Journal of Progressive Agriculture, Vol. 14 No.: 1 April, 2023

A REVIEW ON THE PROBLEM OF INCOMPLETENESS IN SAMPLE SURVEYS

Shilpa Yadav[1] and Pankaj Nagar[2]
[1]Assistant Professor, Department of Statistics, University of Rajasthan, Jaipur
[2]Professor, Department of Statistics, University of Rajasthan, Jaipur
Email: shilpa.y2108@gmail.com

Received: 15.12.22
Accepted: 24.02.23

ABSTRACT

Conventional survey research procedures rely on Neyman's probability-based sampling paradigm, having a number of fundamental assumptions that are increasingly getting very difficult to achieve in today's survey research environment. One such condition being, an equal and fair chance of every unit of the population to be included in the sample. In most of the studies, complete coverage of the population under study is practically impossible for a variety of reasons. On one hand, where it is difficult to have a perfect sampling frame that is independent of duplication as well as has covered the entire population in the frame, it is also not very economic to construct a fresh sampling frame, every time a study is being conducted. On the other hand, even in the rarest of the cases, if we have succeeded to have a perfect sampling frame, the problem of non-response further creates a void in the study. When cost is another factor that is taken into consideration, it's no surprise that alternative sampling methods are becoming more popular. The authors have taken into consideration, two very common sampling errors, viz., incomplete frame problem and non-response. This study is an attempt to examine and review, a number of past and current practices for drawing conclusions from samples that do not fully adhere to the statistical machinery that is currently available.

***Key words*: Incompleteness, Incomplete sampling frame, Imperfect frames, Sampling error.**

Sample surveys provide us with a powerful statistical tool to gain knowledge about a population and its features. Surveys prove to be the foundations of any type quantitative or qualitative research, be it a one-time study or longitudinal research. But with the fruits, come along the spines. There are a lot many numbers of issues that hinder the process of research at the survey stage and thus hamper the entire study, incompleteness being the most evident and challenging of all. Incompleteness in the sample surveys can arise due to numerous factors and at different stages of survey process like omission of important units from the frame itself, ambiguous or unnecessary information collected due to inexperienced investigators, incomplete mailed questionnaires, misprinting or experimental errors causing void in the survey process and many others. Amongst these humongous problems resulting into incompleteness in the survey, one major problem is the problem of incomplete frame, which despite being a very common issue, remains uncovered to a larger extent.

One of the most crucial and key aspects of sample surveys is a well-defined sampling frame that captures the entire target population in a statistical sense. A simple definition of a sampling frame is the set of the source material from which units are selected. The sampling frame has substantial impacts on the cost and the quality of the survey, be it household surveys or otherwise. Therefore, it is of pivotal importance to have a sampling frame that is complete, accurate and up-to-date, a frame in which every element appears on the list separately, once, only once and nothing else appears on the list. Nevertheless, it is essential to strive for them either in constructing a frame from scratch or using one that already exists. The degree to which there is failure to achieve each of the ideal properties produces survey results that are biased in various ways, but often in the direction of under-estimating the target population.

For decades, the traditional methods of probability-based sampling have served as the gold standard for survey research applications. Relying on the statistical machinery developed by Neyman (1934), it has been possible to make measurable inferences about target populations, when sampling units carry known selection probabilities and samples are selected from a well-defined sampling frame. Thus, for application of a probability sampling scheme, i.e., a sampling scheme in which every member of the target population has a known, non-zero probability of being included in the sample, it is cardinal to have a frame that is perfect and clearly defines the target population. However, the survey research industry is currently in a state of flux due to formidable challenges that question the external validity of the statistical machinery we have relied on for decades to develop determinate inferences for population parameters using probability-based samples. Top among such challenges is the problem of imperfect frames. Imperfection in sampling frames may arise due to four primary factors as discussed by Kish (1965). The omission of sampling units that should have been a part of the population, but are not included, leads to the first and most common problem, i.e., the problem of non-coverage or incompleteness in sampling frames, which is the most frequently occurs in important practical problems and contradicts the rule that every element must occur in the frame. Even if the surveyor succeeds in preparing an accurate and complete frame that covers the entire population, by the time the survey starts, it becomes outdated and leads to an incomplete sampling frame problem. The second factor that leads to imperfection is clustering of elements or sampling units, where a group of elements appear together, contradicting the rule that every element should be listed separately. The problem that mostly concerns the cluster sampling. Blank or foreign elements occur in many frames when some listings contain no elements of the target population, contradicting the rule that list contain no element which does not belong to the target population, thus giving rise to empty units generating biased and misleading results. The last factor that gives rise to the problem of imperfect frames being duplicate listings, that gives each sampling unit an unequal and unfair chance of being included in the sample.

Non-Coverage includes the problem of “incomplete frames”, a term that seems to imply omissions in preparing the frame. But it also refers to imply “missed units”, omissions due to faulty execution of survey procedures. For example, in demographic surveys to study infant mortality rates, we obtain the list of infants born and died in the past year from hospital records. No matter how judiciously the frame is prepared, the actual data is still under-covered as many of the infant births and deaths would still not be recorded or registered in any hospital records. In medical surveys, there is rarely a situation when one can obtain a complete list of all the patients suffering from a certain disease desired for any study due to under-registration and incomplete patient information. Incomplete frame problems are the most common when the study collects data for factors that are rather sensitive such as HIV, Sexual crimes, drug use, gambling etc. or related to factors that are considered illegit or illegal such as domestic violence, sexual abuse etc. where people tend to hide their information and thus are excluded from the sampling frames while construction. In rural areas, where there is lack of medical facilities, many of the deaths are not properly diagnosed and thus, never reported. Even in household surveys, real-time immigration and migration make it difficult to prepare a perfect sampling frame which is accurate and up to date, especially in case of a dynamic population. In online surveys, where email lists or phone number lists are used as sampling frames, error may accumulate due to the construction of the lists. In such cases, a particular segment of the population whose email addresses or phone numbers are unavailable may never included in the study, giving rise to sampling frame error.

The incomplete information obtained using an imperfect frame or non-response makes the full survey and its results ambiguous and misleading. Therefore, it is of vital importance to develop new techniques to deal with such situations. On one hand, though it may seem, the problem of imperfect or incomplete frames is not new, but still it is a wide area where extensive research has to be done. Several research has been done on dealing the problem of incomplete sampling frames but it always seems inadequate, as the variety of situations are there to be considered and there are a large number of sample design in which the estimators are to be adjusted for under-coverage or incompleteness. On the contrary, there is extensive research being done in the field of non-response, but there are so many facets which still need deeper exploration.

REVIEW OF LITERATURE

The problems of incompleteness in sample surveys have been discussed by many authors and have been dealt with varying tools, a few have been reviewed here.

Hansen et. al. (1946) discussed about the problem of non-response in mail questionnaire methods used for data collection. They proposed a technique to combine the techniques of mail questionnaires and personal interviews to get joint advantages of both procedures. The principle followed was to send questionnaires to a group of individuals from the target population, the size been larger than the expected number of questionnaires that would be returned and following it by personal interviews for a sample of non-respondents. Following the process, they suggested a composite estimator to estimate the population characteristic, one part of which comes from the respondent group and another from the non-respondents and derived its variance. They also used an illustration to show for a given degree of reliability, the sample sizes of the mailing list for various rates of response and the size for which personal interviews would be undergone to achieve a desired level of precision with the given cost constraints. The optimum values were obtained under different sampling structures.

Yates (1948), in his first edition of the book, described an early account of the principal weaknesses of the frames. He pointed out that the frames are often imperfect, that they sometimes have omissions of important units and sometimes consist of superfluous units, that they have an unknown amount of duplication and sometimes the supplementary information presented for stratification or unequal probability selection may be in error. In his book, he gave some sound practical advices for handling the imperfections, in brief.

Goodman (1952) considered the problem of duplication in the frame by matching a number of lists of names (k). He presented unbiased estimators of the number of names occurring in common between the k lists and the number of names occurring in the k lists. He also used unbiased and an insufficient statistic to obtain minimum variance unbiased estimators. He suggested selection of a sample of $n_i = N_i/g_i$ names from N_i names in the lists (i=1, 2, …, k) and gave the unbiased estimator $d^1_{[t]} = \prod_i \quad g_i e_{[t]}$, where the product was over all the values of i appearing in [t], some subset containing at least two of the integers 1,2, …, k, $d^1_{[t]}$ gave the number of names occurring together in lists [t]. He also suggested an unbiased estimator $\sum_{i=0}^{k-v} \quad (-1)^i C_v^{v+i} \, d^1(v+i)$ for the number of names occurring in v lists, where $d^1(v+i) = \sum d^1_{[t]}$ and another unbiased estimator $\sum_{i=0}^{k-v} \quad (-1)^i C_{v-1}^{v+i-1} \, d^1(v+i)$ for the number of names present in at least v lists.

El-Badry (1956) devised a plan that utilizes the information available to the surveyor to gain maximum efficiency, the estimates which are free from the non-response bias to the maximum extent. He suggested a series of mail questionnaires to be sent out to the target population, the non-respondents being chosen at each stage for further mailing of questionnaires followed by a personal interview to the most stubborn group of non-respondents in the end. He proposed stratification of the population on the basis of responses obtained at each stage before the field interviews were conducted. Combining the results obtained, he constructed an estimator of population mean,

$\underline{X} = c\sum_{h=1}^{n_{11}} \quad X_{11}^h + d \sum_{h=1}^{n_{21}} \quad X_{21}^h + \cdots + g\sum_{h=1}^{n_{m2}} \quad X_{(m+1)1}^h, c, d, \ldots, g$ □□□□□ □□□□□□□□□, $X_{i1}^h (i = 1, 2, \ldots, m+1)$ □□□□□□□ □h□ h□h□□□□□□□□□□□ □□ □h□ □□h□□□□□□□ and obtained the estimates of the constants involved such that the proposed estimator is unbiased as $c = \frac{1}{n_1}, d = \frac{1}{n_1} \cdot \frac{1}{k_2}, e = \frac{1}{n_1} \cdot \frac{1}{k_2 k_3}, \ldots, g = \frac{1}{n_1} \cdot \frac{1}{w \prod_2^m \quad k_i}$.

He also derived the variance of the same as

$$Var\ \underline{X} = \frac{1}{n_1}\left\{\frac{N - n_1}{N}\sigma^2 + \left(\frac{1}{k_2} - 1\right)P_{12}\sigma_{12}^2 + \frac{1}{k_2}\left(\frac{1}{k_3} - 1\right)P_{22}\sigma_{22}^2 + \cdots + \frac{1}{\prod \quad k_i}\left(\frac{1}{w} - 1\right)P_{m2}\sigma_{m2}^2\right\}$$

He further obtained the optimum sample sizes for successive mailing of questionnaires and conducting personal interviews given fixed relative variance, using a suitable cost function thus minimising the total cost of the survey.

Deming et. al. (1959) took into consideration, the problem of duplication and proposed a method of matching the lists, when two or more lists are available to the surveyor for use. They proposed the theory of estimation of the proportions of names, common between any two lists of names, using the sample drawn from lists. They gave the probability distributions, the expected values, variances as well as the third and fourth order moments of the estimates of the proportions duplicated. They also covered testing of hypothesis w.r.t. proportions, optimum allocation of samples, the effect of duplication within a list and possible gains on stratification. Deming et. al. Gave the probability distribution of 'd', where d denotes the number of names that are common between 2 lists or samples, as

$$P(d) = \frac{(D\ d)}{(M\ m)(N\ n)} \sum_{k=d}^{D} \quad (D - d\ k - d)(M - D\ m - k)(N - k\ n - d)$$

where, M, N are the number of names in the two lists, m and n are the number of names from list 1 and 2 in the sample out of which d are common between the two.

Under the two limiting cases, viz., Case 1, when M, N, m, n all increase limitless such that D, m/M, n/N remains constant and Case 2, when M, N, m, n, D, all increase limitless such that mnD/MN remains constant at some value λ, the pdf of d are given as

$$P(d) \to (D\ d)\left(\frac{mn}{MN}\right)^{d}\left(1-\frac{mn}{MN}\right)^{D-d}$$

$$P(d) \to \frac{\lambda^{d}}{d!}\, e^{-\lambda}$$

They derived the variance of the proportion $\hat{p} = \frac{N}{n}\frac{d}{m}$, tending to $\frac{Np}{mn}$. Further, they developed the cases where one of the list is complete and hence obtained the estimated number of duplications and the variance of the proposed estimator. They used suitable examples to support the theory developed.

Seal (1962) took into consideration, the problem of incompleteness in sampling frames and discussed the use of out-dated frames in large scale surveys. He mentioned that the statistic based on the sample drawn on the basis of such frames provides biased and vague estimates of the population characteristic, existing during the time period of the survey. In his paper, Seal suggested some simple methods which could be easily worked out in most of the practical situations on the basis of a reasonable birth and death process. He considered a large scale survey which is launched at some time ‘k’ and is completed at time ‘m’ (s.t. m>k) on a certain set of establishments, chosen randomly. During the survey, data is observed at time ‘t’ for a reference period $t \in [a,b]$, where $a < b \leq k$. He used the example of Labour Bureau’s publications released under “Large Industrial Establishments in India” to illustrate his study. He further developed the study based on the concepts of Stochastic processes, considering the change in the population as a stochastic process arising by births (addition to the population) and death (elimination from population) and hence gave the expected number of establishments M_t at time t, as

$$M_t = \frac{\lambda}{\mu}\left(1-e^{-\mu t}\right) + M_0 e^{-\mu t}$$

Unbiased estimates of μ and λ were obtained as

$$\tilde{\mu} = \frac{1}{p}\left(N_0 - N_{0p}\right)$$

$$\tilde{\lambda} = \tilde{\mu}\,\frac{N_0\left(N_p - N_{0p}\right)}{N_0 - N_{0p}}$$

N_0 □□□ N_{0p} being number of establishments in the existing frame at time 0 and common at time 0 and p respectively. However, assuming some reasonable model, it might be possible to derive the number of births and deaths of units that could be expected based on the data available for two consecutive frames. He described the problem of estimating the total of any characteristic of a dynamic population on the basis of such model. Seal also gave some illustrations in support of the methods suggested in the paper. With his findings, Seal also explained the various limitations for the proposed methods.

Hartley (1962) proposed the use of multiple frames to overcome the problem of incomplete frames in a manner that whole population is covered by the combined use of these frames. Hartley showed that dual-frame or multiple frame surveys can cost far less than a single frame survey and still can achieve the same level of precision. His applications focused on situations, where a single frame despite being complete is expensive to sample, on the other hand, a sample can be drawn economically from multiple frames which are originally incomplete. He suggested a weighted estimator for determining population characteristic using estimates obtained from different frames. He also derived the optimum sample sizes to be taken from different frames using suitable cost functions.

Hansen et. al. (1963) took into consideration the various potential imperfections that sampling frames can suffer from. They advised various procedures for probability sampling in case of incomplete frame problems. They proposed the Predecessor-Successor method to obtain information on the omissions or missing values. They considered a population divided into two groups, one that consists of all the units included in the list and the other being the group of non-included units. They supposed establishment of the principle viz. geographic ordering of the units in the population such that given any one unit, one can uniquely determine its successor by following a definite trail of travel. The procedure given by Hansen et. al. consists in first picking a random sample from the specified frame and then determining successor of each selected unit and checking if it is listed in the frame. If the successor is found on the list, we reject it else we proceed to check the availability of the next successor in the frame. The process is continued till we reach to the unit listed in the frame. All the units that were not listed are then included in the sample. Thus, the final sample will consist of the original sample as well as a sample that was not listed originally but follow the selected units on the trail of travel. The probability of selection of the non-included units is same as that of the unit preceding the listed unit. Thus, the sample size, here, becomes a random variable and thus workload is increased. They also proposed sampling from decentralized lists for dealing with incompleteness of the frame.

Singh (1983) developed mathematical formulation of the Predecessor-Successor method suggested by Hansen et. al. for estimating the total number of units of the target population which were not included in the frame and also the population aggregate of the characteristic under study for the target population. Singh suggested separate estimators for two different situations viz. when the missing units are random and when they differ significantly from the units already listed in the frame. He proposed the use of the unbiased estimator $\hat{Y} = N\left(1 + \underline{m}\right)\underline{y}_n$ for estimation from an incomplete frame, when the units not included in the frame are more or less having the same properties as that of the included ones. The variance of the estimator proposed is given as

$$V(\hat{Y}) = \frac{N(N-n)}{n}\left[\underline{y}_n^2 s_m^2 + \left(1+\underline{m}\right)^2 s_y^2 - 2\left(\frac{1}{n}-\frac{1}{N}\right) s_y^2\, s_m^2\right]$$

For the second case, he considered the estimation procedure when the units missing in the existing frame differ from the ones included in the frame. For this, he proposed the unbiased estimator $\hat{Y} = N\left(\underline{y}_n + \underline{m}\,\underline{y}'_m\right)$, $\underline{y}_n$ and $\underline{y}_m$ being the sample mean based on 'n' included units and 'm' units observed missing in the frame. The variance of the estimator proposed in the second case is given as

$$V(\hat{Y}) = N^2\left(\frac{s_y^2}{n} + \frac{s_y'^{\,2}}{n_1}\,\underline{m}^2 + \frac{s_m^2}{n}\,\underline{y}_m'^{\,2} - 2\frac{s_y^2 s_y'^{\,2}}{nm}\right)$$

the prime denoting the corresponding values of the missing units.

Singh (1989), further developed another theory for sampling from an incomplete sampling frame consisting of some missing units and some alien elements which do not belong to the target population. Using the predecessor-successor method and two situations viz. when the units are random and when they are different from the included units, he developed separate estimators. He considered the distribution of units actually belonging to the population (n_1) to be hypergeometric distribution and hence developed the unbiased estimator of N_1, number of units in the population belonging to the population out of N in the frame $\hat{N}_1 = \frac{n_1}{n} N$, such that

$$E(n_1) = np \text{ and } V(n_1) = \frac{npq(N-n)}{N-1}$$

Unbiased estimator for the total for the character under study, when missing units are random, was given as $\phi = N_1\underline{y} + \underline{m}N_1\underline{y} = T_1 + \underline{m}T_1$, and derived the variance of T_1 as

$$V(T_1) = \frac{N(N-n)}{n(N-1)}\left[(Np-1)S_1^2 + NpqY_1^2\right]$$

$$where, E(\underline{m}) = \underline{M}, V(\underline{m}) = \left(\frac{np+q}{(np)^2} - \frac{1}{Np}\right)S_m^2$$

p being the proportion of units belonging to the target population and q=1-p. Thus obtained the estimate of variance of ϕ as

$$\hat{V}(\phi) = \hat{V}(T_1) + \hat{V}(\underline{m}T_1) + 2\,\widehat{Cov}(T_1, \underline{m}T_1)$$

Under the second case, he proposed another estimator $\phi' = \hat{N}_1\underline{y} + \hat{M}e_1 = T_1 + \hat{M}e_1, e_1$ being the cluster mean of the missing units occurring together between two consecutive included units and obtained its variance as

$$\hat{V}(\phi') = V(T_1) + V(\hat{M})V(e_1) + V(\hat{M})[E(e_1)]^2 + V(e_1)\left[E(\hat{M})\right]^2 + 2E(\hat{M})Cov(T_1, e_1)$$

$$where, Cov\,(T_1, e_1) = \frac{N_1}{N_1-1}\sum_{i=1}^{N_1}\left(y_i - \underline{y}_i\right)\left(\left(\frac{m_i}{\underline{M}}\right)\underline{y}_i' - \underline{Y}_2\right)$$

He gave another estimator considering the group of one included and other missing units as to form a cluster and hence developed another estimator $\phi'' = \hat{N}_1(1+\underline{m}')e_1', e_1'$ being the mean of the cluster so formed. He further obtained the variance of the estimator as

$$V(\phi'') = V(T')\left[(1+\underline{M}')^2 + V(\underline{m}')\right] + \underline{Y}^2 V(\underline{m}')$$

Agarwal et. al. (2008) developed the estimates of population total, mean and its variance and also derived the formulae for estimators in case of simple random sampling technique under with and without replacement schemes. They, in addition to the estimator developed by Singh (1983), they constructed another estimator for estimation in case of simple random sampling without replacement when non-included units are different from the included ones, as $\underline{y}_w = \frac{1}{n}\left(n_1\underline{y}_1 + n_2\underline{y}_2'\right)$, $\underline{y}_1$ being the mean of n_1 units, coming from the existing frame and $\underline{y}_2'$ being the mean of a sample of n_2' units taken from the n_2 units originally missing from the frame and traced using predecessor-successor method. They further obtained the variance of the proposed estimator

$$V\left(\underline{y}_w\right) = \frac{(1-f)S^2}{n_1} + \frac{h-1}{n}W_2 S_2^2\,, W_2 = \frac{K}{N'}$$

K being the number of non-included units in the frame and N' being the aggregate of included and non-included units in the population, f and h were used for sampling fraction and retainment factor respectively. They obtained the results for the same in case of with replacement sampling scheme too. They substantiated the theory with the help of a numerical problem.

Kang (2013), in his review article has given a detailed appraisal on the problem of missing data, the mechanisms by which missing data occurs and the different methods for dealing with missing data.

Rueda et. al (2018) combined sensitive research and multiple frame surveys. In particular, they considered statistical techniques for dealing with sensitive data coming from multiple frame surveys using intricate sampling designs. They estimated the mean of a sensitive variable associated with disagreeable behaviours, when the data are obtained using the randomized response concept. They constructed estimators and investigated their properties theoretically. Along with that they used Jackknife technique to estimate variance. The theory was then supported with Monte-Carlo simulation to assess the performance of the proposed estimator and its accuracy.

Gupta et. al. (2019) proposed a weighted product estimator for estimating population characteristics in case of an incomplete sampling frame. The proposed estimator is a combined linear estimator based on samples drawn from the known and listed population and also on the new information gathered for the unlisted or non-included units. The estimator used for the mean of known population is a product estimator and the mean for the non-included population is obtained through simple random sampling scheme. They also proved the unbiasedness of the proposed estimator, derived its mean square error up to first approximation and established the efficiency of the estimator with respect to the one given by Agarwal (2008). The study was supported with the help of numerical illustration. Similar estimators were developed for ratio and regression methods of estimation by Joshi et. al. (2021) and Gupta et. al. (2021) and their relative efficiencies w.r.t. the estimator proposed by Agarwal (2008) was also given.

CONCLUSION

The problem of incomplete sampling frames, despite being as old as many other concepts and issues in the field of sample surveys, is still being neglected and often confused with non-response and missing problems. The work done so far in the field of incomplete sampling frame still lacks a lot many facets that are still under the wraps and require deep research to unfold the hidden solutions to the problem of the same. The work done can be extended further for more complex sampling designs and techniques to obtain better estimates of the population characteristics.

REFERENCES

Agarwal, B., & Gupta, P. C. (2008). Estimation from Incomplete Sampling Frames in Case of Simple Random Sampling. *Model Assisted Statistics and Applications*.

Deming, W. E., & Glasser, G. J. (1959). On the Problem of Matching Lists by Samples. *Journal of the American Statistical Association*, 403-415.

El-Badry, M. A. (1956). A Sampling Procedure for Mailed Questionnaires. *American Statistical Association*, 209-227.

Goodman, L. A. (1952). On the Analysis of Samples from k Lists. *The Annals of Mathematical Statistics*, 632-634.

Gupta, P. C., Joshi, L., Joshi, V., & Nagar, P. (2019). Weighted Product Estimator for Incomplete Sampling Frame. *Journal of Rajasthan Academy of Physical Sciences*, 233-240.

Gupta, P. C., Joshi, V., Nagar, P., & Singh, A. K. (2021). Linear Regression Estimator in Case of Incomplete Sampling Frame. *Journal of Rajasthan Academy of Physical Sciences*, 49-56.

Hansen, M. H., & Hurwitz, W. N. (1946). The Problem of Non-Response in Sample Surveys. *Journal of the American Statistical Association*, 517-529.

Hansen, M. H., Hurwitz, W. N., & Jabine, T. B. (1963). The Use of Imperfect Lists for Probability Sampling. *U.S. Bureau of*

Census (pp. 497-517). Canada: Int. Stat. Inst. Ottawa.

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample Survey Methods and Theory*. New York: John Wiley and Sons.

Hartley, H. O. (1962). Multiple Frame Surveys. *Social Statistics Section America Statistical Association*, (pp. 203-206). Minnesota.

Joshi, V., Nagar, P., Singh, A. K., & Gupta, P. C. (2021). Use of Ratio Method of Estimation in Incomplete Frames. *International Journal of Agricultural and Statistical Sciences*, 267-272.

Kang, H. (2013). The Prevention and Handling of the Missing Data. *Korean J Anesthesiol*, 402-406.

Kish, L. (1965). *Survey Sampling*. John Wiley and Sons Inc., New York.

Neyman, J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of thw Royal Statistical Society*, 558-625.

Rueda, M., Cobo, B., & Perri, P. F. (2018). Randomized Response Estimation in Multiple Frame Surveys. *International Journal of Computer Mathematics*.

Seal, K. C. (1962). Use of Out-Dated Frames in Large Scale Sample Surveys. *Calcutta Statistical Association Bulletin*, 68-84.

Singh, R. (1983). On the Use of Incomplete Frame in Sample Survey. *Biometrical Journal*, 545-549.

Singh, R. (1989). Method of Estimation for Sampling from Incomplete Frames. *Australian Journal of Statistics*, 269-276.

Yates, F. (1948). *Sampling Methods for Census and Surveys* (First Edition ed.). London: Charles Griffin and Co.