Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 69(2) 2015 161-167

Some Contributions to Successive Sampling: Review

Nazeema Beevi and C. Chandran
Department of Statistics, University of Calicut, Calicut

Received 21 October 2013; Revised 08 April 2015; Accepted 01 June 2015

SUMMARY

In this paper reviewed some work of successive sampling by various authors. The work done so far in successive sampling is known technique that can be used in regression, ratio, super-population model and longitudinal surveys to estimate population parameters and measurements of difference or change of a study variable. Some regression type estimators of population mean have been proposed for successive sampling using information of two or more auxiliary variables. This class includes a number of regression estimators (Sen 1971, 1972 and 1973) and also the class of estimators suggested by Singh *et al.* (2008, 2009, 2010, and 2012).

Keywords: Auxiliary variable, Longitudinal surveys, Ratio estimator, Regression estimator, Successive sampling, Super-population model.

1. INTRODUCTION

In continuing sample surveys conducted at regular intervals (*e.g.* quarterly) for investing the time-dependent characteristics of certain dynamic populations, it is frequently advantageous to use the so-called rotation sampling technique, whereby a scheme of partial replacement of sampling units is developed in such a way that the sampling units to be used will be in the sample consecutively for some finite number of occasions, then they will be replaced by newly selected units. The replacement is done only to a portion of the sample while the other portion is retained for the next occasion. To such plans of sampling were attached different names by various authors, such as sampling on succession occasions with partial replacement of units, rotation sampling, or rotation designs for sampling on repeated occasions or sampling for a time series, or successive sampling. Several interesting problems could fall in this branch of study; some of the examples may be quoted in the different fields of study:

- **Agricultural Field:** To know the agricultural production at different points of time and to know the pattern of variation in agricultural production over period of time.
- **Scientific Field:** To know the growth of telecommunications at different points of time and the pattern of change over given period of time.
- **Demographic Field:** To know the birth rate, death rate, infant mortality rate at different points of time and the pattern of changes in the parameters over given period of time.

Sometime in problems on sampling in studying one or more characteristics of a particular population it becomes desirable to take samples at certain interval (not necessarily equal) from the population and give the estimates of the characteristics under study of the population on each occasion after making use of the information on previous information. If sampling is

Corresponding author: Nazeema Beevi
E-mail address: nazeemathaj@gmail.com

done on more than one occasion then it is known as Successive Sampling. Generally in successive sampling, if the main objective is to estimate the change from period to period, then it is found that it is best to retain the same population on each occasion, to get most precise estimators. But if it is desired to estimate the mean on each occasion to the next, it may be best to retain part of the sample and draw the remainder of the sample afresh.

Change is an inherent behaviour of the nature. Some types of change directly or indirectly affect the quality of living of the human beings. Such changes at different occasions of time at any given occasion of time. Many problems could fall in this branch of study, which call as successive occasions. An important aspect of successive sampling/continuous survey is the structure of the sample on each occasion. The structure of the sample in the following three different possibilities:

i. The same sample may be used on each occasion (Panel Sampling/Fixed Sample).

ii. A new sample may be taken on each occasion (Repeated Sampling/Independent Sample).

iii. A part of the sample may be retained while the remainder of the sample may be drawn afresh (Sampling on Successive Occasions/Partial Replacement).

(i) Panel Sampling/Fixed Sample

In this method, first selecting a group of participants through a random sampling method and then asking that group for information several times over a period of time. Therefore, each participant is interviewed at two or more time points; each period of data collection is called a "wave". This longitudinal sampling-method allows estimates of changes in the population, for example with regard to chronic illness to job stress to weekly food expenditures. Panel sampling can also be used to inform researchers about within-person health changes due to age or to help explain changes in continuous dependent variables such as spousal interaction.

The essential feature of a panel survey is that it provides observations on a set of variables for the same sample units over time. In the panel survey, repeated observations are derived by following a sample of persons (a panel) over time and by collecting data from a sequence of interviews. These interviews are conducted at usually fixed occasions that in most cases are regularly spaced.

The main objective of Panel Sampling is to estimate the change with a view to study effect of the forces acting upon the population, it is better to retain the sample from occasion to occasion. This sampling involves attempting to collect survey data from the same units on multiple occasions. After the initial sample selection, no addition to the sample is made

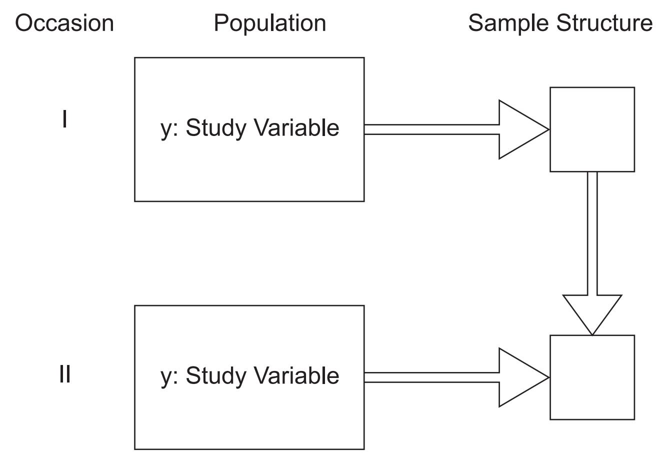


Fig. 1. Sample structures for panel sampling

(ii) Repeated Sampling/Independent Sample

A successive survey means survey on repeated occasions. Many studies have used survey method: example sociological, economical, medical etc. The objective for these are to estimate characteristics of a population on repeated occasions in order to measure time-trend as well as current values of the characteristics (see, Rao and Graham 1964). A partial replacement of units from occasion to occasion is known as repeated sampling (see, Manoussakis 1977).

The idea behind simple random sampling is that the value of sample characteristics is approximately the same as the corresponding values in the population. If we use a fixed sample for a survey then the accuracy of the measurements in this sample will decrease over time due to population changes. It is believed that a repeated sample is better than a fixed sample. In repeated sampling, we need take into account the correlation between current occasion h^{th} and previous occasion $(h-1)^{th}$. We use correlation coefficient ρ to denote this relationship. When $\rho = 1$, the previous sample is not good for estimating of current population characteristics. Contrarily, if $\rho = 0$, that means data from the current occasion has no relationship with the previous occasion. In repeated sampling, sample size and population size are the same at all occasions. But the sample is not the same from occasion to occasion. The correlation between the two occasions will have it will also have an effect on the sample characteristics for the current occasion.

The basic objective of Repeated Sampling is to study the overall average or total, it is better to select a fresh sample each occasion.

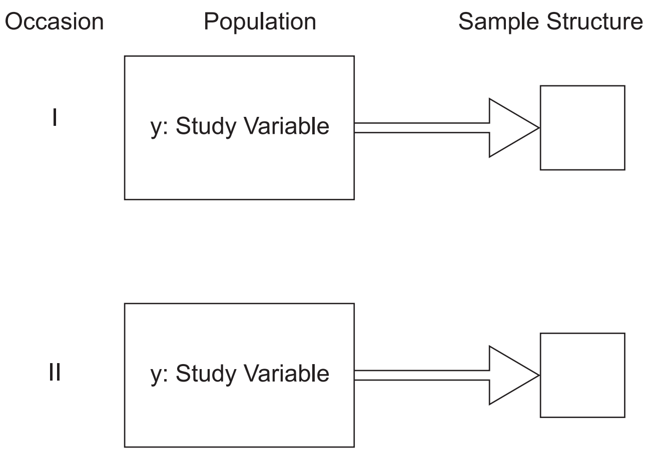
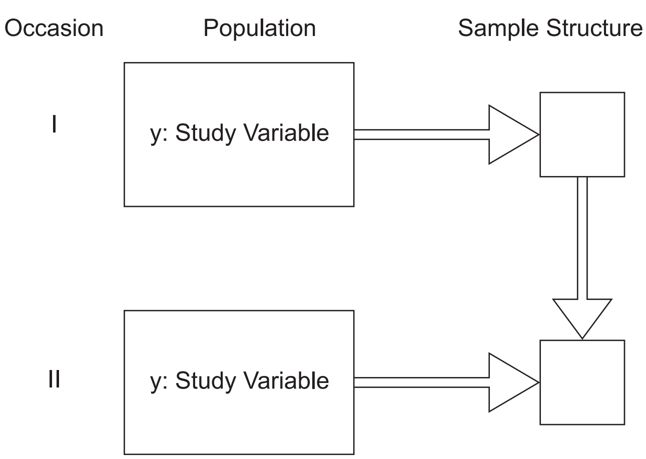


Fig. 2. Sample structures for repeated sampling

(i) Successive Sampling/Partial Replacement

In successive sampling, the objective is to estimate the average value for the most recent (current) occasion, the retention of a part of the sample over occasions provides efficient estimates as compared to other alternatives (see Singh and Chaudhary 1986). In successive sampling, if the characteristic of the population are likely to change rapidly with time. Hence partial replacement of the sample on each occasion leads to a better result.



The main advantages of a technique of sampling where partial replacement of units is part of the overall sampling design over one where there is no partial replacement of units (*i.e.*, taking a new set of units, or using the same set of units every time). Partial replacement units in the sample spread the burden of reporting among more respondents and hence result in better co-operation from respondents. This is very important from the standpoint of maintaining the rate of response when a human is studied.

Experiences from many censuses or survey studies (by complete enumeration or sampling methods) seem to indicate that the respondents tend to become uncooperative during the third or fourth visit if the same sampling units are used. Even with full cooperation, the respondents may be unwilling to give the same type of information time after time, or they may be influenced by the information which they give and receive at earlier interviews and this may make them progressively less representative as time proceeds.

Partial replacement of units in the sample permits the use of data from past samples to improve the current estimate of population characteristics of interest. This can be accomplished by some appropriate methods of estimation which take advantage of past as well as present information to provide an estimate for the present occasion. This theoretical advantage is perhaps the most important reason for using partial replacement units technique when we have to deal with time-series characteristics.

In many social surveys, the same population is sampled repeatedly and the same study variable is measured at each occasion, so that development over time can be followed. For example, labour force surveys are conducted monthly to estimate the number of employed, monthly surveys in which data on price of goods are collected to determine a consumer price index, political opinion surveys conducted at regular intervals to measure voter preferences, etc. In this case, the use of a successive sampling scheme can be an attractive alternative to improve the estimates of level at a point in time or change between two time points. (see, *e.g.*, Cochran 1977).

The first attempt to solve this problem was made by Jessen (1942) as early as 1939 in conducting a survey in which he used the information obtained in 1938. He propounded the idea of partial replacement of units drawn on the first occasion for getting an improvement estimator of the population mean on the second occasion. In fact, he combined two independent estimators of the population mean on the second occasion a double sampling regression estimator based on the units common to both the occasions and a sample mean of new units, to provide an improved estimator to the population mean on the second occasion. Yates (1949) has also stated some of the results on sampling on two or more occasions. The general theory of successive sampling has been studied for one character independently by Patterson (1950) and Tikkiwai (1956) both arriving at similar results. The effect of using matched sampling to estimate the time-conditioned characters of a population from occasion to occasion is to improve the efficiency of the various estimators, based on matched sampling, of the character with increasing occasions reaching a limiting value. This is in addition that it gives more efficient estimators of the various parameters on an occasion as compared to the corresponding simple estimators based on no- matching on that occasion.

Singh (1968), successive sampling procedures using multi-stage sampling design have been developed. It is found that it is generally advantageous to retain a fraction of the sample selected on previous occasions for improving the estimate of a mean on the most recent occasion. This type of partial replacement may sometimes be recommended for estimating a mean over all occasions. This type of partial replacement may sometimes be recommended for estimating a mean over all occasions. This is specially true if the experimenter is interested in not only obtaining an overall estimate for the entire period but also in separate estimates for each occasion. In a sampling enquiry repeated on three occasions, it has been observed that for estimating the mean on the third occasion it would be preferable to repeat the same sample fraction from one occasion to the next, while for estimating the mean over all occasions the sample fraction repeated on the second occasion. The problem of matching primary and secondary stage units in multi-stage sampling on successive occasions was examined by Singh and Kathuria (1969).

Sen (1971), the theory of successive sampling on two occasions for ratio and regression estimators has been developed by using information from the previous occasion on two auxiliary variables whose population means are unknown and Sen (1971), a more general but simpler approach to the theory for regression estimators using P ($P \geq 1$) auxiliary variables is presented. The extension of the theory to sampling on more than two occasions is straightforward.

The theory has been applied to waterfowl mail survey (Sen 1970) in Ontario, Canada. It was found that when the matched proportions of the sample in the second year were optimal for the variates, "kill per hunter" and "days hunted" the relative gain in precision of the estimate of "kill per hunter" using previous season's (i) "kill per hunter" and (ii) "days hunted" from the matched portion as auxiliary variables was 66 percent as compared with simple random sampling with no matching; when "kill per hunter "alone was used as supplementary information, the total gain was reduced to 35 percent.

Auxiliary information is often available in practical problems, so statisticians naturally make use of it in successive sampling (see, Sen 1972, 1973). Sen (1972), the theory developed so far aims at providing the optimum estimate by combining (i) a double-sampling regression estimate from the matched portion of the sample and (ii) a sample mean based on a random sample from the unmatched portion of the sample on the second occasion theory has been generalized in the present note by using a double- sampling multivariate ratio estimate using P auxiliary variates ($P \geq 1$) from the matched portion of the sample.

Sen (1971, 1972, and 1973) investigated successive sampling on two occasions using multi-auxiliary information. He considered all the auxiliary variables to be positively correlated with the study variable and hence used two-phase (double) sampling multivariate ratio estimator. Adhvaryu (1978), the case whom some auxiliary variables are positively correlated and remaining are negatively correlated with the study variables and hence used two-phase sampling multivariate ratio estimator. The more the information, the greater the precision of the estimator achieved. Hence in survey sampling use of auxiliary information has become a common practice, if it is available, especially without additional cost. Singh and Singh (2001) considered successive sampling using information on an auxiliary variable which is available for the current occasion (second occasion), estimator of population mean on second occasion based on the unmatched portion can be improved using information on the supplementary variable and estimator based on matched can be obtained using information on the auxiliary variate at the second occasion and also by using information on the study variate on the previous occasion. Thus the entire information available at the current occasion is made use of to get an estimator of population mean at the current occasion with highest precision.

Biradar and Singh (2001) proposed an estimator for the population mean on the second of two successive occasions utilizing information available on both the occasions on an auxiliary variate with an unknown population mean. Rueda *et al.* (2008) presents successive sampling to estimate quantiles with P-auxiliary variables. Successive sampling has been extensively used in applied and Social Sciences to estimate measures of level change of a linear parameter such as a mean or total and testing of the direction of this change. For example longitudinal surveys are used in the analysis of Social change (Ruspini 1999) and the study of occupational mobility (Solga 2001).

In successive sampling on two occasions, the theory developed so far (Jessen 1942, Patterson 1950) aimed at providing the optimum estimate of the population mean in the second occasion by combining two estimators of this mean; a double sampling regression estimate from the matched portion of the sample when the auxiliary variable is the value of the principle variable in the first occasion, and a simple mean based on a random sample from the unmatched portion on the second occasion. Successive Sampling has also been discussed in some detail by Narain (1953), Adhvaryu (1978), Eckler (1955), Gordon (1983), Arnab and Okafor (1992), Singh and Srivastrava (1973), Singh *et al.* (1992) and Singh (2003), which provides an extensive bibliography of this topic. In all the above studies, the parameter considered is the mean.

For the problem of estimating a population quantile the situation is quite different. The study of economic issues frequently involves variables with extreme values which strongly influence the value of the mean. In these situations, the mean may offer results which are not representative enough because it moves with the direction of the asymmetry. The median, on the other hand, is unaffected by extreme values. Quantiles are used by the national agencies and statistical offices to obtain the low-income proportion and other important measures such as the Gini coefficient or Lorenz-curves. Economics studies also use quantiles in analysis of the relationship between household food expenditure and household income, wages and salaries analysis, impact of behaviour on the birth weight of infants, labour markets discrimination, school quality, demand analysis etc.

Singh and Karna (2009) presents in successive sampling over two occasions with partial replacement of units at current (second) occasion, utilizing the information on an auxiliary character over both the occasion on study character, regression type estimators for estimating the population mean at current (second) occasion have been proposed. Behaviours of the proposed estimators have been studied. Proposed estimators have been compared with the sample mean estimator when there is no matching and the optimum estimator, which is a linear combination of the means of the matched and unmatched portion of the sample at the current (second) occasion. Optimum replacement policy is also discussed through this paper.

Singh and Priyanka (2010), in the present work, information on an auxiliary character, which is readily available on all the occasions, has been used along with the information on study character from the previous and current occasion. Consequently, chain type difference and regression estimators have been proposed for estimating the population mean at current (second) occasion in the two occasions successive sampling. The proposed estimators have been compared with sample mean estimator when there is no matching and the optimum estimator, which is the combination of the means of the matched and unmatched portions of the sample at the second occasion. Optimum replacement policy is also discussed.

In practice non-response is one of the major problems encountered by survey statisticians. Longitudinal surveys are more prone to this problem than single-occasion surveys. For example, for an agricultural production survey it might be possible that crop on certain plots is destroyed due to some natural calamities or disease. So that yield on these plots is impossible to be measured. Such non-response (incompleteness) can have recognized for some time that if the suitable information about the nature of non-response in the population is unknown, the inference concerning population parameters can be spoiled. Many methods are used to reduce the negative impact of non-response in sample survey. Imputation is one which deals with the filling up method of incomplete data for adapting the standard analytic model in Statistics. It is typically used when needed to substitute missing item values with certain fabricated values in the sample surveys. Singh and Horn (2000); suggested a compromised method of imputation. Singh and Priyanka (2007b) also discussed the treatment of non-response in successive sampling.

Motivated with above works Singh *et al.* (2010) have suggested impressive imputation methods to deal with the problems of non-response on the current occasion in two-occasion successive sampling. There may be situations where the non-response may occur on both the occasions in two-occasion successive sampling. Singh *et al.* (2012) intends to develop some imputation methods to reduce the impact of non-response at both the occasions in two-occasion successive sampling. Utilizing the auxiliary information, which is only available at the current occasion, estimators have been proposed for estimating the population mean at the current occasion. Estimators for the current-occasion are also derived as a particular case when there is non-response either on the first occasion or second occasion. Behaviours of the proposed estimators studied and their respective optimum replacement policies are also discussed. To study the effectiveness of the suggested imputation methods, performances of the proposed estimators are compared in two different situations, with and without non-response.

Estimation of population mean at current occasion in successive sampling under a super-population model present the work is an attempt to utilize auxiliary information through a super population linear model as well as in the structure of estimators for estimating the population mean on current occasion in two occasion successive sampling developed by Singh and Priyanka (2010).

Singh and Priyanka (2010), present the work intended emphasize the role of several varying auxiliary variates at both the occasions to improve the precision of estimates at current occasion in two occasion successive sampling. Two different efficient estimators are proposed estimators with the sample mean estimator when there is no matching from previous occasion and the optimum successive sampling estimator when no auxiliary information is used have been incorporated. Singh *et al*. (2012), an attempt to estimate population mean on the current occasion using two-phase successive sampling on two occasions has been made. Two-phase ratio, regression and regression type estimators for estimating population mean on second (current) occasion have been proposed. Optimum replacement policies of the proposed estimators have been discussed. The proposed estimators are compared with sample mean estimator when there is no matching and the optimum estimator which is a linear combination of the means of the matched and unmatched portions of the sample at the current occasion.

REFERENCES

Adhvaryu, D. (1978). Successive Sampling using multi-auxiliary information. *Sankhya,* **40**, 167-173.

Arnab, R. and Okfar, F.C. (1992). A note on double sampling over two occasions. *Pak. J. Statist.,* **8**, 9-18.

Biradar, R.S. and Singh, H.P. (2001). Successive sampling using auxiliary information on both occasions. *Bull. Cal. Statist. Assoc.*, **51**, 243-251.

Cochran, W.G. (1977). *Sampling Techniques,* 3rd ed., Wiley, New York.

Eckler, A.R. (1955). Rotation sampling. *The Ann. Math. Statist.,* **26**, 664-685.

Gordon (1983). Successive sampling in finite populations. *The Ann. Math. Statist.,* **11**, 702-706.

Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. In: *Iowa Agricultural Experiment Station Road Bulletin No. 304*, Ames, USA, 1-104.

Manoussakis, E. (1977). Repeated sampling with partial replacement of units. *The Ann. Stat*., **5(4),** 795-802.

Narain, R.D. (1953). On the recurrence formula in sampling on successive occasions. *J. Ind. Soc. Agril. Statist.,* **5**, 96-99.

Patterson, H.D. (1950). Sampling on successive occasions with partial replacement of units. *J. Roy. Statist. Soc.,* **12**, 241-255.

Rao, J.N.K. and Graham, J.E. (1964). Rotation designs for sampling on repeated occasions. *J. Amer. Statist. Assoc.*, **69,** 492-501.

Rueda, M., Munoz, J. and Arcos, A. (2008). Successive sampling to estimate quantiles with P auxiliary variables. *Quality Quantity,* **42**, 427-443.

Ruspinis, E. (1999). Longitudinal research and the analysis of social change. *Quality Quantity,* **33**, 219-227.

Sen, A.R. (1970). A pilot survey of the characteristics of waterfowl hunters in Ontario, 1968-69. *J. Amer. Statist. Assoc.,* **65**, 1039.

Sen, A.R. (1971). Successive sampling with two auxiliary variables. *Sankhya*, **33**, 371-378.

Sen, A.R. (1972). Successive sampling with P (P ≥ 1) auxiliary variables. *The Ann. Math. Statist.*, **43**, 2031-2034.

Sen, A.R. (1973). Theory and application of sampling on repeated occasions with several auxiliary variables. *Biometrics,* **29**, 381-385.

Singh, G.N. (2003). Estimation of population mean using auxiliary information on recent occasion in h occasions successive sampling. *Stat. Trans.*-new series, **6(4),** 523-532.

Singh, V.P. (1992). A generalized efficient class of estimators of population mean in two-phase successive sampling. *Inter. J. Manage. Sys.,* **8(2),** 173-183.

Singh, D. (1968). Estimates in successive sampling using a multi-stage design. *J. Amer. Statist. Assoc.,* **63(321)**, 99-112.

Singh, S. (2009). A New Method of Imputation in Survey Sampling. *Statistics,* **43(5)**, 499-511.

Singh, S. and Horn, S. (2000). Compromised imputation in survey sampling. *Metrika,* **51**, 267-276.

Singh, D. and Chaudhary. F.S. (1986). *Theory and Analysis of Sample Survey Designs.* New Age International (P) Ltd.

Singh, G.N., Jaishree Prabha Karna (2009). Estimation of population mean on current occasion in two-occasion successive sampling. *Metron*, **LXVII,** 87-103.

Singh, D. and Kathuria, O.P. (1969). On two-stage successive sampling. *Austr. J. Stat*., **11(2)**, 59-69.

Singh, G.N. and Priyanka, K. (2007b). Effect of non-response on current occasion in search of good rotation patterns on successive occasions. *Stat. Trans.*-new series, **8(2)**, 273-292.

Singh, G.N. and Priyanka, K. (2008). Search of good patterns to improve the precision of estimates at current occasion. *Comm. Statist.- Theory Methods*, **37(3)**, 337-348.

Singh, G.N. and Priyanka, K. (2009). Search of effective rotation patterns in presence of auxiliary information in successive sampling over two occasions. *Stat. Trans.*-new series, I, **10(1)**, 59-73.

Singh, G.N. and Priyanka, K. (2010). Use of imputation methods in two-occasion successive sampling. *J. Ind. Soc. Agril. Statist.*, **64(3)**, 417-432.

Singh, G.N. and Priyanka, K. (2010). Estimation of population mean at current occasion in successive sampling under a super-population model. *Model Assist. Statist. Appl.*, II, **4**, 189-200.

Singh, G.N. and Priyanka, K. (2010). Estimation of population mean at current occasion in presence of several varying auxiliary variates in two-occasion successive sampling. *Stat. Trans.*-new series, II, **1**, 105-126.

Singh, G.N. and Singh, V.K. (2001). On the use of auxiliary information in successive sampling. *J. Ind. Soc. Agril. Statist.*, **54(1),** 1-12.

Singh, S. and Srivastrava, A.K. (1973). Use of Auxiliary Information in Two Stage Successive Sampling. *J. Ind. Soc. Agril. Statist.,* **25**, 101-104.

Singh, G.N., Singh, V.K., Priyanka, K., Prasad Shakti and Karna Jaishree Prabha (2012). Rotation patterns under imputation of missing data over two-occasion. *Comm. Statist.- Theory Methods*, **41**, 1857-1874.

Singh, G.N., Majhi, D. and Prasad, S. (2012). Estimation of population mean in two-phase successive sampling. *Recent Adv. Inform., IEEE Xpore.*

Solga, H. (2001). Longitudinal surveys and the study of occupational mobility: Panel and retrospective design in comparison. *Quality Quantity,* **35**, 291-309.

Tikkiwal, B.D. (1956). A further contribution the theory of univariate sampling on successive occasions, *J. Ind. Soc. Agril. Statist.*, **8**, 84-90.

Yates, F. (1949). *Sampling Methods for Census and Survey*, Charles Griffin and Co.Ltd., London.