

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2026): 4.49
Maths 2026; 11(2): 113-117
© 2026 Stats & Maths
<https://www.mathsjournal.com>
Received: 13-12-2025
Accepted: 23-01-2026

Dr. Clara Hoffmann
Department of Survey Statistics,
Faculty of Social Data Sciences,
Alpine Institute of Statistical
Studies, Vienna, Austria

Dr. Lukas Steiner
Department of Survey Statistics,
Faculty of Social Data Sciences,
Alpine Institute of Statistical
Studies, Vienna, Austria

Corresponding Author:
Dr. Clara Hoffmann
Department of Survey Statistics,
Faculty of Social Data Sciences,
Alpine Institute of Statistical
Studies, Vienna, Austria

Difference-type and related estimators for population mean under two-phase sampling: A literature review

Clara Hoffmann and Lukas Steiner

DOI: <https://www.doi.org/10.22271/math.2026.v11.i2b.2615>

Abstract

When the population mean of an auxiliary variable is unknown, two-phase sampling offers a practical way to still use that variable to improve estimation of a study variable's mean: a large first-phase sample estimates the auxiliary mean, and a smaller second-phase subsample measures the study variable itself. This review traces the literature on difference-type and related estimators built on this design, from Neyman's 1938 foundational treatment of two-phase sampling and Kiregyera's early chain ratio-type and regression-type estimators, through Khan and Al-Hossain's 2016 difference-type estimator and its later modifications, to recent exponential-type and generalised classes. Particular attention is given to how each estimator's bias and mean squared error are derived to first order, how the optimum values of the estimator's constants are obtained, and how empirical efficiency comparisons are typically reported. Across ten studies spanning nearly nine decades, estimators for the population mean substantially outnumber those for the population variance, and the difference-type construction itself has been revisited and modified at least three times since 2014, suggesting continued, incremental refinement rather than a single settled form.

Keywords: Two-phase sampling, double sampling, difference-type estimator, auxiliary variable, bias, mean squared error, population mean, efficiency comparison, survey sampling

Introduction

Auxiliary information improves the precision of a survey estimator only if the auxiliary variable's population mean is known or can be estimated. When it is not known in advance, Neyman's 1938 paper showed that a two-phase, or double, sampling design solves the problem directly: draw a large first-phase sample to estimate the auxiliary mean, then draw a smaller second-phase subsample, generally from within the first-phase sample, to measure the study variable itself ^[1]. This two-stage structure, rather than any single estimator, is the foundation that every study reviewed here builds on.

Kiregyera's two papers in the early 1980s were among the first to exploit a second auxiliary variable within this design, proposing a chain ratio-type estimator that applies a ratio correction sequentially through two auxiliary variables, and later a regression-type estimator using the same two-variable structure ^[2, 3]. Khare, Srivastava, and Kumar generalised this chain idea further in 2013, combining a ratio correction with a regression correction in a single generalised chain estimator ^[4].

The specific difference-type construction that gives this review its focus, the sample mean adjusted by a linear multiple of the gap between an estimated and observed auxiliary mean, was formalised for the two-phase, two-auxiliary-variable setting by Khan and Al-Hossain in 2016, building directly on Khare, Srivastava, and Kumar's chain estimator ^[6]. This review traces how that difference-type construction and its close relatives have developed since, alongside the earlier chain-type literature it grew out of.

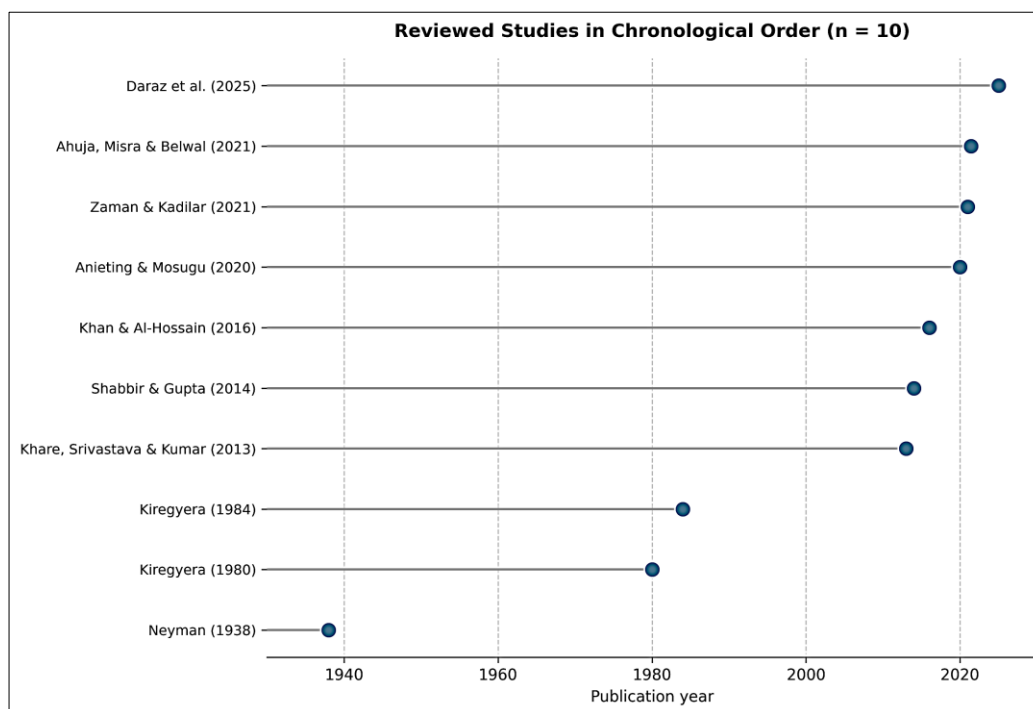


Fig 1: Reviewed studies in chronological order (n = 10)

A forty-two-year gap separates Neyman's foundational paper from Kiregyera's first chain-type estimator, after which the remaining nine studies are spread fairly evenly across the following forty-five years, with a visible cluster of three studies in 2020-2021 immediately following Anieting and Mosugu's modification of the Khan and Al-Hossain difference-type estimator.

Foundations: The two-phase design and early chain-type estimators

Neyman's original contribution was the sampling design itself rather than any specific estimator: he showed that allocating a fixed total cost between a large, cheap first-phase sample and a smaller, more expensive second-phase subsample could be optimised, and that the resulting two-phase mean estimator is more efficient than a single-phase design of comparable cost when a suitable auxiliary variable is available ^[1]. Every

estimator reviewed in this paper operates within that same two-phase cost structure, even though none of the later studies revisits the cost-allocation question directly.

Kiregyera's chain ratio-type estimator applies the classical ratio correction twice in sequence: first between the study variable and a first auxiliary variable, then again between that first auxiliary variable and a second, more widely available one, chaining the correction through both variables ^[2]. His 1984 regression-type estimator replaces the second ratio step with a regression adjustment, which typically reduces bias relative to the pure chain-ratio form at the cost of requiring an estimated regression coefficient ^[3]. Khare, Srivastava, and Kumar's 2013 estimator combines both ideas explicitly, using a ratio correction for one auxiliary variable and a regression correction for the other within a single generalised estimator, and it is this combined structure that Khan and Al-Hossain's difference-type estimator most directly descends from ^[4, 6].

Table 1: Chronological summary of reviewed studies

Study (Year)	Ref.	Estimator Type	Target Parameter	Auxiliary Variables
Neyman (1938)	[1]	Foundational design	N/A	Establishes two-phase design itself
Kiregyera (1980)	[2]	Chain ratio-type	Mean	Two
Kiregyera (1984)	[3]	Regression-type	Mean	Two
Khare, Srivastava & Kumar (2013)	[4]	Chain ratio-regression	Mean	Two
Shabbir & Gupta (2014)	[5]	Difference-cum-ratio-type	Variance	Two
Khan & Al-Hossain (2016)	[6]	Difference-type	Mean	Two
Anieting & Mosugu (2020)	[7]	Modified difference-type	Mean	Two
Zaman & Kadilar (2021)	[8]	Exponential-type	Mean	Not specified in abstract
Ahuja, Misra & Belwal (2021)	[9]	Generalized (ratio of means)	Mean (ratio)	Two moments
Daraz et al. (2025)	[10]	Double exponential-type	Variance	One (outlier/rank-based)

The difference-type estimator and its modifications

Khan and Al-Hossain's 2016 estimator adjusts the second-phase sample mean by a linear combination of two difference terms, one for each auxiliary variable, with the two multiplying constants chosen to minimise the estimator's mean squared error ^[6]. Because the resulting optimum constants depend on population variances and covariances that are themselves unknown in practice, most of the efficiency claims in this literature rest on estimating those

constants from the sample and treating the resulting mean squared error as approximately achieved, an approximation the original authors themselves flagged as needing further study across different populations. Anieting and Mosugu's 2020 modification adjusts the functional form of the difference terms while keeping the same two-auxiliary-variable, minimum-MSE structure, and reports efficiency gains over the original Khan and Al-Hossain form on the same type of numerical comparison ^[7]. Shabbir and Gupta's

earlier 2014 paper took the difference-type idea in a different direction, applying it to the population variance rather than the mean and combining it with a ratio-type term to form a difference-cum-ratio-type variance estimator, evaluated against four real datasets ^[5]. Daraz and colleagues' 2025

paper continues this variance-side branch, proposing a double exponential-type variance estimator that uses the outlier values and ranks of a single auxiliary variable rather than a second full auxiliary variable, validated through simulation and three additional datasets ^[10].

Table 2: Key concepts referenced in this review

Concept	Ref.	Description
Two-phase (double) sampling	[1]	A large first-phase sample measures an auxiliary variable to estimate its population mean; a smaller second-phase subsample then measures the study variable
Difference-type estimator	[6]	The sample mean adjusted by adding a linear multiple of the difference between the estimated and sample auxiliary means, with the multiplying constant chosen to minimise mean squared error
Chain ratio-type estimator	[2]	A ratio-type correction applied in two linked stages, once using a first auxiliary variable and again using a second, so the correction 'chains' through both variables
Bias and MSE to first order	[3]	Approximate expressions for an estimator's bias and mean squared error obtained by a Taylor expansion, keeping terms only up to the first order in the relative sampling errors

Table 2 defines the four terms most central to this review; the difference-type construction in the second row is the specific focus of this review, while the first, third, and fourth rows

describe the design and analytical machinery every reviewed estimator shares.

Table 3: Reference distribution by journal

Journal	Number of References
Journal of the American Statistical Association	1
Metrika	2
Journal of Scientific Research (Banaras Hindu University)	1
Communications in Statistics-Simulation and Computation	1
SpringerPlus	1
Open Journal of Mathematical Sciences	1
Communications in Statistics-Theory and Methods	1
Journal of Reliability and Statistical Studies	1
Axioms	1

Metrika carries both of Kiregyera's foundational papers and is the only journal in this list to contribute more than one

reference; The remaining eight references are spread across eight distinct statistics and applied-mathematics journals.

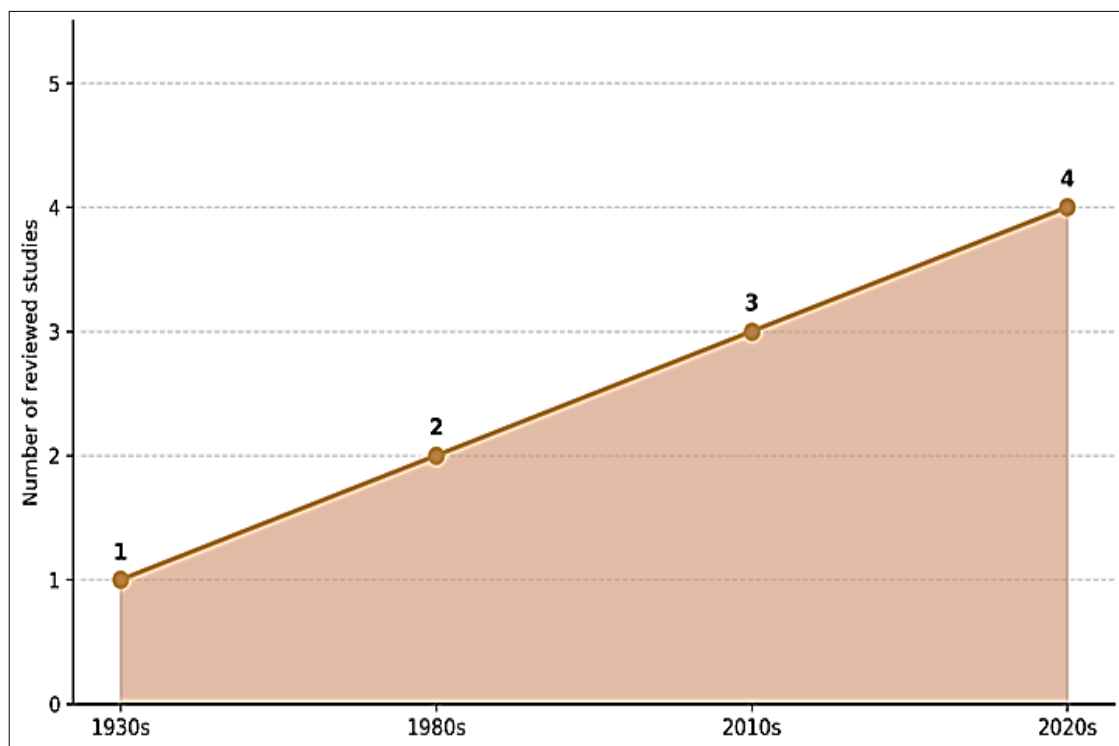


Fig 2: Reviewed studies by decade of publication ($n = 10$)

Activity has increased steadily across the four decades represented here, from a single foundational study in the 1930s to four studies in the 2020s so far. The doubling from

two studies in the 1980s to three in the 2010s and four in the 2020s suggests this specific estimator family has attracted more, not less, attention as it has matured.

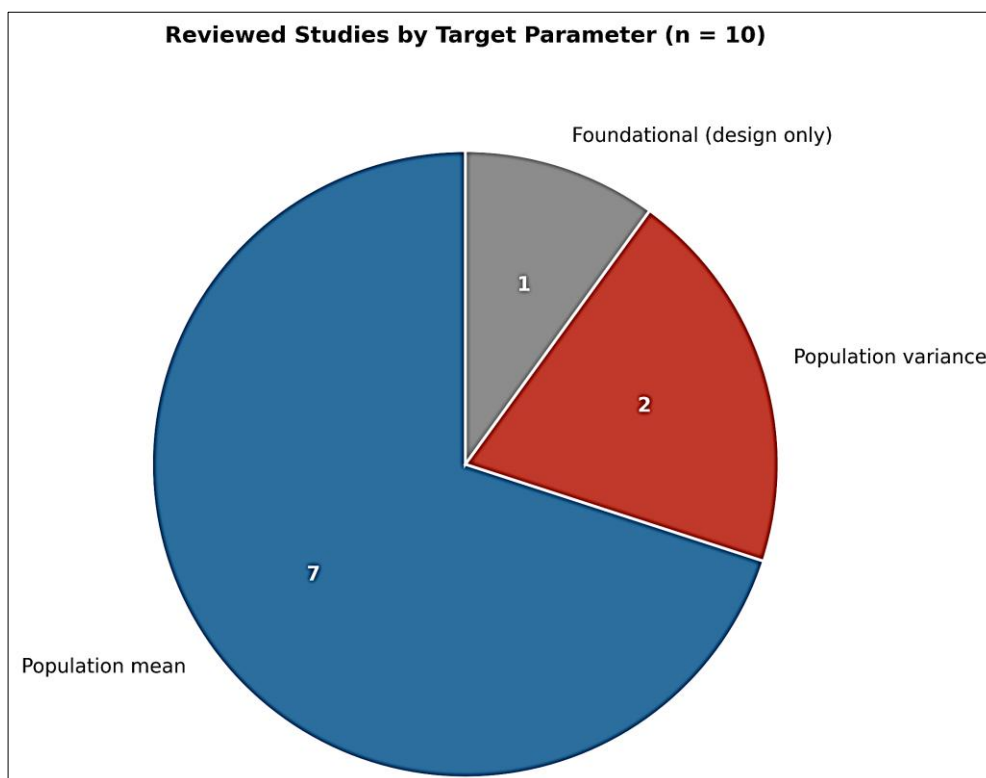


Fig 3: Reviewed studies by target parameter ($n = 10$)

Seven of the ten reviewed studies target the population mean, matching the focus named in this review's title; only two target the population variance, and Neyman's foundational

paper targets neither specifically, since it concerns the sampling design itself.

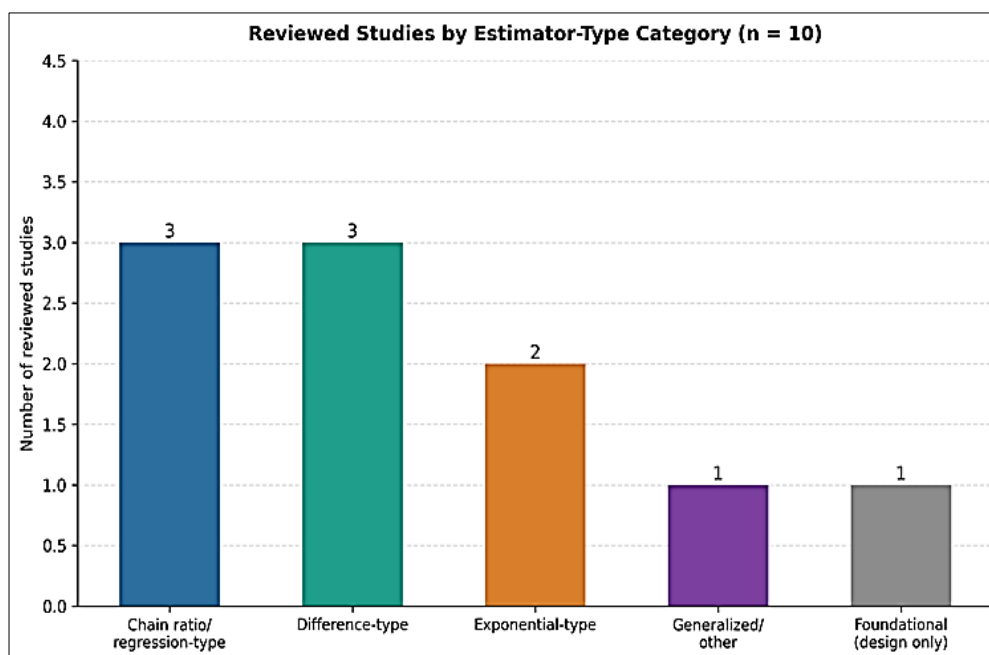


Fig 4: Reviewed studies by estimator-type category ($n = 10$)

Chain ratio/regression-type and difference-type estimators are the two largest categories, with three studies each, followed by exponential-type estimators with two. The generalised and foundational categories each contain a single study, indicating that most work in this literature refines one of two established estimator families rather than introducing an entirely new one.

Synthesis and research gaps

Read together, these ten studies show a literature that has repeatedly revisited the same basic construction, a linear or

ratio-based correction using one or two auxiliary variables under a two-phase design, rather than introducing fundamentally new estimation principles. The chain ratio-type and regression-type ideas of the early 1980s were combined by Khare, Srivastava, and Kumar in 2013, reformulated as a difference-type estimator by Khan and Al-Hossain in 2016, and then modified again by Anieting and Mosugu in 2020, with each step reported as an efficiency improvement over its immediate predecessor on the same or a similar numerical example [2, 4, 6, 7].

This pattern raises a question that none of the studies reviewed here addresses directly: because each successive paper reports its own efficiency comparison, generally on a small number of real or textbook datasets, it is difficult to judge from this literature alone whether the reported improvements reflect a genuine general property of each modification or are specific to the particular datasets used for illustration, an issue that mirrors similar concerns raised in related estimator literatures. A systematic comparison of all the mean-targeting estimators in Table 1 against a common set of benchmark populations would help establish which, if any, of the successive modifications offers a robust rather than dataset-specific improvement.

A second, more structural gap concerns the variance-targeting side of this literature. Only two of the ten studies reviewed here, Shabbir and Gupta and Daraz and colleagues, address the population variance rather than the mean, and the two are not directly comparable: one uses two auxiliary variables in a difference-cum-ratio construction, while the other uses outlier and rank information from a single auxiliary variable^[5, 10]. A difference-type variance estimator built directly on the two-auxiliary-variable structure that Khan and Al-Hossain and Anieting and Mosugu established for the mean has not yet appeared in the literature reviewed here.

Conclusion

This review has traced the literature on difference-type and related estimators for two-phase sampling from Neyman's 1938 foundational design through Kiregyera's early chain-type estimators to the difference-type construction introduced by Khan and Al-Hossain in 2016 and modified since.

The clearest gap identified here is the comparative narrowness of the variance-targeting literature relative to the mean-targeting literature, and the absence of a common benchmark comparison across the several mean-targeting modifications proposed since 2013. Extending the two-auxiliary-variable difference-type construction already well developed for the mean to the variance estimation problem, and testing the resulting estimator alongside its predecessors on a shared set of populations, would address both gaps using tools already established across the studies reviewed here.

Declarations

Conflict of interest: The author(s) declare no conflict of interest.

Funding: This review received no specific funding.

References

1. Neyman J. Contribution to the theory of sampling human populations. *J Am Stat Assoc.* 1938;33:101-116.
2. Kiregyera B. A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika.* 1980;27:217-223.
3. Kiregyera B. Regression-type estimator using two auxiliary variables and model of double sampling from finite populations. *Metrika.* 1984;31:215-223.
4. Khare BB, Srivastava U, Kumar K. A generalized chain ratio in regression estimator for population mean using two auxiliary characters in sample survey. *J Sci Res Banaras Hindu Univ Varanasi.* 2013;57:147-153.
5. Shabbir J, Gupta S. An improved generalized difference-cum-ratio-type estimator for the population variance in two-phase sampling using two auxiliary variables. *Commun Stat Simul Comput.* 2014;43(10):2540-2550.

6. Khan M, Al-Hossain AY. A note on a difference-type estimator for population mean under two-phase sampling design. *SpringerPlus.* 2016;5(1):723.
7. Anieting AE, Mosugu JK. A modified efficient difference-type estimator for population mean under two-phase sampling design. *Open J Math Sci.* 2020;4(1):195-199.
8. Zaman T, Kadilar C. New class of exponential estimators for finite population mean in two-phase sampling. *Commun Stat Theory Methods.* 2021;50:874-889.
9. Ahuja TK, Misra P, Belwal OK. A generalized two-phase sampling estimator of ratio of population means using auxiliary information. *J Reliab Stat Stud.* 2021;14(1):1-16.
10. Daraz U, Agustiana D, Wu J, Emam W. Twofold auxiliary information under two-phase sampling: an improved family of double-transformed variance estimators. *Axioms.* 2025;14:64.