCATION OF THE PEC ENUMERATION AREAS

Figure B.1: Spatial Distribution of the Enumeration Areas



APPENDIX C : SUPPLEMENTARY TABLES ON CONTENT ANALYSIS

Table C.1: Counts of Gender in the PHC and the PEC

PHC	PEC		
	Male	Female	Total
Male	70,867	992	71,859
Female	1,205	73,454	74,659
Total	72,072	74,446	146,518

Table C.2: Counts of Relationship with the Household Head in the PHC and the PEC

PHC	PEC								
	Head	Spouse	Son/ Daughter	Parents	Brother/ Sister	Son/ Daughter-in-law	Other relative	Other non-relative	Total
Head	35,577	690	257	166	61	41	76	161	37,029
Spouse	776	28,279	162	125	36	203	81	63	29,725
Son/Daughter	293	174	53,689	87	184	268	890	129	55,714
Parents	166	129	85	3,910	9	25	281	7	4,612
Brother/Sister	66	38	207	18	1,316	11	313	24	1,993
Son/Daughter-in-law	44	218	166	13	6	4,411	149	10	5,017
Other relative	63	77	854	179	110	111	7,858	185	9,437
Other non-relative	205	105	206	19	36	17	435	1,968	2,991
Total	37,190	29,710	55,626	4,517	1,758	5,087	10,083	2,547	146,518

Table C.3: Counts of Marital Status in the PHC and the PEC

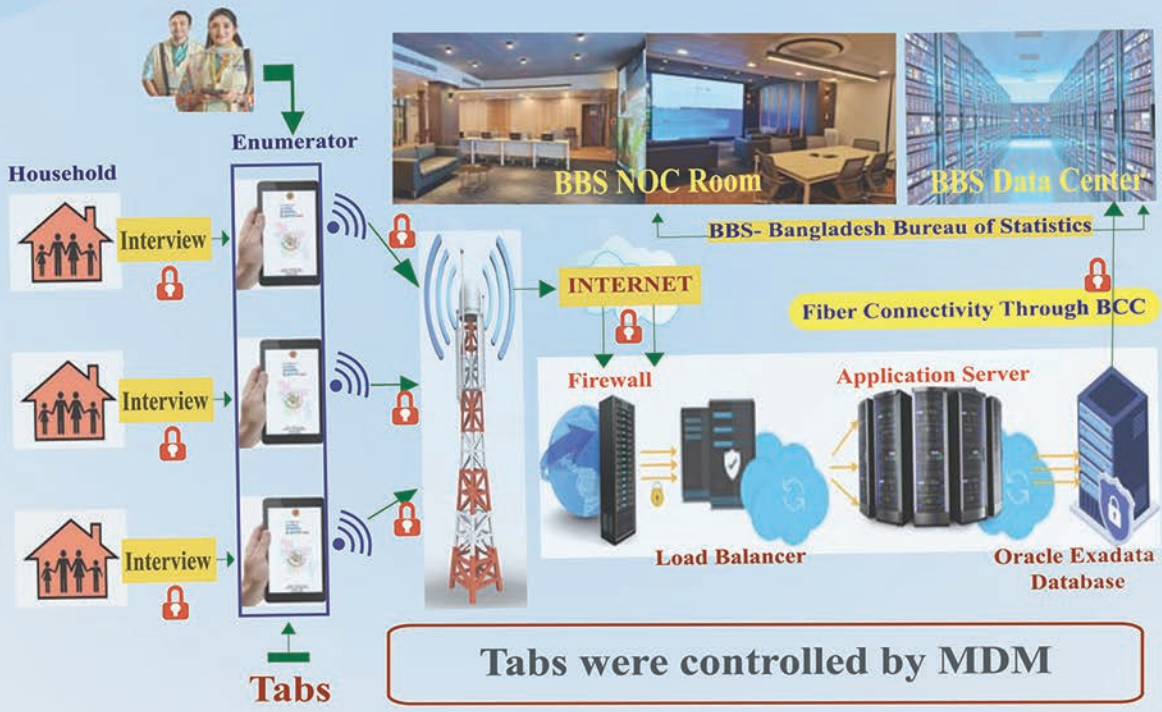
PHC	PEC					
	Never Married	Currently Married	Widow/ Widower	Divorced	Separated	Total
Never Married	58,115	2,269	129	95	49	60,657
Currently Married	1,945	75,611	1,132	160	219	79,067
Widow/ Widower	101	582	5,171	66	65	5,985
Divorced	73	101	61	154	50	439
Separated	36	128	66	56	84	370
Total	60,270	78,691	6,559	531	467	146,518

Table C.4: Counts of Religion in the PHC and the PEC

PHC	PEC					
	Islam	Hinduism	Christian	Buddhism	Others	Total
Islam	131,405	210	17	29	40	131,701
Hinduism	190	12,226	26	6	12	12,460
Christian	11	30	732	5	0	778
Buddhism	5	4	10	1,456	4	1,479
Others	48	44	6	0	2	100
Total	131,659	12,514	791	1,496	58	146,518

Table C.5: Counts of Age Group in the PHC and the PEC

PHC	PEC																
	0-4	**5-9**	**10-14**	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**	**65-69**	**70-74**	**75+**	**Total**
0-4	11,613	1,344	119	74	61	43	22	18	12	6	9	8	8	9	4	6	13.356
5-9	872	11,138	1,365	125	44	40	19	10	8	3	10	7	4	4	3	6	13,658
10-14	114	1,033	11,389	1,525	110	37	28	15	9	6	5	3	8	6	3	7	14,298
15-19	81	93	1,216	10,795	1,876	245	55	27	22	20	19	11	7	9	5	14	14,495
20-24	52	37	114	1,629	8,035	2,140	341	123	48	39	25	23	21	12	10	3	12,652
25-29	47	38	35	213	1,983	6,895	2,149	601	123	68	26	22	18	11	8	8	12,245
30-34	20	25	16	57	326	1,939	5,380	2,350	513	157	54	25	27	11	8	5	10,913
35-39	16	15	19	23	139	568	2,106	5,816	2,043	674	168	67	35	16	7	12	11,724
40-44	14	10	12	11	48	115	448	1,803	4,076	1,849	578	187	86	32	13	10	9,292
45-49	4	5	8	18	27	54	143	566	1,594	3,194	1,257	498	162	70	23	7	7,630
50-54	7	6	10	19	21	37	56	146	438	1,289	3,094	1,334	590	181	57	26	7,311
55-59	6	4	6	9	9	18	32	47	167	421	1,152	2,024	892	357	98	50	5,292
60-64	9	5	7	4	12	24	23	31	63	135	460	961	2,022	961	312	128	5,157
65-69	9	3	2	2	11	15	10	21	15	46	138	342	789	1324	517	217	3,461
70-74	8	6	5	10	14	5	5	9	16	20	53	92	324	519	916	509	2,511
75+	10	3	8	16	21	8	13	9	11	14	22	35	133	251	445	1,524	2,523
Total	12,882	13,765	14,331	14,530	12,737	12,183	10,830	11,592	9,158	7,941	7,070	5,639	5,126	3,773	2,429	2,532	146,518



Bangladesh Institute of Development Studies (BIDS)

E-17, Agargaon, Sher-e-Bangla Nagar, Dhaka 1207

www.bids.org.bd