Survival Analysis of Infants in Small Areas Using Some Indirect Methods.

By

Akhtar Hussain

Registration No: 00000365295 MS 2021-23

Supervisor

Dr. Shakeel Ahmed

Department of Statistics

School of Natural Sciences (SNS), National University of Sciences and Technology (NUST)

Islamabad, Pakistan

September (2023)

Survival Analysis of Infants in Small Areas Using Some Indirect Methods.

By

Akhtar Hussain

Registration No: 00000365295 MS 2021-23

Supervisor

Dr. Shakeel Ahmed

Co-supervisor

Dr. Tahir Mehmood and Dr. Firdos Khan

A thesis submitted in conformity with the requirements for the degree of masters of Science in Statistics

Department of Mathematics and Statistics, School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), Islamabad, Pakistan

September 2023

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by **Akhtar Hussain** (Registration No. **00000365295**), of **School of Natural Sciences** has been vetted by undersigned, found complete in all respects as per NUST statutes/regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/M.Phil degree. It is further certified that necessary amendments as pointed out by GEC members and external examiner of the scholar have also been incorporated in the said thesis.

Signature: ____________

Name of Supervisor: Dr. Shakeel Ahmed

Date: 07/09/2023

Signature (HoD): ____________

Date: 7/9/2023

Signature (Dean/Principal): ____________

Date: 08.09.2023

FORM TH-4

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: **"Akhtar Hussain"** Regn No. **00000365295** Titled: **"Survival Analysis of Infants in Small Areas Using Some Indirect Methods. "** accepted in partial fulfillment of the requirements for the award of **MS** degree.

Examination Committee Members

1. Name: PROF. TAHIR MAHMOOD Signature: ____________

2. Name: DR. FIRDOS KHAN Signature: ____________

Supervisor's Name: DR. SHAKEEL AHMED Signature: ____________

Head of Department

7/9/23
Date

COUNTERSINGED

Date: 08.09.2023

Dean/Principal

Declaration

Akhtar Hussain declare that this thesis is titled **Survival Analysis of Infants in Small Areas Using Some Indirect Methods."** and the work that is given in it is my own and was created by me as a consequence of my own unique research. I confirm that:

1. This work was completed entirely or primarily while pursuing a Master of Science degree at the National University of Science and Technology (NUST), Islamabad, Pakistan.

2. Any instance in which any portion of this thesis has already been submitted to the National University of Science and Technology, Islamabad, Pakistan(NUST or another school for a degree or other qualification has been made clear.

3. In cases when I have consulted other people's published work, this is always explicitly attributed.

4. I always cite the source when I quote from someone else's writing. This thesis is entirely my own work, with the exception of such quotes.

5. I have thanked the primary sources of my assistance.

Akhtar Hussain,

00000365295 MS Statistics 2021

Copyright Notice

This thesis is dedicated to ***my beloved parents.***

Abstract

To ensure the well-being of the masses in a country, policymakers need a clear picture of health indicators at a more granular level. They need to know how to overcome mortality problems related to infants after their birth in the country. This can be done by estimating health outcomes at a small level using data from health surveys and combining them with administrative records. The study's main objective is to suggest some improved estimators of survival function in small areas using direct estimators of survival function, indirect estimators of survival function, synthetic estimators of survival function, composite estimators of survival function, and ratio estimators of survival function, and to obtain separate estimation methods for infant mortality in different sub-populations according to demographic sub-populations demographic characteristics. This study will adopt indirect such as the synthetic estimation method, composite estimation method, and Time-Space Nearest Neighborhood Estimation method to obtain indirect estimates for infant mortality. This study will help public health policymakers make policies for small areas where the required data is limited. Whoever is working on the infant mortality rate, be it the government or NGOs, may get a clearer picture of the mortality rate of infants using the adopted method.

Keywords: *Indicator function, Infant mortality, Granular level, Survival function, Indirect methods, Time-Space Nearest Neighborhood*

Acknowledgments

I want to express my gratitude to God for guiding me through all of my challenges. I have been following your advice every day. You are the one who permitted me to complete my degree. I will continue to put my destiny in your capable hands. Additionally, I want to express my gratitude to my parents and my entire family for their unwavering support and tolerance as I conducted my study and wrote my project. Your supplication for me has kept me going so far. Furthermore, I want to express my gratitude and acknowledgment to my supervisor (Dr. Shakeel Ahmed), who made this work feasible and easier for me, he was the main source of my research, and without his support and guidance, I would never have been able to complete my research. Due to his advice and counsel, I could complete my project. Additionally, I want to thank the members of my committee (Dr. Tahir Mehmood and Dr. Firdos Khan) for making my defense a fun experience and for their insightful remarks and suggestions. Finally, I am in debt of all the sources that I have mentioned in my thesis, I have learned a lot of knowledge from their work, and their work made me able to complete my research.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

SAE	Small area estimation
NGO	Non Government Organization
SYN	Synthetic
CDF	Cumulative distribution function
ADJ	Adjusted
SF	Survival function
TSNN	Time-Space Nearest Neighborhood
IMR	Infant Mortality Rate
SDG's	Sustainable development goals
RE	Relative efficiency
PCoB	Percentage contribution of bias

Chapter 1

Introduction and Motivation

1.1 Background of the study

Information about the population of interest is required when the government is interested in making different policies for the country. Then only taking into consideration the large areas is not a good technique, as there might be a chance of mis-consideration of some areas in great need of the policies. Therefore, the government needs information at more granular administrative levels for better policy-making. For example, while making policies to distribute development funds for areas in need, the government needs granular level data on those needy areas. Furthermore, while addressing health-related problems in different areas, authorities need information about health indicators like the prevalence of disease, vaccine coverage, etc., at granular levels. In the scenario of vaccination, in developed areas, it's easy for individuals to get the facility easily but in backward areas, every citizen can't get vaccinated, so information on more granular levels is necessary to ensure that no one in the population is devoid of getting the facility. Here, thorough knowledge of health indicators in granularity is essential for addressing health-related problems.[1]

To plan, analyze, and evaluate development projects across various sectors, statistical data is vital for policy-making. Particularly when establishing programs to reduce poverty, a comprehensive dataset is essential for well-informed policy formulation. A complete analysis must include precise information on numerous areas to solve this difficulty efficiently. These include things like household income levels, the extent of access to needs like food and shelter, the accessibility of educational possibilities, and the job status of family members. The creation of such a dataset provides policymakers with a thorough grasp of the many facets of poverty, allowing them to make well-informed decisions. Based on this knowledge, policies, and interventions can be developed to successfully reduce poverty and strengthen vulnerable communities. [2]

The same dataset also takes a double significance because it may be used to assess the effectiveness of the implemented program. Policymakers and stakeholders can determine the program's actual impact on the target population by comparing data from before and after implementation. This evaluation offers insightful information about the program's efficacy, enabling iterative changes and deliberate improvements to be made, resulting in a more effective effort to reduce poverty. [2]

The Millennium Development Goals (2000-2015) show how important data is to policy-making on a larger scale. These objectives emphasized the value of statistics in comprehending the top priorities for national and international development. The use of statistical indicators helped to track development, pinpoint problem areas, and direct policy choices. Since then, national statistics agencies and the global data system have strengthened efforts to use data to track development targets and ensure informed, timely, and accurate policy inputs.[2]

The Sustainable Development Goals (SDGs) were introduced in 2015. They offered a worldwide framework for building on the accomplishments of the Millennium Development Goals (MDGs) and addressing the social, economic, and environmental aspects of sustainable development. The SDGs advocate data disaggregation based on sex, income, age, race, culture, resettlement status, geographic area, and other pertinent parameters to ensure inclusivity and effectiveness. This need for more precise data granularity identifies particular sub-populations that might be neglected as development advances, allowing for more equitable and targeted policy-making. However, the requirement for data disaggregation creates difficulties in the data collection procedure. Extensive data collection, which can be expensive, time-consuming, resource-intensive, and dependent on specialized workers, is necessary to reflect segments of the population accurately. For many data collection agencies, conducting large-scale surveys for disaggregation purposes may not be financially feasible. Alternative methods i.e. small area estimate, which incorporates many data sources, should therefore be considered as more economically viable ways to improve data granularity. [2] Such methods give decision-makers access to a more thorough and nuanced understanding of various population groups, enabling the creation of policies that cater to particular needs and encourage inclusive growth.[2]

In the following subsection, we elucidate the key terminologies employed within the scope of this study.

1.2 Definiton of terminologies

This subsection will encompass essential terminologies utilized within the framework of this study.

1.2.1 Small area estimation

Small area estimation is a technique to obtain more precise estimates for areas or sub-populations of a large population, where obtaining direct estimates for the sub-population is not possible because of a very small number of observations. Small area estimation helps to provide accurate estimates in this scenario. It is a very important technique that helps in obtaining estimates for different indicators, for which data is not available at a small level.[3]

1.2.2 Direct and Indirect Estimator

If an estimator is obtained for a specific variable of interest to the population by taking a sample from that population, such an estimator is called a direct estimator.[3]

Precise direct estimates for a small area cannot be possible because of the small sample size, The problem of intolerable standard error takes place, so it is necessary to burrow more information about the variable of interest from related areas with similar characteristics and increasing size. Such estimators are called indirect estimators.[5]

1.2.3 Design based estimator

To make the sample more representative of the population, design-based estimators take into consideration survey weights and conclude the population using a probability distribution obtained by sampling design. In such estimators, the population is held fixed.[4]

1.2.4 Survival Analysis

Survival analysis is a technique that is used for data analysis where the response or the variable of interest is the waiting time for an event to occur. The time in survival analysis can be in years, months, weeks, and days from the start of supervising an individual to the extent when the event occurs. It can also be referred to as the age of the supervised individual when the event just occurs. [6]

“Survival time” can be also used occasionally for the time variable in survival analysis as it denotes the point in time when an individual has survived throughout a specific follow-up duration. The time variable in survival analysis refers to the event as a failure since the event of focus is death, the emergence of illness, or some other negative individual experience. On the other hand, survival time can be "time to resume employment following an elective medical procedure," in which case failing is desirable. [6]

1.3 Infants Mortality

Infant deaths between the ages of 1 day and 1 year are referred to as infant mortality. [7]

1.3.1 Neonatal mortality

The likelihood of passing away before age one. [7]

1.3.2 Infant mortality

The likelihood of passing away before turning one year old. [7]

1.3.3 post-neonatal mortality

. The distinction between newborn and neonatal mortality is Under-five mortality is the likelihood that a child will pass away before turning five. [7]

1.3.4 Child mortality

Is the likelihood that a child will pass away between the ages of 1 and 5? [7]

1.3.5 Terminology and notation for survival analysis

We will discuss some of the basic terminologies and notations related to survival analysis. Firstly, an individual's survival time or the age of the individual at death by a random variable capital T. As T refers to the survival time, time cannot be negative, it means T can take all those whole numbers i.e. value that is greater than or equal to 0. [6] In addition, if we are interested in the survival analysis for a specific time for the random variable T, it will be denoted by lowercase t. For example, if we are interested in knowing whether an individual survives more than 1-year supervised heart disease treatment, then, in this case, our lowercase t will be 1, and we inquire whether T is greater than equal to 1 or not. Finally, a Greek alphabet (δ) will be assigned to the random variable whose values are between 0 and 1. This indicates the process of failure or censorship. [6] Prominently, if the targeted event happens which is a failure, during the study period then (δ) will take the value 1, and if the survival time is censored (the event does not occur during the study period or the survival time is censored at the end of study) then the (δ) will take value 0. Censorship is the only thing that remains for survival time if the targeted event (failure) does not occur during the study period. Censorship takes place if and only if one of the following conditions is fulfilled: an individual survives throughout the study period, is unavailable for follow-up, or leaves the study during the study period. Further, we will discuss the most important concepts which are very important in any survival analysis. The most important term is the survival function which is denoted by S(t). [6] The probability that an individual will survive after a specific time t, will be given to us by S(t). In other words, the probability that the random variable T exceeds the specific time t will be explained by S(t). S(t) has a lot of significance in survival analysis as it helps us to obtain the probabilities for different values of t. Such an ability to get the survival probabilities for numerous time points is an essential tool in obtaining different summary information from survival data. [6]

1.4 Review of literature

Small area estimation balances the survey information with various forms of auxiliary information such as census enumerates, etc., that have the capability to vast the estimator obtained from survey information. Small area estimation incorporates the unification of survey information with auxiliary information of abundant forms, such as overall census enumerates, to step up the efficiency of survey statistics. It boosts the growth of the estimator and escalates its precision at a smaller geographical level by accumulating various data sets. [8] In the sampled areas of the target population, useful information can be given by the survey data. However, in the scenario of smaller units of a population of interest, such as if the population of interest is a region then the small unit can be a sub-region, similarly, it can be a sub-domain, etc. If survey data is applied to such small level data, the results might not be that accurate or precise due to the small sample size or in some scenarios zero sample. [8] A small-area estimation has been practically used very long ago by (Brackstone (1987)), such literature is available in the eleventh century in England and the seventeenth century in Canada (Rao (2003)). However, these small-area statistics at the beginning were based either on census enumerates or any administrative history for obtaining estimates about a variable of interest. Demographers have traditionally used several indirect approaches to estimate the population in small areas and other important features in the years after a census. Sampling is typically not used in traditional demographic approaches. [8] Population analysts have traditionally used numerous indirect approaches at the beginning to estimate the variable of interest in the population and other characteristics of the population in small areas after the census has been done. Such types of approaches did not take into consideration the sampling technique. [8] Demand for small-area statistics has significantly increased globally in recent decades. This is because of their vast exploitation in making policies and programs, the distribution of resources, and planning regionally. National legislative measures have made small-area statistics more necessary, and this will continue increasing in demand in the coming years. The demand for small-area statistics has also increased from the private sector as a result of the reality that local socio-economic, environmental, and many other factors have a major role in business decision-making, particularly in small firms. [8] The most prominent literature on small-area estimation is the contribution of (Rao (2003)), the recent edition has just appeared (Rao and Molina, 2015). Sandal et al (1992) have some portion but do not describe a model-based approach. Richard Valliant, Alan H. Dorfman, Richard M. Royall (2000, section 11.5). the most advanced literature on small-area estimation is from Pfeffermann (2013) [9]. Examples of small area estimation in health-decision making can be seen in (Blakely et al., 2006), (Congdon, 2006, Tomintz et al., 2008), Schmeider et al. (2009), Li et al. (2009), Johnson et al. (2010), and Gutreuter, S., Igumbor, E., Wabiri, N., Desai, M., Durand, L. (2019), Li Z, Hsiao Y, Godwin J, Martin BD, Wakefield J, Clark SJ, et al. (2019)

However, when the sample size of the population of interest is small or zero, the conventional direct estimators are not good at producing reliable estimates (Cochran, 1977). This problem is, somehow, minimized by the Synthetic estimators (Ghosh and Rao, 1994, Gonzalez et al. 1996, Purcel and Kish, 1973,), which utilize larger area estimates to represent relevant small areas, which have the same characteristics as the larger area. ITocreate state estimates of disability and other health characteristics from the National Health Interview Survey, the National Center for Health Statistics in the United States pioneered the use of synthetic estimating in 1968. (National Health Interview Survey). Most sample sizes were too small to produce accurate direct state estimations Gonzalez et al (1973). Kish and Purcell (1973) suggested a synthetic estimator, which is obtained for small areas having similar characteristics to large populations, using the information of larger populations. The classification of a non-homogeneous population into a homogeneous sub-population is called Post-stratification, The process also works in a similar way to the synthetic estimation technique. In this method, estimates are obtained as the weighted averages of the estimates of a homogeneous sub-population (Holt and Smith, 1979). Synthetic methods use population-level auxiliary data, which contributes to the reduction of variance but also causes the bias to be increased in the estimates at a small area level. A hybrid measure of direct and indirect estimation techniques is the composite estimator, which makes a trade-off between the variance and bias at the small-area level. [10]

In the literature on small-area estimation, apart from these significant techniques, numerous works have been done on accumulating data from various sources. Lohar and Rao, (2006), Elliott and Devis (2005), and Lohr and Raghunathan (2017) are the existing literature on small area estimation from multiple surveys using individual-level data. Manzi et al. (2011) and followers contribute in the same direction using area-level estimates rather than individual-level information.

To obtain relatively non-identical (unmatched) samples on multiple consecutive occasions, successive/or consecutive sampling can be used. The special case is two-occasion sampling where information is gained on two consecutive occasions. This strategy allows the prime utilization of information on samples obtained on previous surveys to boost the efficiency of the estimates produced on the recent survey. In other words, this strategy abolishes some elements from the sample of the previous survey and it includes some new elements to the sample of the recent survey. Different strategies have been to obtain different parameters of the population in recent surveys. Jessen (1943) used all the information obtained from the previous survey to produce two new estimators. The first estimate was the only recent survey-based sample mean and the second estimate was a regression estimate based on the sample units measured on both recent and previous surveys by combining the two estimates. Moreover, the usage of auxiliary information in consecutive sampling is the best strategy for improving the efficiency of the estimate of the parameter obtained in the current survey, e.g., Sen et al (1975), Okafor and Arnab (1987), Okafor (1992). Multiple surveys can be repeated over time and the effective sample size can be increased by merging information from two or more consecutive surveys. As an example, the United States National Health Interview Survey (NIHS) is an annual survey that uses non-overlapping samples throughout the year. The level of accuracy for state estimates is studied by Marker (2001) by combining the 1995 NHIS sample with the previous year's sample or the last two years' sample.

Statistically equivalent properties to the study variables may be found in sampling units that are close to the study variables geographically or that are close to other related measures connected with the variable of interest. The survey may have also gathered additional data that can be used as distance measurements, such as GPS coordinates for the sample locations, demographic, social, cultural, and economic indicators, as well as time and space data from earlier polls of the same type that were conducted recently.[16] We combine the sampled sampling units "close" to the small area with the sampled sampling units from the small area to form a group, the nearest neighborhood, and then consider it a survey domain [16].
Depending on how the distance metric is defined, the resulting domain could either be a real survey domain or a pseudo-domain. The domain is a valid survey domain if the distance measure defines a defined subpopulation that is independent of any sample selection outcomes. For instance, a true survey domain is formed by all sampling units that are situated within a specified distance of a fixed geographic point within a restricted area. Since it depends on the outcomes of random selection, which creates a random sub-population, all sampling units within a set distance of one or more sampled sampling units can form a pseudo-domain [16].
This is acceptable because our goal is not to estimate the domain's population's demographics. Instead, we are assuming the local area's population characteristics. We can estimate the demographic characteristics of the

small area by treating them as a survey domain and using all the established statistical inference techniques of survey domain analysis, such as small area total, its variance, and variance estimation. Let Ω_i^+ be the sample that has been enlarged to include the sample from the small region as well as the borrowed sampling units from the areas closest to it. [16] Using sampling units for regions near in terms of geographic location, time-space, or other associated parameters correlated with the research character, Ren (2021) created the SAE approach. To regard the small region as either a survey domain or a pseudo-domain, depending on how far away it is from the larger area, he advised grouping sampled units nearby with sampled units inside the small area itself. The domains are known as true survey domains if the distance measure represents a fixed sub-population that is independent of any sample selection. [16] While a pseudo-domain, which is dependent on random selection and represents a subset of the population, is created by combining sampled units. On the other hand, successive sampling makes use of samples that were taken on two occasions [16].
Two qualities describe the recently accepted SAE approach: (i) Estimates domain parameters using the nearest neighborhood approach. and (ii) It combines data from two surveys that have successive sampling approach features. Our suggestion for determining the survival analysis for the $i-th$ small area is illustrated in this section. When samples are taken on successive occasions from the same population, there are more options for flexible estimation. It is possible to think of the demographic and health surveys (DHS) carried out every five years, as sampling on successive occasions. Using the concept of sampling on two occasions, we first pool the mismatched portion of the data from the prior survey with the sample from the current survey. [16]

Evidence from demographic and health surveys on the effects of previous birth intervals on neonatal, infant, and under-five years mortality and nutritional status in developing nations, is studied by Rutstein (2005). Some important literature on survival function and estimation of survival function in small areas are McPherson RA, Tamang J, Hodgins S, Pathak LR, Silwal RC, Baqui AH, Winch PJ (2010), the process evaluation of a community-based intervention promoting multiple maternal and neonatal care practices in rural Nepal. The study incorporating the training of female community health volunteers was put into place to get around the difficulty of communicating a variety of complex messages intended to improve maternity and neonatal health inside Nepal's terai area. With the help of a flipchart and a picture pamphlet, these volunteers were prepared to offer advice to expectant mothers and their families. The booklet is made up of sequentially arranged illustrated statements on postcard-sized laminated cards that are connected by rings. Based on the information in the brochure, pregnant women were actively urged to have conversations with their families. Sreeramareddy CT, Harsha Kumar HN, Sathian B (2014), studied a secondary data analysis of four demographic and health surveys conducted between 1996 and 2011 revealing the temporal patterns and disparities of under-five mortality in Nepal. A cross-sectional study in the mid-western region of Nepal found that female community health volunteers frequently use services for children's illnesses, indicating that simply improving the quality of healthcare is insufficient, studied by Miyaguchi M, Yasuoka J, Poudyal AK, Silwal RC, Jimba M (2014). Factors influencing infant mortality in Nepal: a comparison of the 2006 and 2011 Nepal Demographic and Health Surveys (NDHS), is studied by Lamichhane R, Zhao Y, Paudel S, Adewuyi EO (2017).

1.5 Problem Statement

Surveys are carried out to get information on different health indicators aggregated level, such surveys are often costly. They are quite useful for aggregated levels but for sub-populations or small domains, this information is not useful. For instance, to implement policy recommendations for a population as a whole survey information is useful but the problem pops up in the case of implementation of policies in sub-populations for a proper picture of that sub-population. Aiming at the effects of the built environment on health outcomes, increasing imbalances, and the developing development of statistical techniques have enabled researchers to explore these connections in a better way. However, in the absence of sufficient data that are recognizable at smaller areas (e.g., census counts), obtaining reliable estimates on health outcome prevalence at these smaller areas (or disaggregated) and likely more neighborhood levels is not possible by geographic health researchers. When surveys are not enough to obtain reliable direct estimates because of the problem of not sufficient information or small sample size, researchers often prefer to depend on the estimates obtained by small area estimation techniques. The researchers need to realize both the primary methods executed to obtain these estimates and their strengths and limitations as the primary and secondary use of small area estimates and geographical health research increase. Therefore, a more resilient, vigorous, and data-driven predictive approach is required to deal with the issue of producing estimates at area levels.

1.6 Objectives of the Study

- The goal of this study is to address the problem of producing accurate estimates for regions where there are few or no data on a particular variable of interest. For statisticians and policymakers, this paucity of information makes it difficult to produce precise estimates for these sectors. By using data from many sources and events, the study suggests a solution. Estimates can be derived for places where data availability is limited by combining various data relating to that area.

- This study's main goal is to provide improved techniques for calculating survival functions in small geographic areas. The study also aims to create unique estimation methods for infant mortality rates across a range of sub-populations classified by demographic traits.

- The study will use indirect approaches such as the synthetic estimation method, composite estimation method, and Time-Space Nearest Neighbourhood Estimation method to accomplish these objectives.

- The findings of this study will be helpful to those who develop public health strategies since they can be used to create successful plans for areas with scarce data sources. This will facilitate a more comprehensive understanding of infant mortality for various stakeholders, including government bodies and non-governmental organizations (NGOs) focused on addressing infant mortality.

1.7 Outline of the study

Till now traditional direct estimates have been used to obtain estimates for a population. Estimates from traditional methods are not accurate and precise if there is a small sample size or a zero sample. The study suggests improved indirect methods to cope with this problem. The methods proposed in this study comprise two methods, i.e., some improved indirect methods, and the nearest neighborhood method. Further, Chapter 2 provides thorough knowledge of the traditional methods and all the newly suggested methods for obtaining survival function. All the simulation tasks, the relative efficiency of the proposed estimators, mean square and biases under bootstrapping will be discussed in Chapter 3. Finally, Chapter 4 will conclude the study with some future recommendations.

Chapter 2

Material and Methods

In this section, firstly the problem under investigation is stated by depicting variables and articulating the information about the study population.
Let U be regarded as the population of interest of size N, can be defined as $U = \{U_1, U_2, \ldots, U_N\}$. The variable of interest related to the population is y with covariate x possessing values x_k and y_k on the kth unit of the population. A certain type of sampling design D is used to draw a sample of size s from the population which results in the non-sampled set $\bar{s} = U - s$. Let U_i for $(i = 1, 2, ..., m)$ be the set of individuals belonging to the ith area (sub-population) of the population. The basic assumption is that at the population level, no information about the area membership is accessible and the task is to obtain area-specific infant mortality. We first discuss some existing methods of SAE for estimating survival function.

2.1 Small area direct estimator for survival function

The small area direct estimator uses small area-specific auxiliary information available in the sample of the population. [1, 12]
Let π_k be the inclusion probability of kth unit in the sample and the sample weight of the kth unit in the sample is given by $w_k = \frac{1}{\pi_k}$.

The direct estimator for the survival function of the ith small area can be written as : [1, 12]

$$\hat{S}_i^{dir}(t) = \frac{\sum_{k\in\Omega_i} w_k I_k}{\sum_{k\in\Omega_i} w_k} \tag{1}$$

where, $I(y_k > t) = I_k = \begin{cases} 1 & \text{if } y_k \geq t \\ 0 & \text{if } y_k < t \end{cases}$
and Ω_i for (i=1,2,3,...m) is the set of units sampled in the ith area. [11]

Assuming simple random sampling as the sample design, we can write the direct estimator for the survival function in the ith area as

$$\hat{S}_i^{dir}(t) = \frac{\sum_{k\in\Omega_i} I_k}{n_i} \tag{2}$$

It is easy to show that the estimator given in (1) is unbiased with variance given by:

$$V\left[\hat{S}_i(t)\right] = \sum_{k\in\Omega_i} \frac{w_k(w_k-1)}{w_k^2} I_k^2 \tag{3}$$

[12]

Under simple random sampling without replacement the variance of the survival function estimator under the assumption $w_k = \frac{N}{n} \approx \frac{N_i}{n_i}$(we assume this when n_i is not available to us) is

$$V\left[\hat{S}_i(t)\right] = \sum_{k\in\Omega_i} \frac{\frac{N}{n}(\frac{N}{n}-1)}{(\frac{N}{n})^2} I_k^2$$

[12]

This estimator may not be suggested as the reliable estimator as the sample is not sufficient to produce accurate estimators for the domains. There could also be a scenario where a domain might not have enough observation or it might have fewer observations to get an accurate sample out of it.

2.1.1 Direct Estimator with Weight Adjustment for Survival Function

Let g_k be the adjustment factor for design weights, $k \in \Omega$ then the final weight for the sample is given by

$$W_k = w_k g_k$$

Now, the weight-adjusted direct survival estimator for the small area i can be written as

$$\hat{S}_i^{ad}(t) = \frac{\sum_{k\in\Omega_i} W_k I_k}{\sum_{k\in\Omega_i} W_k} \tag{4}$$

[1, 12]

The variance of the weight-adjusted direct estimator can be obtained by replacing w_k by W_k in Equation (3), and given by

$$V\left[\hat{S}_i(t)\right] = \sum_{k\in\Omega_i} \frac{W_k(W_k - 1)}{W_k^2} I_k^2 \tag{5}$$

One way to adjust the weights is post-stratification. [12] Let a subgroup of the population be denoted by U_j (for j=1,2,...J) with size N_{+j}. Further, Ω_{+j} is the set of units in the sample belonging to stratum j. A basic direct estimate of N_{+j} is $\hat{N}_{+j} = \sum_{k\in\Omega_{+j}} w_k$

In this case, the adjustment factor g_k will be $g_k = \frac{N_{+j}}{\hat{N}_{+j}}$

Using $\hat{N}_{+j}$ and g_k, we obtain a weight-adjusted estimator for the ith small area as follow

$$\hat{S}_i^{wa}(t) = \frac{\sum_{k\in\Omega_{ij}} W_k I_k}{\sum_{k\in\Omega_{ij}} W_k} \tag{6}$$

The variance of the weight-adjusted estimator can be expressed as

$$V\left[\hat{S}_i^{wa}(t)\right] = \sum_{k\in\Omega_{ij}} \frac{W_k(W_k - 1)}{W_k^2} I_{ki}^2 \tag{7}$$

2.1.2 Weight-Adjusted Ratio Estimator for Survival Function

Let X be the known total of the auxiliary variable x and $\bar{X}_i$ is the population means for a small area i and $\bar{x}$ is the sample mean for a small area i. The direct Horvitz Thompson-type estimator of X is given by

$$\hat{X}_i = \sum_{k\in\Omega_i} \pi_k X_k$$

The adjustment factor in this case is $g_k = \frac{\bar{X}_i}{\bar{x}_i}$ for $k \in \Omega$. [1, 12]
The direct ratio estimator is obtained as

$$\hat{S}^r(t) = \frac{\hat{S}_i(t)}{\bar{x}_i}\bar{X}_i \quad (8)$$

Bias and MSE of the weight-adjusted Ratio Estimator can be obtained as Let $\hat{S}_i(t) = S_i(t)(1 + e_o(t))$, $\bar{X}_i = \bar{X}(1 + e_1(t))$ and $e_{oi}(t) = \frac{\hat{S}_i(t)-S_i(t)}{S_i(t)}$ such that [12]
$E[e_{oi}(t)] = E[e_{1i}(t)] = 0$ and $E[e_{oi}^2(t)] = ...$ $E[e_{1i}^2(t)] = ...$ $E[e_{oi}(t)e_{1i}(t)] = ...$ Now,

$$V(\hat{S}_i(t)) = \frac{E[\hat{S}_i(t) - S_i(t)]^2}{[S_i(t)]^2} = E[e_o(t)]^2$$

$$\hat{S}_{iadj}^{DIR}(t) = \frac{S_i(t)[1 + e_{oi}(t)]}{\bar{X}[1 + e_1(t)]}\bar{X} = S_i(t)[1 + e_o(t)][1 + e_1(t)]^{-1}$$

$$\hat{S}_i(t) = S_i(t) + S_i(t)[e_o(t) - e_1(t) - e_o(t)e_1(t) + (e_1(t))^2+, ...] \quad (9)$$

Ignoring higher-order terms, we have

$$\hat{S}_i(t) - S_i(t) = S_i(t)[e_o(t) - e_1(t) - e_o(t)e_1^2(t) + (e_1(t))^2] \quad (10)$$

Taking expectation on Equation (10)

$$E[\hat{S}_i(t) - S_i(t)] = S_i(t)[E(e_o(t)) - E(e_1(t)) - E(e_o(t)e_1(t) + E(e_1^2(t))] \quad (11)$$

$$
\begin{aligned}
Bias[\hat{S}^r(t)] &= E[\hat{S}_i(t) - S_i(t)] \\
&= S_i(t)[E(e_{1i}^2(t)) - E(e_{oi}(t)e_1(t))] \\
&= S_i(t)\left[\frac{V(\bar{X}_i)}{\bar{X}_i^2} - \frac{Cov(\bar{X}_i, \hat{S}_i(t))}{\bar{X}_i S_i(t)}\right] \\
&= S_i(t)[C_{\bar{x}_i}^2 - C_{\bar{x}_i} C_{\bar{S}_i(t)}] \qquad (12)
\end{aligned}
$$

Square and taking the expectation of (10), we have

$$
\begin{aligned}
MSE(\hat{S}^R(t)) &= E[\hat{S}_i(t) - S_i(t)]^2 \\
&= S_i^2(t)E[e_{oi}(t) - e_{1i}(t)]^2 \\
&= S_i^2(t)E[e_{oi}^2(t) + e_1^2(t) - 2e_{oi}(t)e_{1i}(t)] \\
&= S_i^2(t)[E(e_{oi}^2(t)) + E(e_{1i}^2(t)) - 2E(e_{oi}(t)e_{1i}(t))] \\
&= S_i^2(t)\left[\frac{V(\hat{S}_i(t))}{S_i^2(t)} + \frac{V(\bar{X})}{\bar{X}^2} - 2(\frac{(Cov(\bar{X}, \hat{S}_i(t))}{\bar{X}\hat{S}_i(t)})\right] \\
&= V(\hat{S}_i(t)) + \frac{S_i^2(t)}{\bar{X}^2}V(\bar{X}) - 2\frac{S_i(t)}{\bar{X}}Cov(\bar{X}, \hat{S}_i(t)) \\
&= V(\hat{S}_i(t)) + R^2V(\bar{X}) - 2RCov(\bar{X}, \hat{S}_i(t)) \qquad (13)
\end{aligned}
$$

2.2 Indirect Estimators of the Survival Function

Due to the small sample size and high-cost constraints, indirect approaches are essential for small area estimates. These methods improve accuracy by utilizing auxiliary data and minimizing non-sampling errors.

2.2.1 Synthetic estimator of Survival Function for Small Area

An unbiased estimator for a relatively larger area is derived from a sample survey and that estimator is used to derive estimates for sub-areas on the grounds. It is assumed that the small areas share the same features as the large area and we classify these estimates as synthetic estimates.[1, 13]
We assume the following implicit model that the survival function of the ith area is equal to the overall survival function, which is

$$S_i(t) = S(t) \tag{14}$$

Under Equation (14), the estimator for the ith area is

$$\hat{S}_i(t) = \hat{S}(t)$$

The survival function estimator using sample weight for the small area i can be written as

$$\hat{S}_i^{syn}(t) = \frac{\sum_{k\in\Omega} w_k I_k}{\sum_{k\in\Omega} w_k} \tag{15}$$

Under simple random sampling, the synthetic estimator for the survival function of the population is

$$\hat{S}_i^{syn}(t) = \frac{1}{n}\sum_{k\in\Omega} I(y_k > t) \tag{16}$$

The bias and MSE of the synthetic estimator is given by

$$\begin{aligned} Bias(\hat{S}_i^{syn}(t)) &= E[\hat{S}_i^{syn}(t) - S_i^{syn}(t)]^2 \\ &= E[\hat{S}_i^{syn}(t)] - S_i^{syn}(t) \\ &= S(t) - S_i(t) \end{aligned} \tag{17}$$

[13]

The bias will be zero when the small area's survival function coincides with the overall population's survival function. However, this assumption is not fulfilled in real applications. We may construct a synthetic estimator using a broad area-level estimator.

For domains of large size, synthetic estimators rely on direct estimators. As a result, given the small domain, the design variance of the synthetic estimator is lower than that of the direct estimator. Synthetic estimators are skewed because of their reliance on strict assumptions. Full mean square error, which considers bias and volatility, is hence important. The approximated mean square is

$$\begin{aligned} MSE\left(\hat{S}_i^{syn}(t)\right) &= E\left(\hat{S}_i^{syn}(t) - S_i(t)\right)^2 \\ &= E\left(\hat{S}_i^{syn}(t) - \hat{S}_i^{dir}(t)\right)^2 + 2Cov(\hat{S}_i^{syn}(t), \hat{S}_i^{dir}(t)) - V(\hat{S}_i^{dir}(t)) \\ &\approx E\left(\hat{S}_i^{syn}(t) - \hat{S}_i^{dir}(t)\right)^2 - V(\hat{S}_i^{dir}(t)) \end{aligned} \tag{18}$$

A sample estimate of the MSE of $\hat{S}_i^{syn}(t)$ is given by

$$mse\left[\hat{S}_i^{syn}(t)\right] = \left[\hat{S}_i^{syn}(t) - \hat{S}_i^{dir}(t)\right]^2 - V[\hat{S}_i^{dir}(t)] \tag{19}$$

[13]

Averaging to overall areas, we have

$$mse_a\left(\hat{S}_i^{\ syn}(t)\right) = \frac{1}{m}\sum_{l=1}^{m}\frac{1}{N_l^2}\left(\hat{S}_l^{syn}(t) - \hat{S}_l^{dir}(t)\right) - \frac{1}{m}\sum_{l=1}^{m}\frac{1}{N_l^2}V\left(\hat{S}_l^{dir}(t)\right) \tag{20}$$

[13]

The estimate of the bias term is

$$b^2\left(\hat{S}_i^{\ syn}(t)\right) \approx \frac{1}{m}\sum_{l=1}^{m} b^2(\hat{S}_l^{\ syn}(t)) \tag{21}$$

[13]

2.2.2 Synthetic estimator for Survival Function with Stratification

The population is divided into the strata J (j=1,2,3,..., J) which cut across the area. Let N_{ij} be the intersection of domain i and post-stratum j. An implicit model can be written as

$$S_{ij}(t) = S_{+j}(t) = \frac{1}{N_{+j}} \sum_{i=1}^{m} \sum_{k \in U_{ij}} I(y_{ijk} \geq t) \tag{22}$$

[1, 13, 14]

where $\sum_{k=1}^{K} N_{ik} = N_i$

The relationship can be generalized at the sample level as

$$\hat{S}_{ij}(t) = \hat{S}_{+j}(t) \tag{23}$$

[1, 13, 14]

Where $\hat{S}_{+j}(t)$ can be obtained as $\hat{S}_{+j}(t) = \frac{1}{N_{+j}} \sum_{i=1}^{m} \sum_{k \in \Omega_{ij}} I(y_{ijk} > t)$

An implicit model for ith area can be written as

$$S_i(t) = \frac{1}{N_{i+}} \sum_{j=1}^{J} N_{ij} S_{+j}(t) \tag{24}$$

We have $N_i = N_{i+} = \sum_{j=1}^{J} N_{ij}$ Using relations given in (24), we can write a synthetic as

$$\hat{S}_i^{ps}(t) = \frac{1}{\hat{N}_{i+}} \sum_{j=1}^{J} \hat{N}_{ij} \hat{S}_{+j}(t) \tag{25}$$

[1, 13, 14]

2.3 Composite estimator for Survival Function

To equate the domain's bias from a synthetic estimator and the instability from a direct estimator, we suggest the following composite estimator for the survival function (Royall. 1973, Schaible et al., 1977), which is a weighted combination of the direct and the synthetic estimator. [1, 13]

$$\hat{S}_i^{com}(t) = \psi_i \hat{S}_i^{dir}(t) + (1 - \psi_i) \hat{S}_i^{syn}(t) \tag{26}$$

[1, 13]

where ψ_i is the weight ensuring $0 \leq \psi_i \leq 1$

If $\psi > 0.5$, the direct estimator will be given more weight, and efficiency will be lost with a reduction in bias, and vice versa. [1, 13] The MSE of the composite estimator is obtained following Rao (2003), and given by

$$MSE[\hat{S}_i^{com}(t)] = \psi_i^2 MSE[\hat{S}_i^{dir}(t)] + (1-\psi_i)^2 MSE[\hat{S}_i^{syn}(t)] \\ + 2\psi_i(1-\psi_i)E[\hat{S}_i^{dir}(t) - S_i(t)][\hat{S}_i^{syn}(t) - S_i(t)] \tag{27}$$

2.3.1 Optimal Composite Estimator for Survival Function

Deriving the value of parameter ψ through the minimization of the mean squared error associated with the composite, denoted as $MSE(\hat{S}_i^{com}(t))$, results in the optimal formulation of the composite estimator. [1, 13]
The approximate optimal weights depend on true $MSE's$ of $\hat{S}_i^{syn}(t)$ and $\hat{S}_i^{dir}(t)$ and given by

$$\psi_i^* \approx \frac{MSE(\hat{S}_i^{syn}(t))}{\left[MSE(\hat{S}_i^{dir}(t)) + MSE(\hat{S}_i^{syn}(t))\right]} \approx \frac{1}{1+R_o} \tag{28}$$

where $R_o = \frac{MSE(\hat{S}_i^{dir}(t))}{MSE(\hat{S}_i^{syn}(t))}$. The optimum MSE of $\hat{S}_i^{syn}(t)$ is obtained by inserting ϕ_i^* in Equation (27). An estimate of ψ_i^* can be obtained as

$$\hat{\psi}_i^* = \frac{mse(\hat{S}_i^{syn}(t))}{\left[\hat{S}_i^{syn}(t) - \hat{S}_i^{dir}(t)\right]^2} \tag{29}$$

Readers can find a thorough explanation of the choice of $\hat{\psi}_i^*$ in Rao (2003, page:57). [1, 13]

2.4 Regression Estimator of Survival Function

A general regression estimator, which is a model-assisted estimator, is provided when data on a covariate that is substantially linked with the study variable is available, is given by

$$\hat{S}_i^{reg}(t) = \frac{\sum_{k\in\Omega_i} W_k^c I_k}{\sum_{k\in\Omega_i} W_k^c} \tag{30}$$

[1, 13, 15]

Where, a set of regression weights is given by W_k^c for $k \in \Omega_i$ i.e. $W_k^c = W_k \left(1 + \frac{A_{\Omega_i}^{-1}\bar{X}_k}{v^2(\bar{X}_k)}\right)^t (\bar{X}_i - \hat{\bar{X}}_i^{dir})$ with $A_{\Omega_i}^{-1} = \frac{\sum_{k\in\Omega_i} W_k^c \bar{X}_k \bar{X}_k^t}{v^2(\bar{X}_k)}$. $\hat{\bar{X}}_i^{dir}$ is the covariate $x's$ direct estimator and $v^2(X_k)$ is the regression model's variance function. This estimator satisfies the calibration constrain $\frac{\sum_{k\in\Omega_i} W_k^c \bar{X}_k}{\sum_{k\in\Omega_i} W_k^c} = \bar{X}_i$. The estimator given in (30), can also be written as

$$\hat{S}_i^{reg}(t) = \sum_{k\in\Omega_i} W_k^c I_k + \hat{\beta}_i(\bar{X}_i - \hat{X}_i^{dir}) \tag{31}$$

[1, 13] When the sample size in the ith area is small, the estimator $\hat{\beta}_i$ may not be stable. Using $\hat{\beta}_s$ with the complete sample information is one option to move further, and the regression estimator is written as

$$\hat{S}_i^{reg}(t) = \hat{S}_i^{dir}(t) + \hat{\beta}_s(\bar{X}_i - X_i^{dir}) \tag{32}$$

[1, 13]

It is simple to demonstrate that the regression estimator is unbiased and has the variance of the following form assuming that β is known.

$$\begin{aligned} V(\hat{S}_i^{reg}(t)) &= V(\hat{S}_i^{dir}(t)) + \beta^2 X_i^2 V(\hat{S}_i^{dir}(t)) - 2\beta X_i^2 cov(\hat{S}_i^{dir}(t), X_i^{dir})) \\ &= V(\hat{S}_i^{dir}(t))[1 - \rho_{xyi}^2] \end{aligned} \tag{33}$$

Equation (33) demonstrates the regression estimator's efficiency advantage over the direct estimator. When we have at least two observations in the sample that correspond to the limited area of interest, the estimators mentioned in this section function.

If the variance function $v^2(X_k) = x$, also known as the proportional model since the variation in the target variable is expected to be proportional to the auxiliary variable, is true for a single variable regression model, the regression estimator transforms into a ratio estimator, can be expressed as

$$\hat{S}_i^r(t) = \frac{\sum_{k\in\Omega_i} W_k I_k}{\sum_{k\in\Omega_i} W_k X_k} = R_i^r \bar{X}_i \tag{34}$$

By using Taylor series expansion, the bias and mean square error of the ratio estimator can be obtained, which are given by

$$Bias(\hat{S}_i^r(t)) = S_i(t)\frac{V(X_i^{dir})}{\bar{X}_i^2} - \frac{cov(\hat{S}_i^{dir}(t), \hat{X}_i^{dir})}{S_i(t)\bar{X}_i} \tag{35}$$

$$MSE(\hat{S}_i^r(t)) \approx v(\hat{S}_i^{dir}(t)) + R_{1i}^2 V(\hat{X}^{dir}) - 2R_{1i} cov(\hat{S}_i^{dir}(t), \hat{X}_i^{dir}) \tag{36}$$

where $R_{1i}^2 = \frac{S(t)}{\bar{X}_i}$.
From Equation (36), As can be shown, the ratio estimator outperforms the simple direct estimator., i.e., $MSE(\hat{S}_i^r(t)) \le V(\hat{S}^{dir}(t))$, when $\rho_{xyi} < \frac{R_{1i}SE(\hat{X}^{dir})}{2SE\hat{S}_i^r(t)}$, where, $\rho_{xyi} = \frac{cov(\hat{S}_i^{dir}(t),\hat{X}_i^{dir})}{\sqrt{V(\hat{S}_i^r(t)V(\hat{X}^{dir})}}$, and $SE\hat{S}_i^r(t)$ and $SE(\hat{X}^{dir})$ are the direct estimators' standard errors for the study character y and the auxiliary character x, respectively.
The estimators provided in Equations (32) and (34) both become the form of the synthetic estimator that assumes the same regression model is valid over all small areas when X_i is substituted by X.

When there is a concern about non-response bias or other potential defects in the survey data, as well as when there are auxiliary variables that are highly associated with the variable of interest, the synthetic regression estimator is especially helpful. When direct estimation may be difficult due to small sample sizes or non-response concerns, this method makes use of the link between the auxiliary variables and the target variable to build a model that estimates the target variable more precisely. When the basic assumptions of the regression model are satisfied and you have faith in the accuracy and applicability of the auxiliary variables, you should use the synthetic regression estimator.

By combining clusters of sample units from a single target survey or a comparable survey done in recent years that is close to a small region, Ren (2021) presented a method for creating a domain made up of a small area. In the following part, we provide several new strategies for nearest-neighbor estimators that incorporate data from two successive surveys. [16]

2.5 Time Space Nearest Neighborhood Method Survival Function

Let $\Omega_i^{(1)}$ and $\Omega_i^{(2)}$ be the samples for the *ith* area, chosen on the first and second occasions respectively. Further, $\Omega_{i_m}^{(1)}$ and $\Omega_{i_m}^{(2)}$ be the sets of matched samples in $\Omega_i^{(1)}$ and $\Omega_i^{(2)}$ respectively such that $\Omega_{i_m}^{(1)} = \Omega_{i_m}^{(2)} = \Omega_{i_m}$ Further, $\Omega_{i_u}^{(1)}$ and $\Omega_{i_u}^{(2)}$ be the sets of unmatched parts in the survey conducted on two occasions such that $\Omega_{i_u}^{(h)} = \Omega_i^{(h)} - \Omega_{i_m}^{(h)}$ for $(h = 1, 2)$.
We suggest different strategies for estimating the survival function estimate of the *ith* area and categorize them into four major groups.

2.5.1 Group-1 (G1) Estimator of Survival Function-Direct pooling of the samples on two occasions.

These Group-1 (G1) estimators were introduced in Ren (2021) [16], which tells, that the data from the previous occasion can be used for pooling the unmatched data with the sample selected for the current occasion. Let Ω_i^+ be the broad sample consisting of units on the *ith* area in the current sample plus the unmatched units from the previous sample such that $\Omega_i^+ = \Omega_i^{(2)} U \Omega_{i_u}^{(1)}$.
For the *ith* small area under successive samples, an improved estimator can

be obtained as

$$\hat{S}_i^{G1}(t) = \frac{\sum_{k\in\Omega_i^+} W_k^* I_k}{\sum_{k\in\Omega_i^+} W_k^*} \tag{39}$$

where $W_k^*,\ k \in \Omega_i^{(2)}$ is a set of expansion weights associated with the kth sampling unit on the second occasion.

$\hat{S}_i^{G1}(t)$ is a ratio-type mean estimator that utilizes the domain inference tool and the linearization method. Assuming simple random sampling in both surveys $\left(W_k^* = \frac{N_i}{n_i^+}\ for\ k \in \Omega_i^{(+)}\right)$, we can show that $\hat{S}_i^{G1}(t)$ is unbiased estimator under model given in Equation (39) i.e.,

$$E[\hat{S}_i^{G1}(t)] = \frac{\sum_{k\in\Omega_i^+} W_k^* E[I_k]}{\sum_{k\in\Omega_i^+} W_k^*} = S_i(t) \tag{40}$$

with variance

$$V[\hat{S}_i^{G1}(t)] = \left(\frac{1}{n_i^+} - \frac{1}{N_i}\right)\sigma_i^2 \tag{41}$$

$V[\hat{S}_i^{G1}(t)] \to 0$ as $n_i^+ \to \infty$.

However, the problem of a small sample size or zero sample size can be overcome by G1. For area-specific estimation, it suggests equal weights to be given to the units in both samples.

2.5.2 Group-2 (G2) Class of Estimator of Survival Function-Weighted average of estimators on two occasions

Another way to obtain an estimator for the survival function of ith area is to find the weighted survival function estimate of the unmatched samples on the first occasion $\Omega_u^{(1)}$ and the full sample from the second occasions $\Omega^{(2)}$. Let $\hat{S}_i^{(1)}(t)$ and $\hat{S}_i^{(2)}(t)$ be the estimators of the survival function for $i-th$ area at first and second occasions respectively. Giving relatively high weight to the units from the current occasion seems logical as we intended to obtain an estimate of the population on the second occasion.

The area-specific estimator is given by

$$\hat{S}_i^{G2}(t) = \theta^{(1)}\hat{S}_i^{(1)}(t) + \theta^{(2)}\hat{S}_i^{(2)}(t) \tag{42}$$

$\hat{S}_i^{G2}(t)$ is a class of estimators of the ith area with weights $\theta^{(1)}$ and $\theta^{(2)}$ for the estimator at the first and second occasions respectively such that $\theta^{(1)}+\theta^{(2)} = 1$. where $\hat{S}_i^{(1)}(t)$ and $\hat{S}_i^{(2)}(t)$ are the estimates obtained from the previous survey and the current survey respectively. If $\theta > 0.5$, the survey conducted the previous year will be given more weight and vice versa. The choice of θ can be obtained by minimizing the $MSE(\hat{S}_i^{G2}(t))$. The inference will be simple, assuming that the polls completed on two occasions are independent. The calibration method can increase the aggregate survival estimate's proximity to the target estimate or previous survey estimates. The bias of $\hat{S}_i^{G2}(t))$ is expressed as

$$\begin{aligned} Bias(\hat{S}_i^{G2}(t)) &= \theta^{(1)}Bias(\hat{S}_i^{(1)}(t)) + \theta^{(2)}Bias(\hat{S}_i^{(2)}(t)) \\ &= \theta^{(1)}E[\hat{S}_i^{(1)}(t) - S_i(t)] + \theta^{(2)}E[\hat{S}_i^{(2)}(t) - S_i(t)] \end{aligned} \tag{43}$$

where $S_i(t)$ is the population total for a certain area during the current time. The estimate $\hat{S}_i^{G2}(t)$ will only be unbiased if the population characteristics stay stable over the time between the two surveys, which is feasible when the two estimators are unbiased estimators of the area-specific total in the population at the current occasion.

Similarly, the MSE of $\hat{S}_i^{G2}(t)$ can be expressed as

$$MSE(\hat{S}_i^{G2}(t)) = \theta^{(1)2}MSE(\hat{S}_i^{(1)}(t)) + \theta^{(2)2}MSE(\hat{S}_i^{(2)}(t)) \tag{44}$$

The covariate term vanishes as we select the unmatched parts from the previous survey, hence the two samples are independent.

The optimum value of $\theta^{(1)}$ is obtained by minimizing $MSE(\hat{S}_i^{G2}(t))$ with subject to the constraint $\theta^{(1)}+\theta^{(2)} = 1$ and given by $\theta^{(1)opt} = \frac{MSE(\hat{S}_i^{(2)}(t))}{MSE(\hat{S}_i^{(1)}(t))+MSE(\hat{S}_i^{(2)}(t))}$

The optimal choice of $\theta^{(1)}$ can also be expressed as

$$\theta_{opt}^{(1)} = \frac{1}{1 + RE_{12}}$$

where $RE_{12} = \frac{MSE(\hat{S}_i^{(1)}(t))}{MSE(\hat{S}_i^{(2)}(t))}$ A higher value of RE_{12} shows higher variation in the estimate obtained from the first occasion and suggests the relatively lower choice of $\theta^{(1)}$ and vice versa. The optimum MSE of $\hat{S}_i^{G2}(t)$ is obtained after substituting the optimum value of $\theta^{(1)}$ and $\theta^{(2)}$ in Equation (**??**) i.e.,

$$MSE\left[\hat{S}_i^{G2}(t)\right]_{opt} = \frac{MSE[\hat{S}_i^{(2)}(t)]MSE[\hat{S}_i^{(1)}(t)]}{MSE[\hat{S}_i^{(1)}(t)] + MSE[\hat{S}_i^{(2)}(t)]} \tag{46}$$

Equation (42) can be thought of as a class of successive sample area-specific estimators because it is possible to generate different members of the group by substituting alternative estimators for $\hat{S}_i^{(1)}(t)$ and $\hat{S}_i^{(2)}(t)$. Following are the group of the G2 estimators for successive samples that will be covered:

Members of the G2 Class of estimator for Survival Function

Obtaining direct estimators from the two samples, a successive sample area-specific estimator is given by

$$\hat{S}_{1_i}^{G2}(t) = \theta_1^{(1)}\hat{S}_i^{dir(1)}(t) + \theta_1^{(2)}\hat{S}_i^{dir(2)}(t) \tag{47}$$

We obtain the following estimator by combining the synthetic estimator from the sample on the current occasion with the direct estimator on the previous occasion.

$$\hat{S}_{2_i}^{G2}(t) = \theta_2^{(1)}\hat{S}_i^{dir(1)}(t) + \theta_2^{(2)}\hat{S}_i^{syn(2)}(t) \tag{48}$$

Similarly, the following estimator is obtained using the composite estimator on the current occasion follow

$$\hat{S}_{3_i}^{G2}(t) = \theta_3^{(1)}\hat{S}_i^{dir(1)}(t) + \theta_3^{(2)}\hat{S}_i^{com(2)}(t) \tag{49}$$

The bias (if exists) and the MSE of the estimator given in (49) are similar to the derivations of (43)-(45) The synthetic and composite estimators' mean squared errors are expressed similarly to those in Section 2's definition. One can employ various direct and indirect estimators from the sample of the current occasion in Group-2. Due to the fact that data from previous surveys is only used when we have a sufficient sample size, resulting in a valid direct area level estimate, we confine the estimator on the first occasion to the direct approach only. On the current occasion, Group-2 focuses on using synthetic and composite methods. However, a comparable estimator can be built to estimate the mean or total of the auxiliary variable, which can then be used to estimate area-specific study variable parameters.

2.5.3 Group-3 (G3) Class of Estimators of Survival Function-Synthetic estimators using G2 class of estimates

This group of estimators focuses on getting area-level estimates using the information on some associated auxiliary variable X with sample observations in the previous and current surveys under G2, and known area-specific mean from current population $\bar{X}_i$. Assuming a linear relationship between the survey variable y on the auxiliary covariate(S), we propose a class of linear regression-type estimator based on successive samples, presuming that the regression coefficient β_i of y on x is known for the ith area on the current occasion.

$$\hat{S}_i^{G3}(t) = \hat{S}_i^{G2}(t) + \beta_i^{(2)}(X_i - \hat{X}_i^{G2}) \tag{50}$$

When β_i is not known at area level an estimate $\hat{\beta}_i^{(2)}$ is obtained from the second occasions sample. We also propose the following ratio estimator, presuming that the variability in the outcome variable is proportionate to the auxiliary variable on the current occasion.

$$\hat{S}_i^{G3}(t) = \hat{S}_i^{G2}(t)\frac{\bar{X}_i}{\hat{\bar{X}}_i^{G2}} \tag{51}$$

The estimator $\hat{\hat{X}}_i^{G2}$ is obtained after replacing y by x in Equation (42). The bias of $\hat{S}_i^{G3}(t)$ is obtained by introducing the following error terms. Let

$$\hat{S}_i^{G2}(t) = S_i^{(2)}(t)(1 + e_{0i}^*) \quad ; \quad \hat{X}_i^{G2} = X_i^{(2)}(1 + e_{0i}^*) \tag{52}$$

such that $e_{0i}^* = e_{1i}^* \approx 0$, exact equality holds when the estimators $\hat{S}_i^{G2}(t)$ and $\hat{X}_i^{G2}$ are unbiased estimators of $S_i(t)$ and $X_i^{(2)}$ respectively.
Further $E(e_{0i}^{*2}) = \frac{Var(\hat{S}_i^{G2}(t))}{S_i^2(t)}$, $E(e_{1i}^{*2}) = \frac{Var(\hat{X}_i^{G2})}{X_i^2}$; $E(e_{i0}^{*2}e_{i1}^{*2}) = \frac{Cov(\hat{S}_i^{G2}(t),\hat{X}_i^{G2})}{S_i(t)X_i}$ It is simple to demonstrate that the regression-type estimator for area total under Group-3 is unbiased with variance provided by when the coefficient for $\beta_i^{(2)}$ is known.

$$\begin{aligned} V(\hat{S}_{1i}^{G3}(t)) &= V(\hat{S}_i^{G2}(t)) + \beta^2\bar{X}_i^2 V(\hat{S}_i^{G2}) - 2\beta\bar{X}^2 Cov(\hat{S}_i^{G2}(t), \hat{\hat{X}}_i^{G2})) \\ &= V(\hat{S}_i^{G2}(t))[1 - \rho_{xyi}^{*2}] \end{aligned} \tag{53}$$

where, $\rho_{xyi}^{*2} = \frac{Cov(\hat{S}_i^{G2}(t),\hat{X}_i^{G2})}{\sqrt{V(\hat{S}_i^{G2}(t))V(\hat{X}_i^{G2})}}$.
Equation (53) shows the superiority of the regression estimator under G3 over the direct estimator under G2 regarding efficiency. The approximate bias and mean squared error of the ratio estimator under Group-3 are given by

$$Bias(\hat{S}_{2i}^{G3}(t)) \approx S_i(t)\left[\frac{V(\hat{\hat{X}}_i^{G2})}{\bar{X}_i^2} - \frac{Cov(\hat{S}_i^{G2}(t), \hat{X}_i^{G2})}{S_i(t)\bar{X}_i}\right] \tag{54}$$

$$MSE(\hat{S}_{2i}^{g3}(t)) \approx V(\hat{S}_i^{G2}(t)) + R_i^2 V(\hat{\hat{X}}_i^{G2}) - 2R_i Cov(\hat{S}_i^{G2}(t), \hat{\hat{X}}_i^{G2}) \tag{55}$$

where, $R_i = \frac{S_i(t)}{X_i}$. From Equation (55), we can see that the ratio estimator under Group-3 performs better than the direct estimator under Group-2 i.e. $MSE(\hat{S}_{2i}^{G3}(t)) \leq V(\hat{S}_i^{G2}(t))$ when $\rho_{xyi}^* < \frac{R_i SE(\hat{X}_i^{G2})}{2SE(\hat{S}_i^{G2}(t))}$ where $\hat{S}_i^{G2}(t)$ and $SE(\hat{X}_i^{G2})$ are the standard errors of the direct estimators obtained under Group-3 corresponding to the study character y and the auxiliary character x respectively. The members of both ratio-type and regression-type classes of estimators are obtained by substituting the estimators $\hat{S}_i^{G2}(t)$,$SE(\hat{X}_i^{G2})$ or both by their respective members given under Group-2.

2.5.4 Group-4 (G4) Estimator of Survival Function-Synthetic estimators using G1 estimates

To make the class of estimates closer to the target survey, we can utilize available information on some correlated auxiliary variable x with sample observations in the previous and current surveys and known area-specific total from current population X_i. Assuming the same linear relationship between the study variable and one or more covariates, we suggest a class of linear regression-type estimators based on successive samples assuming that the regression coefficient β_i^2 of y on x is known for the ith area. The class of estimators under G3 uses the auxiliary information using the estimators given in G2. However, we can utilize the sample auxiliary information from the combined sample with population information on a current occasion. The estimator for a population total of ith area is known and the same for the two occasions, we have

$$\hat{S}_{1i}^{G4}(t) = \hat{S}_i^{G4}(t) + \hat{\beta}_i^{+}(\bar{X}_i - \hat{\hat{X}}_i^{G4}) \tag{56}$$

When β_i is not known at the area level, an estimate $\hat{\beta}_i^{+}$ is obtained from the combined sample. Further assuming that the variation in the dependent variable is proportional to the auxiliary variable on the previous and current occasion, we introduce the following ratio estimator

$$\hat{S}_{2i}^{G4}(t) = \hat{S}_i^{G1}(t)\frac{X_i}{\hat{\hat{X}}_i^{G1}} \tag{57}$$

The estimator $\hat{\hat{X}}_i^{G1}$ given in Equation (57) is obtained after replacing y by x in Equation (39). The bias and mean squared error of $\hat{S}_{1i}^{G4}(t)$ and $\hat{S}_{2i}^{G4}(t)$ are obtained by introducing the following error terms. Let

$$\hat{S}_i^{G1}(t) = S_i(t)(1 + e_{0i}) \quad , \quad \hat{X}_i^{G1} = X_i(1 + e_{0i})$$

such that $e_{0i}^{*} = e_{1i}^{*} \approx 0$, exact equality holds when the estimators $\hat{S}_i^{G1}(t)$ and $\hat{X}_i^{G1}$ are unbiased estimators of Y_i and X_i respectively.
Further $E(e_{0i}^2) = \frac{Var(\hat{S}_i^{G1}(t))}{S_i^2(t)}$, $E(e_{1i}^2) = \frac{Var(\hat{X}_i^{G1})}{X_i^2}$, $E(e_{0i}^2 e_{i1}^2) = \frac{Cov(\hat{S}_i^{G1}(t),\hat{X}_i^{G1})}{S_i(t)X_i}$

When the coefficient for $\hat{\beta}_i^{+}$ is known, it is easy to show that the regression-type estimator for area total under Strategy II is unbiased with variance given by

$$\begin{aligned} V(\hat{S}_{1i}^{G4}(t)) &= V(\hat{S}_i^{G1}(t)) + \beta^2 \bar{X}_i^2 V(\hat{S}_i^{G1}) - 2\beta \bar{X}^2 Cov(\hat{S}_i^{G1}(t), \hat{\bar{X}}_i^{G1})) \\ &= V(\hat{S}_i^{G1}(t))[1 - \rho_{xyi}^{*2}] \end{aligned} \tag{58}$$

where, $\rho_{xyi}^{*2} = \frac{Cov(\hat{S}_i^{G1}(t),\hat{\bar{X}}_i^{G1})}{\sqrt{V(\hat{S}_i^{G1}(t))V(\hat{\bar{X}}_i^{G1})}}$.
Equation (40) shows the superiority of the regression estimator under G3 over the direct estimator under G2 regarding efficiency. The approximate bias and mean squared error of the ratio estimator under G3 are given by

$$Bias(\hat{S}_{2i}^{G1}(t)) \approx S_i(t)\left[\frac{V(\hat{\bar{X}}_i^{G1})}{\bar{X}_i^2} - \frac{Cov(\hat{S}_i^{G1}(t), \hat{\bar{X}}_i^{G1})}{S_i(t)\bar{X}_i}\right] \tag{59}$$

$$MSE(\hat{S}_{2i}^{G1}(t)) \approx V(\hat{S}_i^{G1}(t)) + R_i^2 V(\hat{\bar{X}}_i^{G1}) - 2R_i Cov(\hat{S}_i^{G1}(t), \hat{\bar{X}}_i^{G1}) \tag{60}$$

where, $R_i = \frac{S_i(t)}{\bar{X}_i}$. From Equation (**??**), we can see that the ratio estimator under Group-3 performs better than the direct estimator under Group-1 i.e. $MSE(\hat{S}_{2i}^{G4}(t)) \leq V(\hat{S}_i^{G1}(t))$ when $\rho_{xyi}^{*} < \frac{R_i SE(\hat{\bar{X}}_i^{G1})}{2SE(\hat{S}_i^{G1}(t))}$ where $SE(\hat{S}_i^{G1}(t))$ and $SE(\hat{\bar{X}}_i^{G1})$ are the standard errors of the direct estimators obtained under G1 corresponding to the study character y and the auxiliary variable respectively. The synthetic versions of the estimators given in Equation (56) and (57) are obtained by replacing area-specific sample quantities of auxiliary characters with the quantities obtained from the full sample.

Chapter 3

Results and Discussions

3.1 Efficiency Comparison of Estimators

To evaluate the performance of the proposed small area estimation strategies, we use reproductive health data from two consecutive surveys i.e. PDHS 2012-13 and PDHS-special 2017-18. A bootstrapped study is conducted by considering the two surveys as the study populations on two successive occasions with 50,238 and 50,495 children of different ages in PDHS 2012-13 and PDHS-special 2017-18 respectively. A description of the variables used in our analysis is given in Table 3.1.

Table 3.1: Some important variables used in the study

DHS code	Variable name	Description	Usage
B7	Age at death (Data-1 and Data-2)	Age of individuals at time of death	Response
B8	Current age (Data-1 and Data-2)	Age of individuals at the time of data collection	Response
V101	Region (Data-1 and Data-2)	Region where data collected	Area membership
V102	Residence type (Data-1 and Data-2)	Type of residence (e.g., urban, rural) at data collection	Stratification
V005	Weight of child (Data-1 and Data-2)	Weight of the child at the time of data collection	Auxiliary variable

An indicator function is also used in this study which is

$$I(y_k > t) = I_k = \begin{cases} 1 & \text{if } y_k \geq t \\ 0 & \text{if } y_k < t \end{cases}$$

The indicator function is obtained from the "Age at death" and "Current age" variables. Table 3.2 illustrates the indicator function used in this study. Those whose "age at death" is less than 12 months and whose "current age" is not given are considered as not surviving. Those whose "age at death" is greater than 12 months and whose "current age" is not given are considered as survived. Those whose "age at death" is not given and whose "current age" is less than 12 months are excluded from the study. Those whose "age at death" is not given and whose "current age" is greater than 12 months are considered as survived.

Age at Death (In months)	≤ 12	$12 \leq$	NA	NA
Current Age (In months)	NA	NA	≤ 12	≥ 12
Survival Indicator (In months)	0	1	censord	1

Table 3.2: Survival Indicator Function

The bootstrapping algorithm is conducted with the following steps.

Step-1:

Select a random sample of size n_1 and n_2 without replacement from both data sets (PDHS-2012-13 and PDHS-special 2017-18) and obtain the mean estimators for the two samples separately corresponding to each region.

Step-2:

Obtain the estimates under $G2$ and $G3$ using the estimators in Step-1 with appropriate choices of $\theta_l^{(t)}$ (for t = 1, 2 and 1 = 1,2,3).

Step-3:

Obtain the estimates under $G1$ and $G4$ after combining the two samples obtained in Step-1.

Step-4:

Iterate Steps 1-3, Q times, and obtain the MSE, percentage contribution of bias (PCoB) in MSE, and relative efficiency (RE) of the survival function using the empirical density function approach. The expressions used for computing MSE, PCoB, and RE are given by

$$MSE(\hat{S}_i^{SK(\cdot)}(t)) = \frac{\sum_{q=1}^{Q}\left[\hat{S}_i^{SK(\cdot)}(t) - S_i(t))\right]^2}{Q} \tag{3.1.1}$$

$$PCoB(\hat{S}_i^{SK(\cdot)}(t)) = \frac{\sum_{q=1}^{Q}\left[\hat{S}_i^{SK(\cdot)}(t) - S_i(t))\right]^2}{QMSE(\hat{S}_i^{SK(\cdot)}(t))} \times 100 \tag{3.1.2}$$

$$RE(\hat{S}_i^{SK(\cdot)}(t)) = \frac{MSE\left[\hat{S}_i^{(S1)}(t)\right]}{MSE\left[\hat{S}_i^{SK(\cdot)}(t)\right]} \tag{3.1.3}$$

where K = 1, 2, 3, 4 and ·=reg[1], r[2], creg[3], cr[4].

[1]Reggression

[2]Ratio

[3]Composite Reggression

[4]Composite Ratio

3.2 Comparison of strategies using bootstap

For numerical comparison, we use Region as the membership variable and age at death in months as the response variable. The results including MSE, PCOB and RE are given in Tables 1-4 for the four provinces (Punjab, Sindh, KPK and Balochistan). The value of weight $\theta_l^{(1)} = 1 - \theta_l^{(2)}$ is varied from 0.1 to 0.4. The relative efficiency and PCoB both supported the cases with smaller values of weight, which supports the fact that higher importance should be given to the estimators obtained on the current occasion. Further, the relative efficiency of the estimators is higher for larger sample sizes in the majority of the cases for G2 and G3. Among all competing small area estimators, the regression, ratio, composite regression, and composite ratio estimators under G2 and G3 provide higher RE and smaller PCOB than their counterparts even for larger choices of $\theta_l^{(t)}$ for (t = 1,2, l= 1,2,3). Further, the PCOB in MSE tends to increase with the increase in $\theta_l^{(1)}$ which also suggests using smaller weights to the estimators obtained on the first occasion. We observe that the mean estimators under G2 and G3 perform well in terms of both PCOB and RE for all sample sizes when $\theta_l^{(1)}$=1-$\theta_l^{(2)} < 0.5$ i.e. giving more weight to the estimators obtained on the current occasion.

3.2.1 Punjab Results

From table 3.3 it can be seen that the PCoB of the direct estimator in group-1 is small, so this estimator can be used as the best estimator from group-1. Similarly, the direct estimator in group-2 has minimum PCoB, so this estimator is the best estimator from group-2. Likewise, in group-3, the composite estimator has the minimum contribution of MSE, so in this group the best estimator is the composite estimator of group-3. Similarly, ratio estimator is the best estimator from group-4. These four estimators are the best estimators from the four groups for punjab region of Punjab.

Table 3.3: MSE, PCoB, RE in MSE of Total Estimators for Punjab under different groups

	$n^{(1)} = n^{(2)} = 500$				$n^{(1)} = n^{(2)} = 1000$			
$\theta^{(1)}$	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
$MSE(\hat{S}_i^{G1}(t))$	0.0009	-	-	-	0.0006	-	-	-
PCoB in MSE	43.6395	-	-	-	61.2758	-	-	-
$MSE(\hat{S}_i^{G1(w)}(t))$	0.0013	-	-	-	0.0009	-	-	-
PCoB in MSE	47.3883	-	-	-	64.7302	-	-	-
$MSE(\hat{S}_i^{G1(ps)}(t))$	0.0009	-	-	-	0.0006	-	-	-
PCoB in MSE	43.6395	-	-	-	61.2758	-	-	-
$MSE(\hat{S}_i^{G2}(t))$	0.0009	0.0008	0.0007	0.0007	0.0004	0.0004	0.0004	0.0005
PCoB in MSE	1.3451	6.7452	15.6881	27.6000	2.9152	12.8925	28.0004	43.4962
$MSE(\hat{S}_i^{G2(ps)}(t))$	0.0009	0.0008	0.0007	0.0007	0.0004	0.0004	0.0004	0.0005
PCoB in MSE	1.3451	6.7452	15.6881	27.6000	2.9152	12.8925	28.0004	43.4962
$MSE(\hat{S}_i^{G2(w)}(t))$	0.0011	0.0009	0.0007	0.0007	0.0005	0.0004	0.0003	0.0003
PCoB in MSE	2.6786	9.5696	19.3260	31.4152	5.6814	17.7791	33.2997	48.0217
$MSE(\hat{S}_i^{G2(reg)}(t))$	0.0003	0.0003	0.0003	0.0003	0.0001	0.0001	0.0001	0.0001
PCoB in MSE	47.0526	46.8834	46.5167	46.1357	64.3445	64.2322	57.6868	57.6300
$RE(\hat{S}_i^{G2(reg)}(t))$	1.3665	1.3570	1.3901	1.3872	1.3345	1.3347	1.6141	1.5994
$MSE(\hat{S}_i^{G2(r)}(t))$	0.0007	0.0007	0.0006	0.0006	0.0004	0.0004	0.0004	0.0004
PCoB in MSE	42.3203	42.0969	41.3049	40.8321	60.0210	59.8856	52.7867	52.7563
$RE(\hat{S}_i^{G2(r)}(t))$	1.3952	1.3856	1.4275	1.4240	1.4069	1.4074	1.6931	1.6768
$MSE(\hat{S}_i^{G2(creg)}(t))$	0.0006	0.0006	0.0006	0.0006	0.0003	0.0003	0.0003	0.0003
PCoB in MSE	26.4150	26.0606	26.0968	25.4897	41.9877	41.7625	34.9272	35.0440
$RE(\hat{S}_i^{G2(creg)}(t))$	1.5527	1.5432	1.5747	1.5681	1.7755	1.7762	1.9985	1.9807
$MSE(\hat{S}_i^{G2(cr)}(t))$	0.0006	0.0006	0.0006	0.0006	0.0003	0.0003	0.0003	0.0003
PCoB in MSE	22.7656	22.3985	22.0479	21.4363	37.4194	37.1831	30.3725	30.5158
$RE(\hat{S}_i^{G2(cr)}(t))$	1.7329	1.7378	1.7533	1.7635	1.4089	1.4144	1.5778	1.5624
$MSE(\hat{S}_i^{G3(reg)}(t))$	0.0006	0.0005	0.0004	0.0003	0.0004	0.0003	0.0002	0.0002
PCoB in MSE	48.1041	48.607	46.3964	43.9074	65.3242	65.8012	58.4586	56.4613
$RE(\hat{S}_i^{G3(reg)}(t))$	1.6355	1.9507	2.4476	2.9169	1.5857	1.8938	2.7767	3.2967
$MSE(\hat{S}_i^{G3(r)}(t))$	0.0005	0.0004	0.0003	0.0003	0.0004	0.0003	0.0002	0.0001
PCoB in MSE	42.4903	41.8811	47.4366	33.1462	60.1858	59.6233	49.4968	45.3408
$RE(\hat{S}_i^{G3(r)}(t))$	1.6899	2.0448	2.6267	3.1724	1.7024	2.0781	3.1150	3.7875
$MSE(\hat{S}_i^{G4(reg)}(t))$	0.0002	-	-	-	0.0001	-	-	-
PCoB in MSE	37.2425	-	-	-	52.4747	-	-	-
$RE(\hat{S}_i^{G4(reg)}(t))$	3.5225	-	-	-	3.8613	-	-	-
$MSE(\hat{S}_i^{G4(r)}(t))$	0.0002	-	-	-	0.0001	-	-	-
PCoB in MSE	24.4560	-	-	-	37.9505	-	-	-
$RE(\hat{S}_i^{G4(r)}(t))$	3.8159	-	-	-	4.5502	-	-	-

3.2.2 Sindh Results

Similarly, Table 3.4 illustrates the efficiency comparison of different estimators in the region Sindh of Pakistan. it can be seen that the PCoB of the direct estimator in group-1 is small, so this estimator can be used as the best estimator from group-1. Similarly, the ratio estimator in group-2 has minimum PCoB, so this estimator is the best estimator from group-2. Likewise, in group-3, the composite ratio estimator has the minimum contribution of MSE, so in this group the best estimator is the composite estimator of group-3. Similarly, ratio estimator is the best estimator from group-4. These four estimators are the best estimators from the four groups for Sindh region of Pakistan.

Table 3.4: MSE, PCoB, RE in MSE of Total Estimators for Sindh under different groups

$\theta^{(1)}$	$n^{(1)}=n^{(2)}=500$				$n^{(1)}=n^{(2)}=1000$			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
$MSE(\hat{S}_i^{G1}(t))$	0.0011	-	-	-	0.0008	-	-	-
PCoB in MSE	48.5115	-	-	-	65.5577	-	-	-
$MSE(\hat{S}_i^{G1(w)}(t))$	0.0016	-	-	-	0.0012	–	–	–
PCoB in MSE	49.8887	–	–	–	66.9233	–	–	–
$MSE(\hat{S}_i^{G1(ps)}(t))$	0.0011	–	–	–	0.0008	–	–	–
PCoB in MSE	48.4003	–	–	–	65.7037	–	–	–
$MSE(\hat{S}_i^{G2}(t))$	0.0009	0.0008	0.0008	0.0009	0.0004	0.0004	0.0005	0.0006
PCoB in MSE	2.1031	9.3027	21.0216	34.3832	4.4023	16.8818	34.9945	51.4904
$MSE(\hat{S}_i^{G2(ps)}(t))$	0.0009	0.0008	0.0008	0.0009	0.0004	0.0004	0.0005	0.0006
$MSE(\hat{S}_i^{G2(w)}(t))$	0.0013	0.0010	0.0009	0.0008	0.0006	0.0005	0.0004	0.0004
PCoB in MSE	5.6173	13.6348	24.8962	36.8263	11.1278	24.2323	40.1371	54.1587
$MSE(\hat{S}_i^{G2(reg)}(t))$	0.0004	0.0004	0.0004	0.0004	0.0002	0.0002	0.0002	0.0002
PCoB in MSE	0.3322	0.3059	0.0162	0.0277	0.5886	0.6895	2.0031	2.0492
$RE(\hat{S}_i^{G2(reg)}(t))$	2.7653	2.7695	2.8368	2.8322	4.1819	4.1181	3.9845	4.0262
$MSE(\hat{S}_i^{S2(r)}(t))$	0.0002	0.0004	0.0004	0.0004	0.0004	0.0002	0.0002	0.0002
CoB in MSE	0.2725	0.2543	0.0083	0.0176	0.4977	0.5998	1.7630	1.7922
$RE(\hat{S}_i^{G2(r)}(t))$	2.7243	2.7315	2.7876	2.7877	4.1345	4.0737	3.9481	3.9900
$MSE(\hat{S}_i^{G2(creg)}(t))$	0.0004	0.0004	0.0004	0.0004	0.0002	0.0002	0.0002	0.0002
CoB in MSE	0.1407	0.1372	0.0073	0.0147	0.2248	0.3189	0.8893	0.8648
$RE(\hat{S}_i^{G2(creg)}(t))$	2.3978	2.4100	2.4304	2.4427	3.6217	3.5709	3.5311	3.5676
$MSE(\hat{S}_i^{G2(cr)}(t))$	0.0005	0.0005	0.0005	0.0005	0.0002	0.0002	0.0002	0.0002
CoB in MSE	0.1128	0.1119	0.0037	0.0096	0.1841	2725	0.7638	0.7358
$RE(\hat{S}_i^{G2(cr)}(t))$	2.1975	2.2073	2.2212	2.2174	2.2084	2.2111	2.1773	2.1755
$MSE(\hat{S}_i^{G3(reg)}(t))$	0.0003	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
CoB in MSE	0.0666	1.6694	4.8228	9.5060	0.1818	3.1614	5.4248	14.1596
$RE(\hat{S}_i^{G3(reg)}(t))$	3.3808	4.0345	4.6762	4.9825	5.1199	5.9219	6.7137	6.9378
$MSE(\hat{S}_i^{G3(r)}(t))$	0.0003	0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
CoB in MSE	0.1461	2.2242	6.4345	12.3317	0.3487	4.1762	7.7533	18.2957
$RE(\hat{S}_i^{G3(r)}(t))$	3.3067	3.9075	4.4184	4.6056	5.0175	5.7180	6.3348	6.0950
$MSE(\hat{S}_i^{G4(reg)}(t))$	0.0002	-	-	-	0.0001	-	-	-
CoB in MSE	19.9539	-	-	-	33.0173	-	-	-
$RE(\hat{S}_i^{G4(reg)}(t))$	4.6028	-	-	-	5.7039	-	-	-
$MSE(\hat{S}_i^{G4(r)}(t))$	0.0003	-	-	-	0.0001	-	-	-
CoB in MSE	23.7884	-	-	-	37.8645	-	-	-
$RE(\hat{S}_i^{G4(r)}(t))$	4.1221	-	-	-	4.9932	-	-	-

3.2.3 KPK Results

Similarly, Table 3.5 illustrates the efficiency comparison of different estimators in the region KPK of Pakistan. it can be seen that the PCoB of the direct estimator in group-1 is small, so this estimator can be used as the best estimator from group-1. Similarly, the ratio estimator in group-2 has minimum PCoB, so this estimator is the best estimator from group-2. Likewise, in group-3, the regression estimator has the minimum contribution of MSE, so in this group, the best estimator is the composite estimator of group-3. Similarly, the ratio estimator is the best estimator from group-4. These four estimators are the best estimators from the four groups for the Punjab region of Punjab.

Table 3.5: MSE, PCoB, RE in MSE of Total Estimators for KPK under different groups

	$n^{(1)} = n^{(2)} = 500$				$n^{(1)} = n^{(2)} = 1000$			
$\theta^{(1)}$	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
$MSE(\hat{S}_i^{G1}(t))$	0.0006	-	-	-	0.0003	-	-	-
PCoB in MSE	11.1012	-	-	-	20.2676	-	-	-
$MSE(\hat{S}_i^{G1(w)}(t))$	0.0011	-	-	-	0.0006	-	-	-
PCoB in MSE	19.5740	-	-	-	33.1643	-	-	-
$MSE(\hat{S}_i^{G1(ps)}(t))$	0.0006	-	-	-	0.0003	-	-	-
PCoB in MSE	11.1012	-	-	-	20.2676	-	-	-
$MSE(\hat{S}_i^{G2}(t))$	0.0009	0.0007	0.0006	0.0006	0.0004	0.0003	0.0003	0.0003
PCoB in MSE	0.2615	1.1516	3.1178	5.9964	0.4634	2.3109	6.0539	11.6688
$MSE(\hat{S}_i^{G2(ps)}(t))$	0.0009	0.0007	0.0006	0.0006	0.0004	0.0003	0.0003	0.0003
$MSE(\hat{S}_i^{G2(w)}(t))$	0.0018	0.0015	0.0012	0.0010	0.0008	0.0007	0.0006	0.0005
PCoB in MSE	8.1375	10.3629	13.2944	16.3755	15.0877	19.3435	23.9961	28.5767
$MSE(\hat{S}_i^{G2(reg)}(t))$	0.0004	0.0004	0.0004	0.0004	0.0002	0.0002	0.0002	0.0002
PCoB in MSE	7.2295	7.0667	3.7623	3.5607	13.4820	13.5491	18.0983	18.2351
$RE(\hat{S}_i^{G2(reg)}(t))$	1.3111	1.2945	1.3084	1.2997	1.3586	1.3566	1.2885	1.2792
$MSE(\hat{S}_i^{G2(r)}(t))$	0.0004	0.0004	0.0004	0.0004	0.0002	0.0002	0.0002	0.0002
CoB in MSE	6.7059	6.5406	3.4309	3.2379	12.5513	12.6164	16.8138	16.94
$RE(\hat{S}_i^{G2(r)}(t))$	1.3237	1.3085	1.3147	1.3072	1.3852	1.3831	1.3206	1.3118
$MSE(\hat{S}_i^{G2(creg)}(t))$	0.0005	0.0005	0.0005	0.0005	0.0002	0.0002	0.0002	0.0002
CoB in MSE	3.2882	3.1697	1.6921	1.5712	6.2759	6.2886	8.6183	8.7563
$RE(\hat{S}_i^{G2(creg)}(t))$	1.2110	1.2017	1.1994	1.1946	1.3075	1.3025	1.2670	1.2637
$MSE(\hat{S}_i^{G2(cr)}(t))$	0.0005	0.0005	0.0005	0.0005	0.0002	0.0002	0.0002	0.0002
CoB in MSE	2.9550	2.8388	1.4938	1.3808	5.6412	5.6527	7.7040	7.8336
$RE(\hat{S}_i^{G2(cr)}(t))$	2.1654	2.1554	2.1224	2.1214	2.0964	2.1078	2.0452	2.0236
$MSE(\hat{S}_i^{G3(reg)}(t))$	0.0002	0.0003	0.0002	0.0002	0.0004	0.0001	0.0001	0.0001
CoB in MSE	18.2689	13.3476	10.1286	12.6061	10.0476	23.2519	35.6876	41.4849
$RE(\hat{S}_i^{G3(reg)}(t))$	1.5689	1.7949	2.1552	2.3836	1.5551	1.7663	1.7647	1.8235
$MSE(\hat{S}_i^{G3(r)}(t))$	0.0002	0.0003	0.0002	0.0002	0.0004	0.0001	0.0001	0.0001
CoB in MSE	9.5010	12.8186	9.9915	12.6666	17.3526	23.1293	34.7380	40.7597
$RE(\hat{S}_i^{G3(r)}(t))$	1.5684	1.8096	2.1486	2.3661	1.5972	1.7939	1.7956	1.8451
$MSE(\hat{S}_i^{G4(reg)}(t))$	0.0002	-	-	-	0.0001	-	-	-
CoB in MSE	21.1087	-	-	-	43.9907	-		-
$RE(\hat{S}_i^{G4(reg)}(t))$	2.3793	-	-	-	1.8792	-	-	-
$MSE(\hat{S}_i^{G4(r)}(t))$	0.0002	-	-	-	0.0001	-	-	-
CoB in MSE	21.2419	-	-	-	43.7283	-	-	-
$RE(\hat{S}_i^{G4(r)}(t))$	2.3492	-	-	-	1.8768	-	-	-

3.2.4 Balochistan Results

Similarly, Table 3.6 illustrates the efficiency comparison of different estimators in the region Balochistan of Pakistan. it can be seen that the PCoB of the direct estimator in group-1 is small, so this estimator can be used as the best estimator from group-1. Similarly, the ratio estimator in group-2 has minimum PCoB, so this estimator is the best estimator from group-2. Likewise, in group-3, the regression estimator has the minimum contribution of MSE, so in this group the best estimator is the composite estimator of group-3. Similarly, ratio estimator is the best estimator from group-4. These four estimators are the best estimators from the four groups for punjab region of Punjab.

Table 3.6: MSE, PCoB, RE in MSE of Total Estimators for Balochistan under different groups

$\theta^{(1)}$	$n^{(1)} = n^{(2)} = 500$				$n^{(1)} = n^{(2)} = 1000$			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
$MSE(\hat{S}_i^{G1}(t))$	0.0018	-	-	-	0.0014	-	-	-
PCoB in MSE	59.2783	-	-	-	74.8251	-	-	-
$MSE(\hat{S}_i^{G1(w)}(t))$	0.0022	-	-	-	0.0016	-	-	-
PCoB in MSE	55.9830	-	-	-	59.4856	-	-	-
$MSE(\hat{S}_i^{G1(ps)}(t))$	0.0018	-	-	-	0.0014	-	-	-
PCoB in MSE	59.2783	-	-	-	74.8251	-	-	-
$MSE(\hat{S}_i^{G2}(t))$	0.0011	0.0010	0.0011	0.0012	0.0005	0.0006	0.0007	0.0009
PCoB in MSE	2.7310	12.5855	26.908	42.2570	5.4696	22.7007	43.4763	59.6809
$MSE(\hat{S}_i^{G2(ps)}(t))$	0.0011	0.0010	0.0011	0.0012	0.0005	0.0006	0.0007	0.0009
$MSE(\hat{S}_i^{G2(w)}(t))$	0.0020	0.0016	0.0014	0.0013	0.0009	0.0008	0.0007	0.0006
PCoB in MSE	1.4644	7.8561	18.3718	31.0599	2.8296	14.6401	31.7096	47.4874
$MSE(\hat{S}_i^{G2(reg)}(t))$	0.0005	0.0005	0.0005	0.0005	0.0002	0.0002	0.0002	0.0002
PCoB in MSE	8.6726	9.1930	10.0751	10.4568	16.5111	16.8509	4.6246	4.4712
$RE(\hat{S}_i^{G2(reg)}(t))$	3.0875	3.0250	2.9752	2.9688	4.5246	4.5414	5.2367	5.2038
$MSE(\hat{S}_i^{G2(r)}(t))$	0.0005	0.0006	0.0006	0.0006	0.0003	0.0003	0.0002	0.0002
CoB in MSE	8.1042	8.6153	9.2866	9.6538	15.5793	15.9267	4.3229	4.1550
$RE(\hat{S}_i^{G2(r)}(t))$	3.1337	3.0727	3.0290	3.0263	4.6421	4.6533	5.2777	5.2496
$MSE(\hat{S}_i^{G2(creg)}(t))$	0.0006	0.0006	0.0006	0.0006	0.0003	0.0003	0.0003	0.0003
CoB in MSE	3.9339	4.3188	6.6430	4.8717	7.8994	8.2417	2.1427	1.9504
$RE(\hat{S}_i^{G2(creg)}(t))$	2.9308	2.8838	2.8552	2.8562	4.5243	4.5218	4.8284	4.8162
$MSE(\hat{S}_i^{G2(cr)}(t))$	0.0006	0.0006	0.0006	0.0006	0.0003	0.0003	0.0003	0.0003
CoB in MSE	3.5589	3.9239	4.1214	4.3306	7.2005	7.5368	1.9515	1.7606
$RE(\hat{S}_i^{G2(cr)}(t))$	2.1437	2.1174	2.0827	2.0824	2.0502	2.0517	2.1707	2.1695
$MSE(\hat{S}_i^{G3(reg)}(t))$	0.0005	0.0005	0.0005	0.0005	0.0003	0.0003	0.0002	0.0003
CoB in MSE	17.0451	28.3981	37.0631	47.0547	29.7664	44.5596	42.7283	55.6620
$RE(\hat{S}_i^{G3(reg)}(t))$	3.4228	3.5357	3.6713	3.5137	4.6492	4.4650	5.5062	4.8032
$MSE(\hat{S}_i^{G3(r)}(t))$	0.0005	0.0005	0.0005	0.0005	0.0003	0.0003	0.0002	0.0003
CoB in MSE	16.5973	28.2556	37.4442	47.8474	29.1798	44.4529	43.7458	56.8511
$RE(\hat{S}_i^{G3(r)}(t))$	3.4490	3.5265	3.5900	3.3746	4.7229	4.4661	5.3050	4.5382
$MSE(\hat{S}_i^{G4(reg)}(t))$	0.0007	-	-	-	0.0005	-	-	-
CoB in MSE	60.2962	-	-	-	68.3958	-	-	-
$RE(\hat{S}_i^{G4(reg)}(t))$	2.8306	-	-	-	3.0548	-	-	-
$MSE(\hat{S}_i^{G4(r)}(t))$	0.0007	-	-	-	0.0005	-	-	-
CoB in MSE	61.0909	-	-	-	74.1180	-		-
$RE(\hat{S}_i^{G4(r)}(t))$	2.6390	-	-	-	2.8293	-	-	-

Graphical comparison of different estimators:

Using the best estimator from the four groups in the previous step of efficiency comparison. The graphical comparison of these estimators is obtained by comparing the estimates with respect to MSE's and PCoB's. The comparison of estimates can be seen in the following figure i.e. 3.1, 3.2, 3.3, 3.4. These figures give the efficiency comparisons for estimators for different regions of Pakistan for sample sizes 500 and 1000.

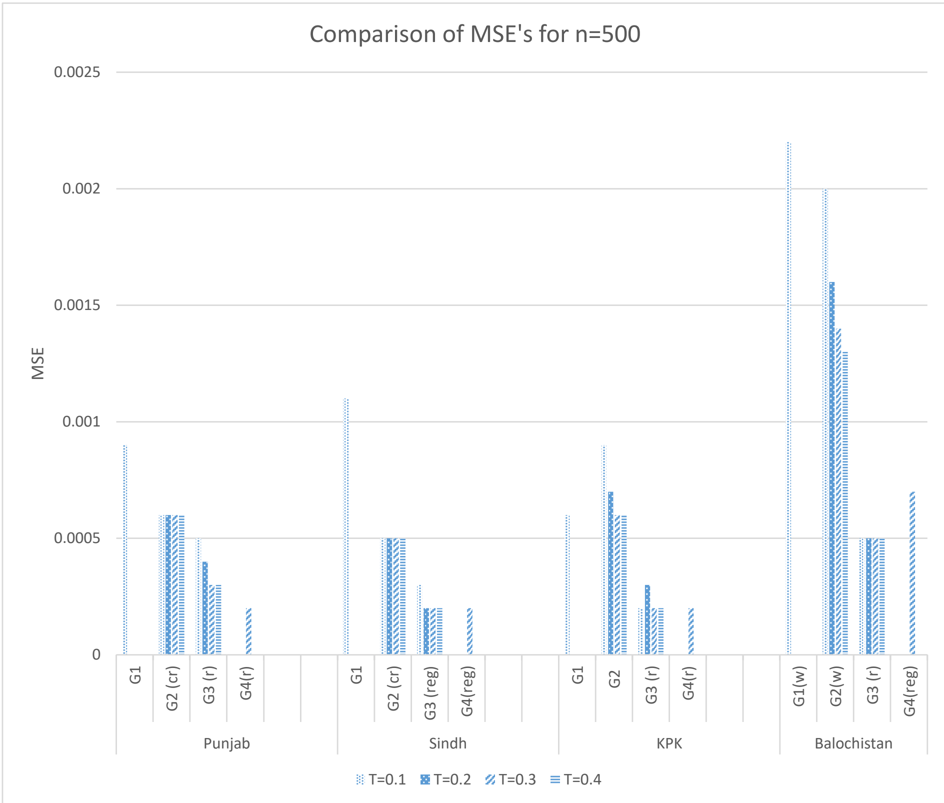

Figure 3.1: Graphical comparison of different estimators of the proposed group with respect to MSE.

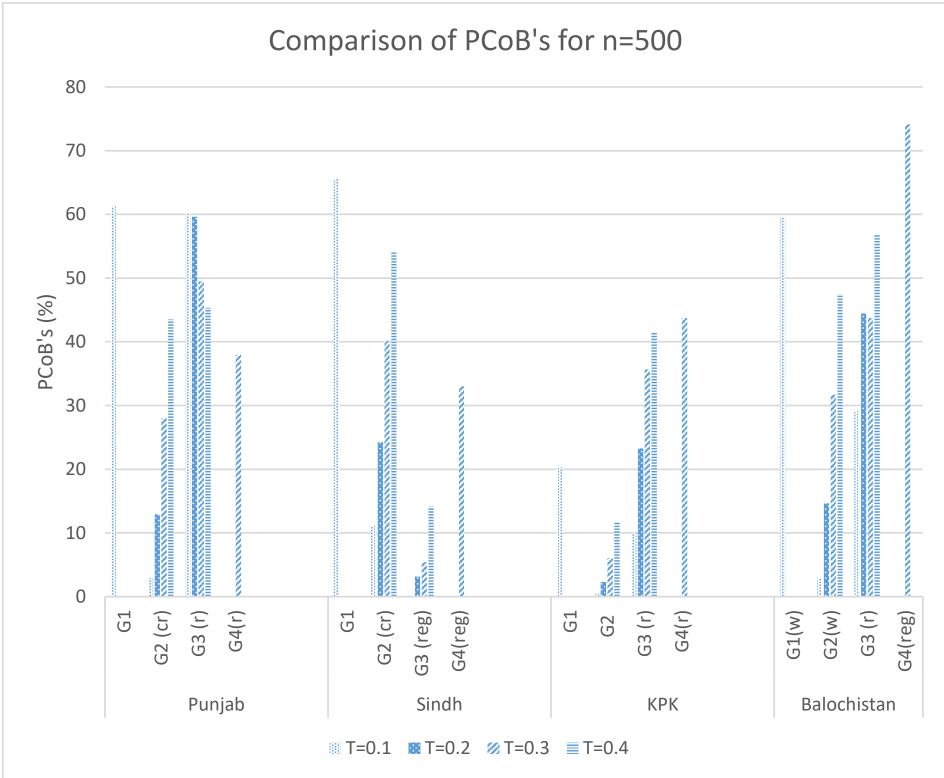

Figure 3.2: Graphical comparison of different estimators of the proposed group with respect to PCoB's.

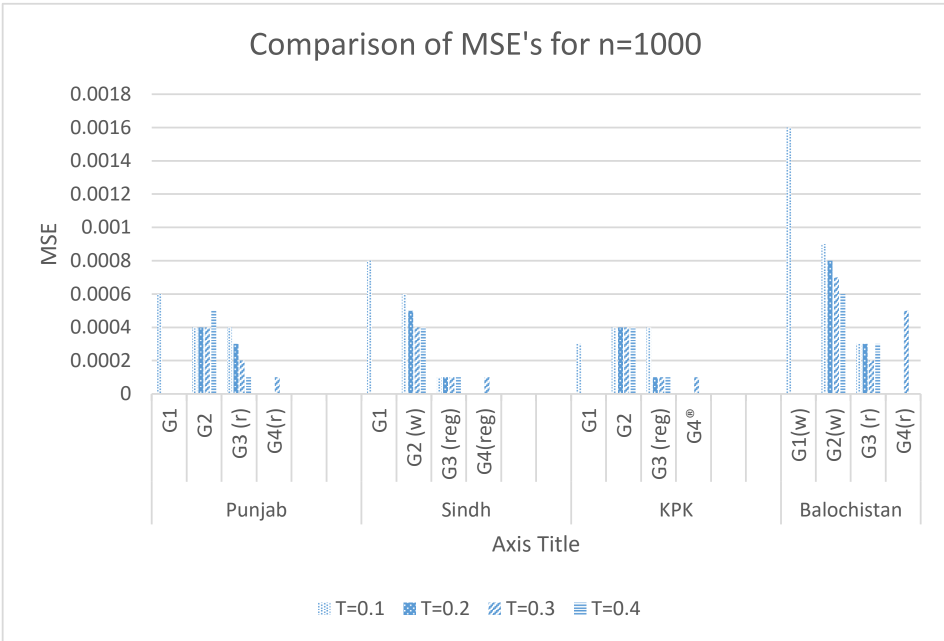

Figure 3.3: Graphical comparison of different estimators of the proposed group with respect to MSE.

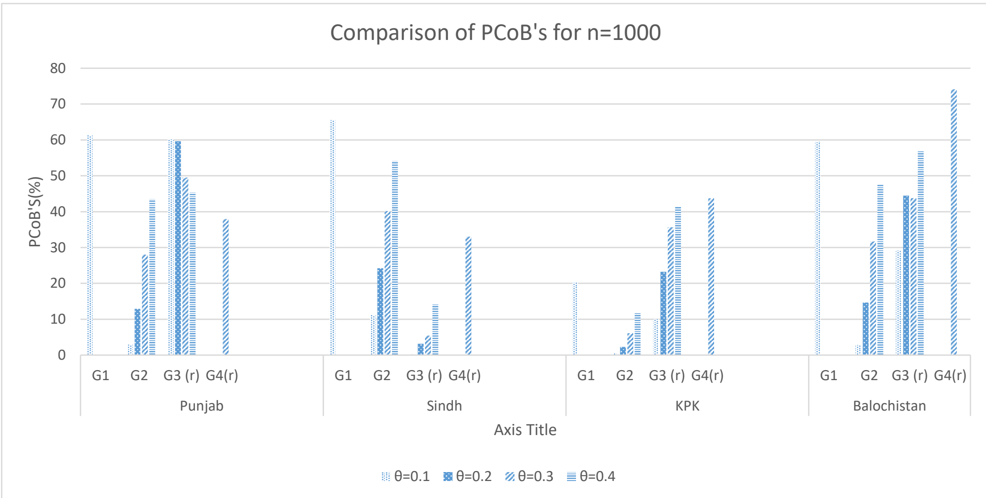

Figure 3.4: Graphical comparison of different estimators of the proposed group with respect to PCoB's.

3.3 Application to Health-related Parameters Estimations in Districts

DHS reports contain estimates on survival health indicators at regional levels which cover four provinces, Punjab, Sindh, KPK, and Balochistan. However, for policy implementation authorities need estimates of the major indicators of survival. The suggested strategies in this article help in producing district-level estimates using auxiliary characters at the population level and combining the samples obtained from multiple surveys. To check the implementation of the strategies, we select samples of larger sizes i.e. n = n2 = 2000 from the two surveys and used a bootstrap algorithm to obtain expected sample sizes at different districts. Table 3.7 provides expected sample sizes in different districts for the sample obtained from the current survey (ESS2) and the combined survey (ESSC) in Columns 2, 5, 8 and 3, 6, 9 respectively. Accumulation of two surveys produced a significant increase in sample size on the current occasion. Due to unavoidable circumstances, DHS contains no observation in District Kalat, Gwadar, and Jafarabad for PDHS 2017-18-special. For similar reasons, the expected sample sizes in District Matiari, Mandi Bahaudin, Shikarpur, Jhelum and Batagram are less than 5 for the current occasion. This is due to the fact that some samples consist ($n_i < 2$).

The table 3.7 below gives information on the expected sample sizes (ESS2)for the current occasion and the expected sample size (ESSC) for the combined occasion for different districts in Pakistan. The table lists various districts of Pakistan and their current occasion sample size and combined occasion sample size, reflecting the country's geographical variety and administrative divisions.

The goal of this strategy is to overcome the problem of limited sample sizes by making use of data from previous events. The smaller (ESS2) values suggest a relatively small sample size for estimation on this occasion. Small sample values, however, may result in estimates that are less accurate and have lower statistical significance.

The strategy of merging data from current and previous occasions (ESSC) is used to address the problem of limited sample sizes. This facilitates obtaining bigger sample sizes and, as a result, more reliable estimations.

Researchers can achieve a balance between the drawbacks of using small sample sizes and the advantages of using a greater, diverse set of observations by merging data.

Table 3.7: Expected Sample Size for Different Districts of Pakistan

District	ESS2	ESSC	District	ESS2	ESSC	District	ESS2	ESSC
Attock	66	127	Korangi	117	155	Killa Saifullah	92	215
Bahawalnagar	94	235	Larkana	142	268	Kohlu	44	106
Bahawalpur	130	332	Matiari	32	214	Lasbela	264	567
Bhakkar	73	202	Mirpurkhas	70	223	Lehri	71	281
Chakwal	59	138	Naushahro Firoze	88	255	Loralai	37	68
Chiniot	67	143	Sanghar	77	212	Mastung	61	121
D.G.Khan	67	240	Shaheed Benazirabad	96	233	Pishin	144	311
Faisalabad	214	663	Shikarpur	31	238	Quetta	37	159
Gujranwala	130	342	Sujawal	97	183	Sherani	85	224
Gujrat	84	193	Sukkur	72	142	Sibi	74	160
Hafizabad	74	102	Tando Allahyar	46	192	Sohbatpur	0	143
Jhang	83	244	Tando MK	37	178	Washuk	62	229
Jhelum	30	62	Tharparkar	82	178	Zhob	0	0
Kasur	117	271	Thatta	86	86	Astore	0	93
Khanewal	85	227	Umerkot	38	38	Diamer	95	595
Khushab	45	87	Abbottabad	76	206	Ghanche	286	324
Lahore	243	598	Bannu	56	215	Ghizer	28	90

Continued on next page

Table 3.7 continued from previous page

District	ESS2	ESSC	District	ESS2	ESSC	District	ESS2	ESSC
Layyah	96	209	Batagram	34	122	Gilgit	25	25
Lodhran	66	139	Buner	44	99	Hunza	23	47
Mandi Bahaudin	21	94	Charsadda	108	294	Kharmang	62	159
Mianwali	29	110	Chitral	44	133	Nagar	49	49
Lahore	243	598	Bannu	56	215	Ghizer	28	90
Layyah	96	209	Batagram	34	122	Gilgit	25	25
Lodhran	66	139	Buner	44	99	Hunza	23	47
Mandi Bahaudin	21	94	Charsadda	108	294	Kharmang	62	159
Mianwali	29	110	Chitral	44	133	Nagar	49	49
Nankana S	22	140	Haripur	86	187	Islamabad	154	622
Narowal	59	121	Karak	31	125	Bajaur Agency	122	393
Okara	83	232	Kohat	101	203	Bannu	307	587
Pakpattan	47	177	Lakki Marwat	0	0	Khyber Agency	42	459
Rahim Yar Khan	184	422	Lower Dir	65	186	Kurram Agency	42	150
Rajanpur	41	71	Malakand (PA)	82	279	Mohmand Agency	59	59
Rawalpindi	130	341	Mansehra	56	112	North Waziristan	66	66
Sahiwal	69	186	Mardan	73	252	Orakzai Agency	187	187
Sargodha	101	297	Nowshera	136	500	South Waziristan	920	2048

Continued on next page

Table 3.7 continued from previous page

District	ESS2	ESSC	District	ESS2	ESSC	District	ESS2	ESSC
Sheikhupura	75	266	Peshawar	174	375	Bagh	131	131
Sialkot	107	291	Shangla	551	1146	Bhimber	127	127
Toba Tek Singh	79	286	Swabi	33	174	Hattian Bala	320	320
Vehari	70	221	Swat	116	316	Haveli	188	188
Badin	103	280	Tank	172	428	Kotli	75	75
Dadu	97	268	Tor Ghar	81	168	Mirpur	199	199
Ghotki	113	305	Upper Dir	38	82	Muzaffarabad	33	33
Hyderabad	86	404	Awaran	78	194	Neelum	119	119
Jacobabad	121	316	Dera Bugti	36	36	Poonch	172	172
Jamshoro	48	91	Gwadar	0	43	Sudhonti	89	89
Shahdadkot	95	347	Jafarabad	0	97	Karachi West	219	268
Karachi Central	151	364	Jhal Magsi	36	104	Kashmore	47	261
Karachi East	138	276	Kachhi (B)	85	85	Khairpur	163	292
Karachi Malir	107	229	Kalat	0	110	Kharan	31	384
Karachi South	78	326	Kech (Tur)	39	74	Khuzdar	33	85

Figure 3.5 displays the expected sample sizes in different districts for the pooled sample. In the legends, there are three dots, i.e., mini, small, and big. The mini dot represents the small sample size, the small dot represents the relatively large sample size and the big dot represents the large sample size, which can be easily seen using the map.

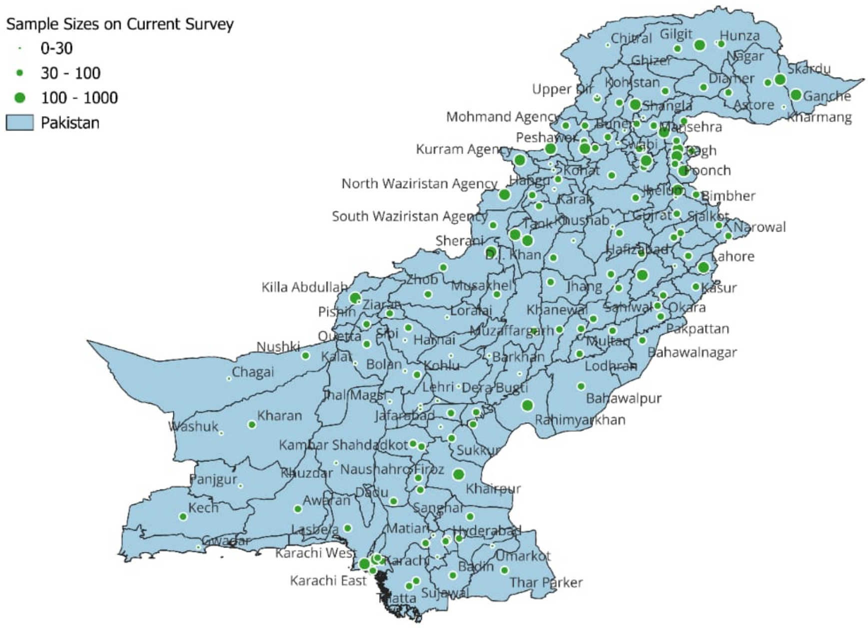

Figure 3.5: Expected sample size for different districts of Pakistan

Similarly, Figure 3.6 displays the expected sample sizes in different districts for the pooled sample for combined occasions. By using these maps it is easy to find all those districts of Pakistan where sample size is limited to obtain reliable estimates. We can increase the sample size by combining the two data sets.

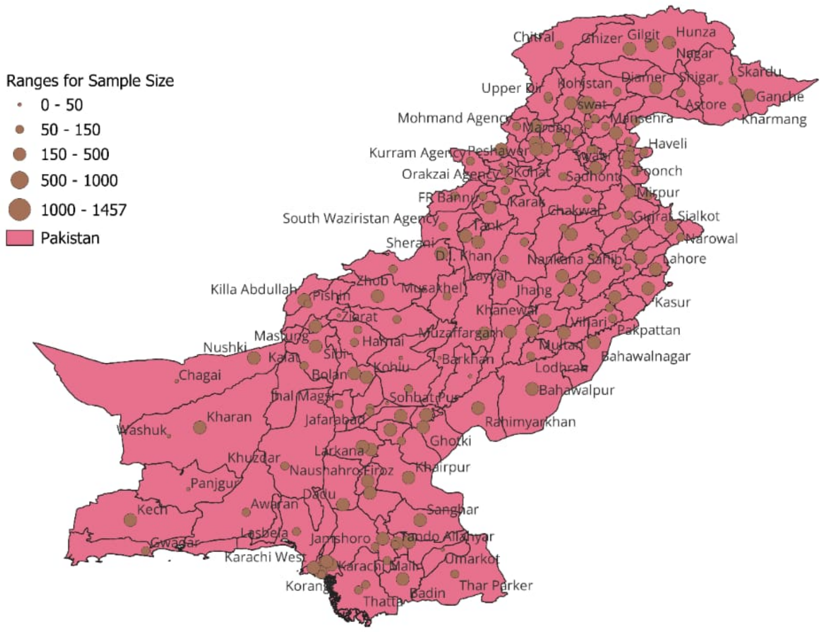

Figure 3.6: Expected sample size for both occasions different districts of Pakistan

3.4 Estimations of Parameters Districts

The bootstrap study produces district-level estimates under different strategies. Based on the previous numerical comparison, the estimated average number of infants that survived in different districts is provided for the cases with the highest accuracy along with root mean squared error (RMSE). Assuming asymptotic normality of the estimators i.e. $\frac{\left[\hat{S}_i^{SK(\cdot)} - S_i(t)\right]}{RMSE\left[\hat{S}_i^{SK(\cdot)}\right]} = Z$, we can construct a (1 - a) confidence interval using the four best estimators from the four groups, the table (3.8) gives a thorough overview of the estimated average infant mortality in different districts of Pakistan. The table displays four distinct estimators from different groups. The best estimator from the four groups is selected by using the percentage contribution of bias (PCoB) or the relative efficiency (RE). Four estimators were selected from the four groups and were used to find estimates for the district level. The root mean squared error (RMSE), which measures how well are the estimators, is provided along with each assessment. The districts are listed in the leftmost column, giving Pakistan's major geographic regions. The estimated infant mortality for each district based on the four estimators is shown in the following columns. Lower (RMSE) values signify a better match between the predicted results and the real data, and they provide insights into the estimator's predictive ability.

Table 3.8: Estimated Average Number of Infant Mortality Rate in Different Districts of Pakistan

District	$\hat{S}_i^{S1}(t)$ (RMSE)	$\hat{S}_i^{S2(reg)}(t)$ (RMSE)	$\hat{S}_i^{S3(creg)}(t)$ (RMSE)	$\hat{S}_i^{S4(reg)}(t)$ (RMSE)
Attock	0.8770 (0.0205)	0.8774 (0.0186)	0.8694 (0.0167)	0.8677 (0.0174)
Bahawalnagar	0.7833 (0.0213)	0.8493 (0.0575)	0.8499 (0.0566)	0.8502 (0.0566)
Bahawalpur	0.8024 (0.0153)	0.8717 (0.0691)	0.8626 (0.0592)	0.8578 (0.0544)
Bhakkar	0.8295 (0.0376)	0.8675 (0.0186)	0.8675 (0.0128)	0.8675 (0.0113)
Chakwal	0.8581 (0.0204)	0.8850 (0.0390)	0.8830 (0.0338)	0.8822 (0.0324)
Chiniot	0.8655 (0.0344)	0.8726 (0.0284)	0.8801 (0.0188)	0.8826 (0.0164)
Dera Ghazi Khan	0.7983 (0.0544)	0.8546 (0.0196)	0.8508 (0.0124)	0.8478 (0.0099)
Faisalabad	0.8233 (0.0566)	0.8890 (0.0141)	0.8749 (0.0077)	0.8652 (0.0149)
Gujranwala	0.8531 (0.0470)	0.8919 (0.0147)	0.8769 (0.0231)	0.8687 (0.0305)
Gujrat	0.8805 (0.0296)	0.9028 (0.0146)	0.8988 (0.0119)	0.8972 (0.0123)
Hafizabad	0.8327 (0.0266)	0.8611 (0.0370)	0.8620 (0.0367)	0.8617 (0.0363)
Jhang	0.8218 (0.0387)	0.8747 (0.0249)	0.8643 (0.0133)	0.8575 (0.0089)
Jhelum	0.9172 (0.0579)	0.9230 (0.0528)	0.8985 (0.0738)	0.8911 (0.0808)
Kasur	0.8507 (0.0200)	0.8717 (0.0158)	0.8700 (0.0110)	0.8692 (0.0099)
Khanewal	0.8257 (0.0212)	0.8631 (0.0525)	0.8639 (0.0517)	0.8644 (0.0518)
Khushab	0.8013 (0.0677)	0.9011 (0.0432)	0.8770 (0.0209)	0.8720 (0.0179)
Lahore	0.8800 (0.0142)	0.8941 (0.0095)	0.8882 (0.0068)	0.8853 (0.0080)

Continued on next page

Table 3.8 – Continued from previous page

District	$\hat{S}_i^{G1}(t)$ (RMSE)	$\hat{S}_i^{G2(reg)}(t)$ (RMSE)	$(\hat{S}_i^{G3(creg)}(t))$ (RMSE)	$\hat{S}_i^{G4(r)}(t)$ (RMSE)
Layyah	0.8320 (0.0180)	0.8649 (0.0383)	0.8688 (0.0401)	0.8701 (0.0412)
Lodhran	0.7786 (0.0351)	0.8378 (0.0868)	0.8464 (0.0941)	0.8491 90.0966)
Mandi Bahaudin	0.7848 (0.0969)	0.8693 (0.0349)	0.8589 (0.0284)	0.8493 (0.0320)
Mianwali	0.8707 (0.0438)	0.8857 (0.0370)	0.8735 (0.0398)	0.8633 (0.0473)
Multan	0.8317 (0.0137)	0.8931 (0.0556)	0.8870 (0.0487)	0.8829 (0.0444)
Muzaffargarh	0.8268 (0.0602)	0.8934 (0.0172)	0.8787 (0.0121)	0.8653 (0.0216)
Nankana	0.8025 (0.1018)	0.9093 (0.0324)	0.8891 (0.0237)	0.8669 (0.0369)
Narowal	0.8472 (0.0352)	0.8916 (0.0274)	0.8787 (0.0149)	0.8751 (0.0130)
Okara	0.8538 (0.0176)	0.8727 (0.0216)	0.8754 (0.0177)	0.8771 (0.0175)
Pakpattan	0.8221 (0.0191)	0.8347 (0.0261)	0.8465 (0.0282)	0.8564 (0.0356)
Rahim	0.8039 (0.0194)	0.8595 (0.0429)	0.8581 (0.0406)	0.8575 (0.0399)
Rajanpur	0.8367 (0.0427)	0.8244 (0.0311)	0.8632 (0.0595)	0.8658 (0.0619)
Rawalpindi	0.8767 (0.0311)	0.8901 (0.0200)	0.8808 (0.0261)	0.8756 (0.0308)
Sahiwal	0.8109 (0.0364)	0.8752 (0.0382)	0.8685 (0.0293)	0.8647 (0.0250)
Sargodha	0.8610 (0.0138)	0.8737 (0.0199)	0.8758 (0.0178)	0.8772 (0.0181)
Sheikhupura	0.8062 (0.0341)	0.8789 (0.0466)	0.8642 (0.0304)	0.8525 (0.0184)
Sialkot	0.8905 (0.0303)	0.8944 (0.0273)	0.8892 (0.0304)	0.8862 (0.0330)
Toba tek s	0.8354 (0.0776)	0.9024 (0.0176)	0.8831 (0.0300)	0.8709 (0.0414)

Continued on next page

Table 3.8 – Continued from previous page

District	$\hat{S}_i^{G1}(t)$ (RMSE)	$\hat{S}_i^{G2(reg)}(t)$ (RMSE)	$(\hat{S}_i^{G3(creg)}(t))$ (RMSE)	$\hat{S}_i^{G4(r)}(t)$ (RMSE)
Vehari	0.8473 (0.0188)	0.8724 (0.0395)	0.8684 (0.0330)	0.8655 (0.0294)
Badin	0.8286 (0.0193)	0.8592 (0.0452)	0.8660 (0.0502)	0.8700 (0.0537)
Dadu	0.8768 (0.0831)	0.9094 (0.0514)	0.8877 (0.0718)	0.8748 (0.0845)
Ghotki	0.7567 (0.0191)	0.8316 (0.0660)	0.8307 (0.0640)	0.8301 (0.0632)
Hyderabad	0.8394 90.0176)	0.8667 (0.0227)	0.8694 (0.0205)	0.8720 (0.0211)
Jacobabad	0.7974 (0.0216)	0.8525 (0.0430)	0.8547 (0.0436)	0.8558 (0.0444)
Jamshoro	0.8654 (0.0250)	0.9126 (0.0467)	0.9095 (0.0412)	0.9090 (0.0405)
Kambar	0.9137 (0.0119)	0.9016 (0.0151)	0.8968 (0.0137)	0.8930 (0.0156)
Karachi Central	0.8858 (0.0458)	0.9106 (0.0223)	0.8957 (0.0352)	0.8889 (0.0419)
Karachi East	0.8877 (0.0441)	0.9038 (0.0281)	0.8923 (0.0384)	0.8894 (0.0412)
Karachi Malir	0.8954 (0.0139)	0.8879 (0.0162)	0.8811 (0.0182)	0.8788 (0.0199)
Karachi South	0.8991 (0.0400)	0.8964 (0.0444)	0.8989 (0.0400)	0.9011 (0.0370)
Karachi West	0.8752 (0.0268)	0.8980 (0.0100)	0.8900 (0.0115)	0.8944 (0.0092)
Kashmore	0.8338 (0.0835)	0.8935 (0.0292)	0.8862 (0.0319)	0.8786 (0.0382)
Khairpur	0.8613 (0.0185)	0.8920 (0.0222)	0.8856 (0.0153)	0.8849 (0.0147)
Korangi	0.9145 (0.0429)	0.9096 (0.0470)	0.8916 (0.0642)	0.8986 (0.0572)
Larkana	0.8735 (0.0170)	0.8780 (0.0190)	0.8843 (0.0220)	0.8853 (0.0228)
Matiari	0.8180 (0.0612)	0.8758 (0.0264)	0.8659 (0.0196)	0.8547 (0.0239)

Continued on next page

Table 3.8 – Continued from previous page

District	$\hat{S}_i^{G1}(t)$ (RMSE)	$\hat{S}_i^{G2(reg)}(t)$ (RMSE)	$(\hat{S}_i^{G3(creg)}(t))$ (RMSE)	$\hat{S}_i^{G4(r)}(t)$ (RMSE)
Mirpurkhas	0.8508 (0.0336)	0.8970 (0.0228)	0.8821 (0.0104)	0.8715 (0.0124)
Naushahro Firoze	0.8437 (0.0487)	0.8869 (0.0145)	0.8844 (0.0111)	0.8828 (0.0112)
Sanghar	0.8487 (0.0620)	0.8917 (0.0240)	0.8823 (0.0285)	0.8767 (0.0331)
Shaheed Benazirabad	0.8638 (0.0164)	0.8613 (0.0158)	0.8645 (0.0122)	0.8660 (0.0121)
Shikarpur	0.8395 (0.0517)	0.8839 (0.0279)	0.8756 (0.0213)	0.8658 (0.0246)
Sujawal	0.8697 (0.0280)	0.8696 (0.0261)	0.8654 (0.0280)	0.8646 (0.0286)
Sukkur	0.8323 (0.0352)	0.8414 (0.0404)	0.8667 (0.0630)	0.8726 (0.0687)
Tando Allahyar	0.8512 (0.0176)	0.8719 (0.0345)	0.8759 (0.0331)	0.8795 (0.0347)
Tando Mohammad Khan	0.8297 (0.0181)	0.8739 (0.0497)	0.8678 (0.0398)	0.8619 (0.0320)
Tharparkar	0.8755 (0.0424)	0.8940 (0.0253)	0.8894 (0.0271)	0.8879 (0.0284)
Thatta	0.8982 (0.0243)	0.8823 (0.0228)	Na [5] (Na)	0.8823 (0.0228)
Umerkot	0.9418 (0.0295)	0.9006 (0.0466)	Na (Na)	0.9006 (0.0466)
Abbottabad	0.8684 (0.0270)	0.8962 (0.0193)	0.8907 (0.0121)	0.8875 (0.0104)
Bannu	0.9283 (0.0403)	0.8764 (0.0256)	0.8897 (0.0138)	0.9011 (0.0143)
Batagram	0.8929 (0.0626)	0.8940 (0.0687)	0.8937 (0.0636)	0.8934 (0.0615)
Buner	0.9112 (0.0458)	0.8690 (0.0230)	0.8878 (0.0234)	0.8950 (0.0284)
Charsadda	0.8607 (0.0192)	0.8907 (0.0220)	0.8837 (0.0133)	0.8796 (0.0095)

Continued on next page

[5] NaN

Table 3.8 – Continued from previous page

District	$\hat{S}_i^{G1}(t)$ (RMSE)	$\hat{S}_i^{G2(reg)}(t)$ (RMSE)	$(\hat{S}_i^{G3(creg)}(t))$ (RMSE)	$\hat{S}_i^{G4(r)}(t)$ (RMSE)
Chitral	0.8418 (0.1192)	0.9215 (0.0434)	0.9002 (0.0606)	0.8858 (0.0744)
D.I Khan	0.8630 (0.0256)	0.8519 (0.0170)	0.8666 (0.0276)	0.8730 (0.0336)
Hangu	0.8509 (0.0231)	0.8845 (0.0376)	0.8802 (0.0282)	0.8771 (0.0235)
Haripur	0.8754 (0.0239)	0.8758 (0.0238)	0.8682 (0.0268)	0.8655 (0.0289)
Karak	0.9214 (0.0269)	0.8877 (0.0308)	0.8952 (0.0184)	0.9018 (0.0117)
Kohat	0.8745 (0.0164)	0.8908 (0.0256)	0.8908 (0.0234)	0.8908 (0.0232)
Lakki Marwat	Na (Na)	Na (Na)	Na (Na)	Na (Na)
Lower Dir	0.9106 (0.0254)	0.9114 (0.0262)	0.9016 (0.0322)	0.8954 (0.0376)
Malakand (PA)	0.9070 (0.0311)	0.9068 (0.0327)	0.8997 (0.0374)	0.8943 (0.0422)
Mansehra	0.8996 (0.0199)	0.9202 (0.0244)	0.9162 (0.01740	0.9152 (0.0162)
Mardan	0.8886 (0.0350)	0.8791 (0.0451)	0.8825 (0.0401)	0.8852 (0.0370)
Nowshera	0.8690 (0.0388)	0.9072 (0.0119)	0.8962 (0.0129)	0.8872 (0.0202)
Peshawar	0.8622 (0.0294)	0.8792 (0.0143)	0.8810 (0.0108)	0.8816 (0.0101)
Shangla	0.8984 (0.0072)	0.8952 (0.0064)	0.8933 (0.0047)	0.8928 (0.0047)
Swabi	0.8424 (0.1190)	0.8847 (0.0786)	0.8715 (0.0896)	0.8579 (0.1025)
Swat	0.8762 (0.0133)	0.8691 (0.0186)	0.8758 (0.0106)	0.8796 (0.0076)
Tank	0.8731 (0.0224)	0.8905 (0.0114)	0.8879 (0.0088)	0.8866 (0.0089)
Tor Ghar	0.9301 (0.0259)	0.8973 (0.0197)	0.8946 (0.0172)	0.8938 (0.0172)

Continued on next page

Table 3.8 – Continued from previous page

District	$\hat{S}_i^{G1}(t)$ (RMSE)	$\hat{S}_i^{G2(reg)}(t)$ (RMSE)	$(\hat{S}_i^{G3(creg)}(t))$ (RMSE)	$\hat{S}_i^{G4(r)}(t)$ (RMSE)
Upper Dir	0.8565 (0.0572)	0(8957 (0.0253)	0.8732 (0.0379)	0.8655 (0.0447)
Awaran	0.9347 (0.0130)	0.9065 (0.0343)	0.9071 (0.0322)	0.9074 (0.0317)
Dera Bugti	0.9525 (0.0272)	0.8961 (0.0598)	Na (Na)	0.8961 (0.0598)
Gwadar	0.9571 (Na)	Na (Na)	Na (Na)	0.9176 (Na)
Jafarabad	0.9878 (Na)	Na (Na)	Na (Na)	0.9173 (Na)
Jhal Magsi	0.8924 (0.0855)	0.9052 (0.0737)	0.8974 (0.0795)	0.8924 (0.0841)
Kachhi (Bolan)	0.9481 (0.0181)	0.9187 (0.0328)	Na (Na)	0.9187 (0.0328)
Kalat	0.9011 (Na)	Na (Na)	Na (Na)	0.8882 (Na)
Kech (Turbat)	0.9103 (0.0493)	0.9159 (0.0443)	0.9130 (0.0442)	0.9124 (0.0445)
Kharan	0.8255 (0.0908)	0.8822 (0.0427)	0.8763 (0.0424)	0.8687 (0.0473)
Khuzdar	0.8589 (0.0622)	0.8315 (0.0399)	0.8512 (0.0518)	0.8613 (0.0607)
Killa Abdullah	0.8155 (0.0939)	0.9060 (0.0125)	0.8865 (0.0233)	0.8743 (0.0348)
Killa saifullah	0.8780 (0.0163)	0.8951 (0.0202)	0.8944 (0.0162)	0.8941 (0.0153)
Kohlu	0.8118 (0.1110)	0.8900 (0.0375)	0.8563 (0.0656)	0.8406 (0.0806)
Lasbela	0.8128 (0.0327)	0.8632 (0.0217)	0.8551 (0.0132)	0.8524 (0.0106)
Lehri	0.9240 (0.0405)	0.9236 (0.0425)	0.9111 (0.0530)	0.9004 (0.0632)
Loralai	0.9531 (0.0208)	0.9334 (0.0378)	0.9209 (0.0465)	0.9193 (0.0479)
Mastung	0.9725 (0.0164)	0.9214 (0.0658)	0.9206 (0.0657)	0.9205 (0.0658)

Continued on next page

Table 3.8 – Continued from previous page

District	$\hat{S}_i^{G1}(t)$ (RMSE)	$\hat{S}_i^{G2(reg)}(t)$ (RMSE)	$(\hat{S}_i^{G3(creg)}(t))$ (RMSE)	$\hat{S}_i^{G4(r)}(t)$ (RMSE)
Pishin	0.8792 (0.0169)	0.8842 (0.0136)	0.8893 (0.0082)	0.8910 (0.0075)
Quetta	0.9060 (0.0380)	0.9468 (0.0253)	0.9289 (0.0197)	0.9125 (0.0303)
Sherani	0.7740 (0.0639)	0.8749 (0.0435)	0.8573 (0.0250)	0.8477 (0.0158)
Sibi	0.8931 (0.0245)	0.8979 (0.0288)	0.8993 (0.0269)	0.8998 (0.0268)
Sohbatpur	0.8342 (NaN)	NaN (NaN)	NaN (NaN)	0.8615 (NaN)
Washuk	0.8473 (0.0428)	0.8898 (0.0193)	0.8794 (0.0145)	0.8708 (0.0187)
Zhob	Na (Na)	Na (Na)	Na (Na)	Na (Na)
Astore	0.9210 (Na[6])	Na (Na)	Na (Na)	0.8895 (Na)
Diamer	0.8447 (0.1092)	0.9089 (0.0462)	0.8916 (0.0624)	0.8725 (0.0812)
Ghanche	0.8983 (0.0154)	0.8844 (0.0101)	0.8972 (0.0114)	0.8881 (0.0081)
Ghizer	0.9328 (0.0692)	0.9593 (0.0453)	0.9386 (0.0628)	0.9236 (0.0773)
Gilgit	0.8434 (0.0557)	0.8602 (0.0373)	NaN (NaN)	0.8602 (0.0373)
Hunza	0.9290 (0.0509)	0.9313 (0.0528)	0.9523 (0.0683)	0.9574 (0.0728)
Kharmang	0.7759 (0.1898)	0.8993 (0.0672)	0.8609 (0.1042)	0.8407 (0.1244)
Nagar	0.8308 (0.0402)	0.8486 (0.0299)	Na (Na)	0.8486 (0.0299)
Shigar	0.8542 (0.0263)	0.9070 (0.0389)	0.8884 (0.0196)	0.8761 (0.0108)
Skardu	0.8208 (0.0594)	0.8910 (0.0206)	0.8797 (0.0104)	0.8683 (0.0124)

Continued on next page

[6] Not available

Table 3.8 – Continued from previous page

District	$\hat{S}_i^{G1}(t)$ (RMSE)	$\hat{S}_i^{G2(reg)}(t)$ (RMSE)	$(\hat{S}_i^{G3(creg)}(t))$ (RMSE)	$\hat{S}_i^{G4(r)}(t)$ (RMSE)
Islamabad	0.8046 (0.0112)	0.8452 (0.0396)	0.8576 (0.0503)	0.8687 (0.0609)
Bajaur Agency	0.8075 (0.0995)	0.8932 (0.0186)	0.8701 (0.0371)	0.8533 (0.0534)
Bannu	0.8950 (0.0136)	0.8956 (0.0128)	0.8898 (0.0166)	0.8887 (0.0176)
Khyber Agency	0.8859 (0.0518)	0.8962 (0.0458)	0.8949 (0.0438)	0.8933 (0.0439)
Kurram Agency	0.8738 (0.0621)	0.8719 (0.0614)	0.8763 (0.0634)	0.8798 (0.0663)
Mohmand Agency	0.8426 (0.0364)	0.8766 (0.0380)	Na (Na)	0.8766 (0.0380)
North Waziristan	0.8401 (0.0340)	0.8851 (0.0498)	Na (Na)	0.8851 (0.0498)
Orakzai Agency	0.8661 (0.0190)	0.8818 (0.0199)	Na (Na)	0.8818 (0.0199)
South Waziristan	0.9082 (0.0054)	0.8956 (0.0167)	0.8950 (0.0168)	0.8948 (0.0170)
Bagh	0.9498 (0.0144)	0.9179 (0.0335)	Na (Na)	0.9179 (0.0335)
Bhimber	0.9028 (0.0201)	0.8944 (0.0153)	Na (Na)	0.8944 (0.0153)
Hattian Bala	0.9111 (0.0121)	0.8865 (0.0256)	Na (Na)	0.8865 (0.0256)
Haveli	0.9300 (0.0140)	0.9206 (0.0132)	Na (Na)	0.9206 (0.0132)
Kotli	0.9475 (0.0190)	0.9280 (0.0253)	Na (Na)	0.9280 (0.0253)
Mirpur	0.9327 (0.0134)	0.9035 (0.0294)	Na (Na)	0.9035 (0.0294)
Muzaffarabad	0.9874 (0.0152)	0.9301 (0.0612)	Na (Na)	0.9301 (0.0612)
Neelum	0.9594 (0.0135)	0.9007 (0.0598)	Na (Na)	0.9007 (0.0598)
Poonch	0.8678 (0.0193)	0.8759 (0.0144)	Na (Na)	0.8759 (0.0144)

The proportion of survivors in different districts of Pakistan using the G1 estimator can be seen in Figure 3.7. There are three dots in the legend, mini, small, and big. Mini dot tells us the regions where the proportion of survivors is low. The small dot tells us the regions where the proportion of survivors is moderate, and the big dot tells us the regions where the proportion of survivors is high.

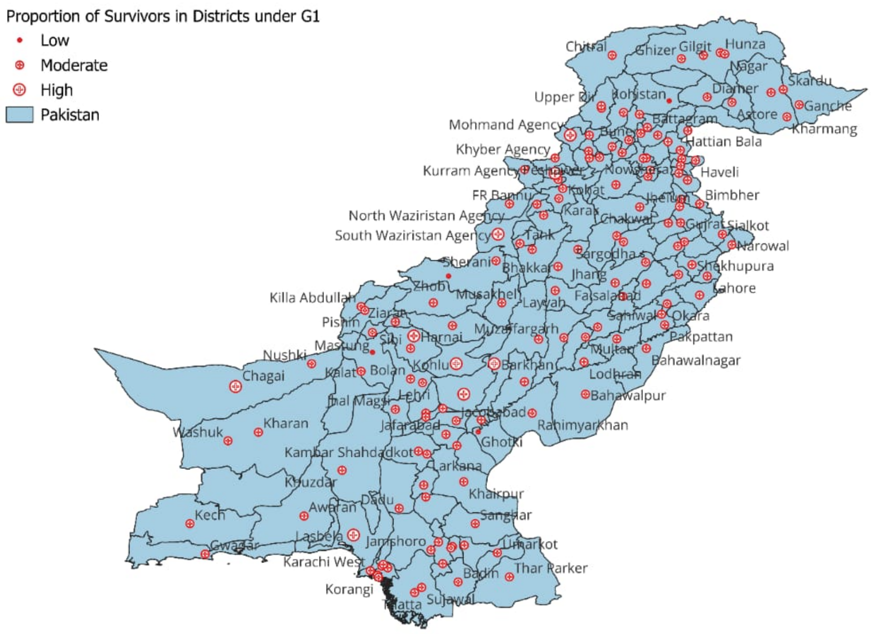

Figure 3.7: Proportion of survivors under G1 for different districts of Pakistan

The proportion of survivors in different districts of Pakistan using the G2 estimator can be seen in Figure 3.8.

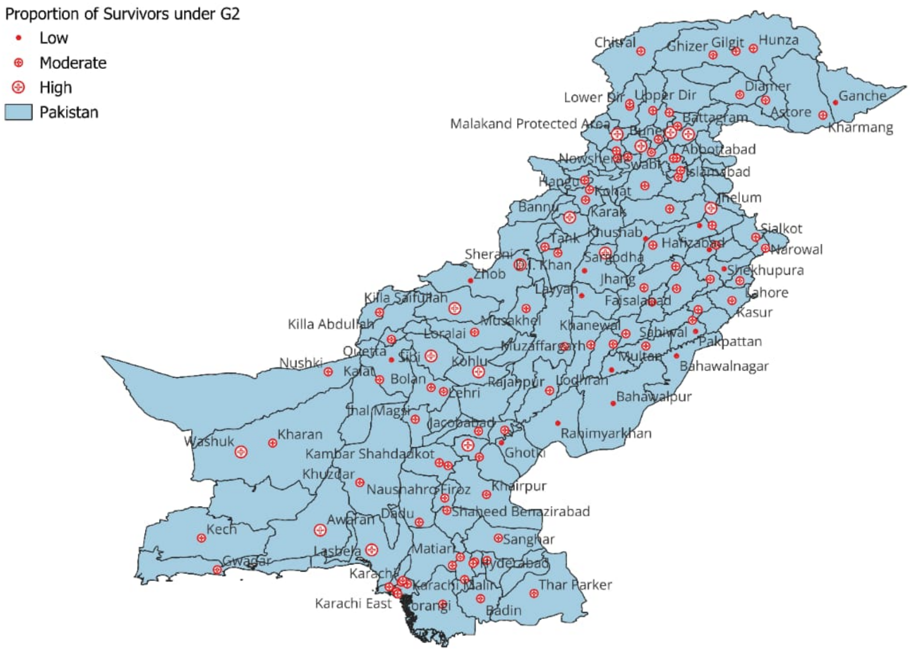

Figure 3.8: Proportion of survivors under G2 for different districts of Pakistan

The proportion of survivors in different districts of Pakistan using the G3 estimator can be seen in Figure 3.9.

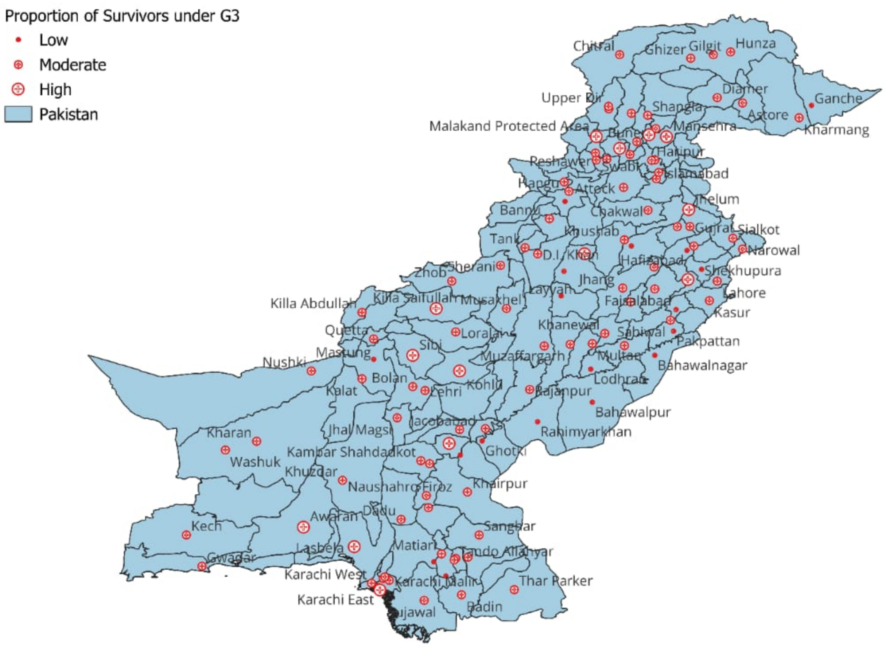

Figure 3.9: Proportion of survivors under G3 for different districts of Pakistan

The proportion of survivors in different districts of Pakistan using the G4 estimator can be seen in Figure 3.10.

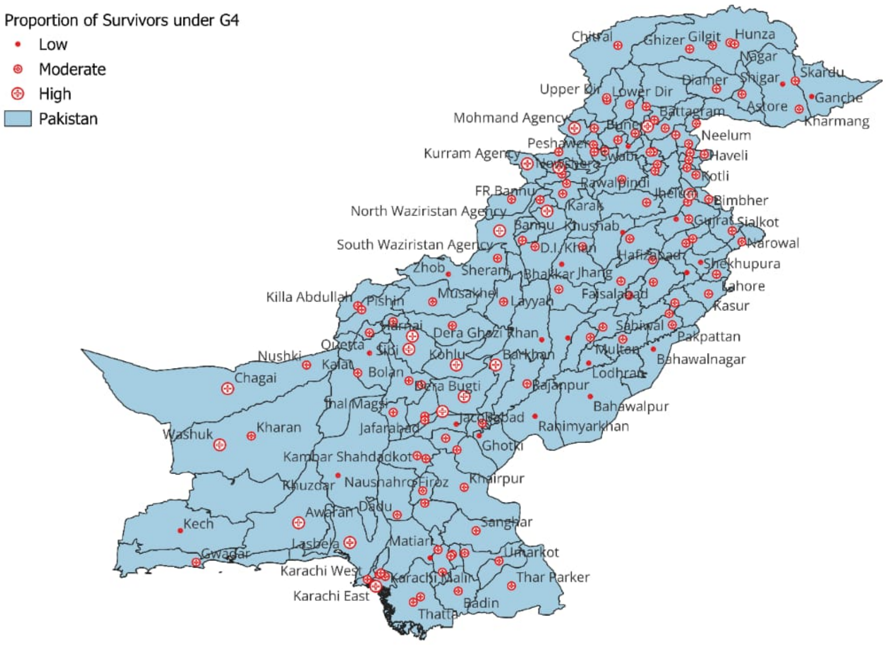

Figure 3.10: Proportion of survivors under G4 for different districts of Pakistan

Chapter 4

Conclusion and Recommendation

4.1 Conclusion

- **Improved Estimators for Survival Functions:** The study addresses the challenge of producing accurate estimates for regions with limited or no data on specific variables of interest, particularly in the context of survival functions. The improved estimators developed in this study have shown promise in providing more accurate insights into survival patterns within small areas.

- **Geographic Variation in Survival:** The findings of this study have demonstrated that survival functions can vary significantly by geographic traits. This variation underscores the importance of considering geographic sub-populations when estimating survival functions, especially in the context of infant mortality.

- **Data-Driven Policy Insights:** The study has contributed valuable insights into areas with limited or no available data. By providing more accurate estimates for these under-served regions, this study can guide public health policymakers in making informed decisions to improve health outcomes and allocate resources effectively.

4.2 Recommendation

1. **Further Research:** On the basis of the improved estimators, consider further research to refine and validate these methods. Explore potential enhancements or modifications that could make them even more robust in capturing survival patterns in diverse geographic sub-populations.

2. **Data Collection and Collaboration:** Encourage collaboration among researchers, public health agencies, and local communities to improve data collection efforts in areas with limited data availability. Increased data collection and sharing can further enhance the accuracy of your estimators.

3. **Tailored Interventions:** Recommend that public health policymakers use the insights provided by this study to tailor interventions and strategies for different geographic sub-populations. Recognize that one-size-fits-all approaches may not be effective, and targeted interventions can yield better results.

4. **Policy Implementation:** Work closely with policymakers to ensure that the recommendations derived from this research are translated into actionable policies and programs. Monitor and evaluate the impact of these policies to assess their effectiveness in improving health outcomes in small areas.

Bibliography

[1] @bookrao2015small, title=Small area estimation, author=Rao, John NK and Molina, Isabel, year=2015, publisher=John Wiley & Sons

[2] @onlineADP, author = Asian development bank, title = INTRODUCTION TO SMALL AREA ESTIMATION TECHNIQUES A Practical Guide for National Statistics Offices, year = may 2020, url = https://www.adb.org/sites/default/files/publication/609476/small-area-estimation-guide-nsos.pdf, note = Accessed on Feb 4, 2023

[3] @inbookSAE:ch1.1, title=Small area estimation, edition=2, author=J.N.K Rao, year=2015, publisher=A John Wiley and sons, inc, address=Canada, chapter=1.1

[4] @inbookSAE:ch1.2, title=Small area estimation, edition=2, author=J.N.K Rao, year=2015, publisher=A John Wiley and sons, inc, address=Canada, chapter=1.2

[5] @inbookSAE:ch1.4, title=Small area estimation, edition=2, author=J.N.K Rao, year=2015, publisher=A John Wiley and sons, inc, address=Canada, chapter=1.4

[6] @articlejenkins2005survival, title=Survival analysis, author=Jenkins, Stephen P, journal=Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK, volume=42, pages=54–56, year=2005, publisher=Citeseer

[7] @articleworld2020infant, title=Infant mortality, author=World Health Organization and others, year=2020, publisher=OECD

[8] @articlepfeffermann2002small, title=Small area estimation-new developments and directions, author=Pfeffermann, Danny, journal=International Statistical Review, volume=70, number=1, pages=125–143, year=2002, publisher=Wiley Online Library

[9] @articlepfeffermann2013new, title=New important developments in small area estimation, author=Pfeffermann, Danny, year=2013

[10] @incollectiondavis2017some, title=Some principles of stratification, author=Davis, Kingsley and Moore, Wilbert E, booktitle=Kingsley Davis, pages=221–231, year=2017, publisher=Routledge

[11] @bookvalliant2000finite, title=Finite population sampling and inference: a prediction approach, author=Valliant, Richard and Dorfman, Alan H and Royall, Richard M, number=04; QA276. 6, V3., year=2000, publisher=John Wiley New York

[12] @miscraosmall, title=Small Area Estimation Methods, Applications and Practical Demonstration, author=Rao, JNK and Molina, Isabel

[13] @articlesmith1986comparison, title=A comparison of direct and indirect methods for estimating environmental benefits, author=Smith, V Kerry and Desvousges, William H and Fisher, Ann, journal=American journal of Agricultural economics, volume=68, number=2, pages=280–290, year=1986, publisher=Wiley Online Library

[14] @articlebradburn2003survival, title=Survival analysis part II: multivariate data analysis–an introduction to concepts and methods, author=Bradburn, Mike J and Clark, Taane G and Love, Sharon B and Altman, Douglas Graham, journal=British journal of cancer, volume=89, number=3, pages=431–436, year=2003, publisher=Nature Publishing Group

[15] @articledabrowska1987non, title=Non-parametric regression with censored survival time data, author=Dabrowska, Dorota M, journal=Scandinavian Journal of Statistics, pages=181–197, year=1987, publisher=JSTOR

[16] @articleren2021design, title=Design-based small area estimation: an application to the DHS surveys, author=Ren, Ruilin, journal=SAE2021 BIG4small Book of short papers, pages=99, year=2021