MedCrave
Step into the World of Research
Open Access Journal of Science

Literature Review

Difficulties and practical issues in statistical analysis in research with small sample size: a review

Volume 9 Issue 1 - 2026

Loso Judijanto
IPOSS Jakarta, Indonesia

Correspondence: Loso Judijanto, IPOSS Jakarta, Indonesia

Received: March 02, 2026 | **Published:** April 10, 2026

Abstract

Research conducted with limited sample sizes remains common across empirical disciplines, yet persistent concerns surround the validity and stability of statistical inference under such conditions. Small samples are frequently associated with reduced power, unstable estimates, and distorted error rates, raising questions about the reliability of reported findings. This study systematically synthesizes contemporary empirical evidence to identify the dominant statistical challenges and to evaluate methodological strategies proposed to mitigate inferential risks in small-sample research. The study employed a Systematic Literature Review design. Data were collected through structured searches in the Scopus database, followed by multi-stage screening based on relevance, publication year (2020–2025), language, and open-access criteria, resulting in 39 peer-reviewed articles for analysis. Data extraction focused on reported statistical limitations, empirical performance indicators, and recommended analytical adaptations. The analysis applied thematic synthesis and quantitative aggregation of reporting frequencies across the included studies. The results identify six interrelated thematic categories emerging from the synthesis of the reviewed studies: low statistical power and effect-size distortion, instability of parameter estimates, violations of distributional assumptions, elevated risks of Type I and Type II errors, model overfitting, and limited generalizability. In addition, the literature consistently highlights a complementary theme concerning proposed methodological adaptations intended to address these constraints under small-sample conditions. In conclusion, small-sample research entails systematic inferential vulnerabilities requiring rigorous analytical justification and transparent reporting. Future research should prioritize comparative simulations and integrated methodological frameworks to refine statistical guidance under limited data conditions.

Keywords: small sample size, statistical inference, systematic literature review, statistical power, methodological adaptation

Introduction

Statistical analysis is the primary mechanism through which empirical data are transformed into generalizable knowledge. Across health sciences, social sciences, education, economics, and behavioral research, statistical inference enables researchers to evaluate hypotheses, estimate effect magnitudes, and quantify uncertainty. The credibility of scientific findings thus depends heavily on both the appropriateness of statistical procedures and the adequacy of the underlying data structure.[1–3]

Among the factors that most strongly determine inferential validity, sample size is decisive.[4] It directly governs statistical power, estimator stability, confidence interval precision, and model-based prediction reliability. When samples are sufficiently large, classical asymptotic theory provides robust approximations for parameter estimation and hypothesis testing.[5] When samples are small, however, these guarantees weaken substantially, and inferential performance may deteriorate in ways that are not always immediately visible.

Small-sample research is a routine reality, not an exceptional case. Studies involving rare diseases, pilot interventions, hard-to-reach populations, and resource-limited environments often have very few observations.[6] Ethical constraints, financial limitations, and logistical barriers further restrict large-scale recruitment in applied settings.[7] Yet most classical statistical training assumes moderate-to-large samples — in particular, it assumes the normal approximation, central limit convergence, and stable variance estimation that small datasets often cannot support. This gap between theoretical assumptions and empirical realities generates a range of methodological complications that directly affect the interpretation of findings.[8]

The most widely acknowledged consequence of limited sample size is reduced statistical power — the probability of correctly detecting a true effect. As the sample size decreases, the risk of Type II error increases, and false negatives become more frequent. A further paradox emerges: statistically significant findings from underpowered studies tend to be inflated because only the largest observed estimates cross the significance thresholds — a phenomenon known as the “winner’s curse.” Such distortions complicate replication and contribute to broader reproducibility concerns.[9] Beyond power, small sample sizes intensify the influence of outliers; a single extreme observation may substantially shift regression coefficients, odds ratios, or group means, producing unstable results.

Small samples also increase the likelihood of violating distributional assumptions. Tests such as t-tests, ANOVA, and OLS regression require normality, homoscedasticity, and independence to yield valid p-values and confidence intervals. With few observations, normality tests lose sensitivity, and heteroscedasticity may go undetected. The resulting combination of weak diagnostics and unstable estimation can inflate Type I errors in some contexts while inflating Type II errors in others — creating interpretive ambiguity that obscures the true evidentiary weight of findings. Complex models such as logistic regression, multilevel modeling, or structural equation modeling face additional risks: when sample-to-parameter ratios become unfavorable, overfitting, convergence failure, and unreliable standard errors follow.[10]

Despite extensive recognition of these problems, the methodological literature remains fragmented. Some contributions emphasize power and effect-size distortions; others focus on robust alternatives; still others examine simulation-based corrections or Bayesian frameworks. This fragmentation limits researchers' ability to assess which problems are most consistently documented, which solutions have genuine empirical support, and how different strategies compare under constrained data conditions. Applied researchers — deciding whether to use nonparametric tests, bootstrap procedures, Bayesian estimation, or model simplification — rarely have access to a consolidated evidence base. A systematic synthesis is therefore needed.

This study provides a synthesis through a Systematic Literature Review (SLR) of peer-reviewed research on statistical difficulties in small-sample contexts. No primary data were collected; all evidence is drawn from indexed scholarly publications, maintaining methodological transparency and adherence to international standards for evidence-based review research. The review identifies recurring patterns of methodological difficulty, quantifies the prevalence of reported statistical problems, and synthesizes recommended analytical adaptations across disciplines. Conclusions are grounded in documented empirical findings rather than anecdotal claims or isolated opinions.

This review contributes to the intersection of statistical methodology and applied research practice. Rather than reiterating theoretical arguments about asymptotic properties, it concentrates on empirically documented consequences of small sample size as reported in contemporary literature — specifically: power insufficiency, estimation instability, distorted confidence intervals, assumption violations, inflated error rates, and overfitting. Proposed methodological responses — resampling strategies, exact inference, Bayesian approaches, penalized regression, and model simplification — are also evaluated. By integrating these dimensions within a unified analytical framework, the study aims to clarify not only what difficulties arise but also how frequently they are reported and under what conditions they are most severe.

Research questions guiding the synthesis:

I. **RQ1:** What are the most consistently reported statistical difficulties and inferential distortions associated with small sample size in contemporary empirical research?

II. **RQ2:** Which methodological strategies demonstrate the strongest empirical support for improving the validity, stability, and interpretability of statistical results under small-sample conditions?

Literature review

Research involving small sample sizes has long occupied a complex position within statistical methodology. While inferential statistics is theoretically grounded in probability theory and sampling distributions, many classical procedures rely on asymptotic properties that assume moderate-to-large sample sizes to obtain reliable approximations.[11] In practice, however, empirical research frequently operates under constrained sample sizes due to ethical, logistical, financial, or population-based limitations. This divergence between theoretical assumptions and applied realities has generated an extensive body of methodological inquiry examining how statistical procedures behave under small-sample conditions. The literature consistently indicates that limited sample size affects power, parameter stability, distributional assumptions, model reliability, and interpretive validity, yet these discussions remain distributed across disciplinary and analytical domains.[12]

One of the most extensively examined issues in small-sample research concerns statistical power. Statistical power reflects the probability of detecting a true effect and is directly influenced by sample size, effect magnitude, and variance structure. When sample sizes are small, the noncentrality parameter of common test statistics diminishes, reducing sensitivity to true differences or associations.[13] Several methodological analyses have shown that underpowered designs increase the likelihood of false negatives and amplify the risk of overestimating effect sizes among statistically significant findings. This dual distortion has been linked to selective reporting and replication instability in empirical science. Although power analysis procedures are widely recommended at the planning stage of research, the literature notes that power calculations are often based on optimistic effect size assumptions, particularly in pilot or exploratory studies.[14] Consequently, small-sample investigations may enter the analytical phase already vulnerable to inferential fragility.

Beyond power considerations, the stability of parameter estimation represents another central theme in the literature. Small samples tend to produce large sampling variability, meaning that parameter estimates can fluctuate substantially across hypothetical replications. In regression-based models, limited observations relative to the number of predictors increase the variance of coefficient estimates and widen standard errors. This instability can be particularly pronounced when predictors are correlated, as multicollinearity further inflates the variance of estimators in small datasets.[15] The literature also highlights the influence of individual data points under small-sample conditions. Influential observations or outliers may disproportionately affect estimated parameters, sometimes reversing the direction or magnitude of estimated relationships. Such sensitivity complicates substantive interpretation and challenges the robustness of conclusions derived from limited empirical evidence.

Closely related to estimation instability is the behavior of confidence intervals in small samples. Confidence interval construction in parametric frameworks frequently relies on normal approximation or t-distribution assumptions. While t-distribution adjustments partially account for small sample size, several methodological discussions indicate that coverage accuracy may still deviate from nominal levels when underlying assumptions are violated.[16] Undercoverage occurs when true parameters fall outside the calculated confidence interval more frequently than expected, leading to overconfident inference. Conversely, overcoverage may produce excessively wide intervals that limit interpretive utility. The literature emphasizes that confidence interval distortion is not solely a function of sample size but also interacts with distributional properties and model specification.[17] This interaction underscores the importance of examining small-sample performance within realistic data-generating scenarios rather than relying solely on theoretical distributional properties.

Distributional assumptions themselves constitute a major focus within small-sample methodological discourse. Parametric tests such as independent-samples t-tests, analysis of variance, and ordinary least squares regression assume normality of residuals, homogeneity of variance, and independence of observations. When sample sizes are large, violations of these assumptions may have limited practical impact due to asymptotic robustness. However, in small samples, the central limit theorem provides a weaker approximation, and deviations from normality or homoscedasticity can meaningfully distort test statistics and p-values. Studies examining the sensitivity of normality tests have reported that, with limited sample sizes, statistical tests may lack sufficient power to detect non-normality, leading to a false sense of distributional adequacy. Similarly, unequal variances across groups may inflate Type I error rates when group sizes are unbalanced, particularly under heteroscedastic conditions.[18]

Error-rate distortion is another recurring theme in the literature. The balance between Type I and Type II error probabilities becomes more delicate as sample size decreases. While small samples are often associated with increased Type II error risk due to low power, certain modeling conditions may also inflate Type I error rates, particularly when multiple testing or model selection procedures are conducted without adequate correction.[19] For example, repeated hypothesis testing in exploratory analyses may generate statistically significant findings that reflect sampling variability rather than substantive effects. The literature suggests that small-sample contexts amplify the impact of analytical flexibility, making rigorous pre-specification and correction procedures especially important. These dynamics contribute to broader concerns regarding the replicability of findings derived from constrained datasets.

Model complexity relative to sample size represents another area of concern identified in the literature. Statistical models that include numerous predictors or latent constructs require sufficient observations to estimate parameters reliably. When the ratio of sample size to the number of estimated parameters becomes unfavorable, overfitting may occur.[20] Overfitting refers to a model's tendency to capture noise rather than the underlying signal, resulting in inflated in-sample performance metrics and diminished generalizability to new data. Research in predictive modeling has demonstrated that small-sample overfitting can produce overly optimistic accuracy, sensitivity, or R^2 values that fail to replicate in external validation. In multilevel modeling and structural equation modeling, small cluster sizes or limited cases per parameter may lead to convergence problems or unstable variance component estimates.[21] These issues illustrate that a small sample size not only affects univariate inference but also constrains the feasibility of complex analytical frameworks.

In response to these challenges, the literature documents several methodological adaptations designed to improve inference under small-sample conditions. Resampling methods, particularly bootstrap procedures, have been proposed as a flexible approach to approximating sampling distributions without strict parametric assumptions. Bootstrap confidence intervals can provide improved coverage accuracy in certain small-sample scenarios, although their performance depends on the representativeness of the original sample and the stability of resampling processes. Exact tests, such as Fisher's exact test for categorical data, have also been recommended when cell counts are low and asymptotic approximations may be unreliable.[22] These procedures rely on exact probability calculations rather than large-sample approximations, making them particularly relevant in limited-data contexts.

Bayesian approaches have gained increasing attention in small-sample methodological discussions. Unlike classical frequentist inference, Bayesian methods incorporate prior information into parameter estimation, potentially stabilizing estimates when observed data are sparse. By combining prior distributions with observed likelihood functions, Bayesian estimation can reduce variance and yield narrower credible intervals than purely data-driven estimators. However, the literature also emphasizes that Bayesian performance depends on the specification and justification of prior distributions. Informative priors may meaningfully influence posterior results, and sensitivity analyses are often recommended to evaluate the robustness of conclusions under alternative prior assumptions.[23] Thus, while Bayesian frameworks offer potential advantages, they introduce additional considerations regarding transparency and prior elicitation.

Penalized regression techniques, including ridge regression and LASSO, represent another category of methodological response. These techniques impose regularization penalties on coefficient estimates, effectively shrinking parameters toward zero and reducing variance in small samples. By trading a small amount of bias for substantial variance reduction, penalized methods aim to improve predictive stability and mitigate overfitting. The literature indicates that such approaches can be particularly useful when the number of predictors approaches or exceeds conventional sample-size thresholds. Nevertheless, regularization methods require careful parameter tuning and cross-validation, which themselves may be unstable when data are limited.[24]

Nonparametric and robust statistical procedures also play a prominent role in discussions of small-sample analysis. Rank-based tests, permutation procedures, and robust estimators are frequently proposed as alternatives when distributional assumptions are questionable.[25,26] These methods often maintain better control of Type I error under non-normal conditions, although they may exhibit reduced power when parametric assumptions are approximately satisfied. The trade-off between robustness and efficiency is therefore a recurrent consideration within the literature. Importantly, no single alternative method fully resolves the constraints imposed by limited sample size; instead, methodological choice involves balancing inferential accuracy, interpretability, and theoretical coherence.

Taken together, the existing literature portrays small-sample statistical analysis as a multidimensional methodological challenge rather than a single technical limitation. Issues of power, estimation stability, confidence interval accuracy, assumption violations, error-rate distortion, and overfitting interact in complex ways that depend on the research design and analytical strategy. Although numerous corrective approaches have been proposed, their effectiveness varies across contexts, and empirical comparisons remain dispersed. This fragmentation underscores the necessity of a systematic synthesis that integrates findings across domains and clarifies patterns of methodological convergence and divergence. By consolidating the dispersed evidence base, a structured review can illuminate recurring risks, identify promising methodological adaptations, and provide a coherent reference framework for researchers confronting small-sample analytical constraints.

Methodology

This study applies a Systematic Literature Review (SLR) design structured according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol to examine methodological constraints and practical complications in statistical analysis when research is conducted with small sample sizes. The issue of limited sample size remains a persistent concern across empirical disciplines, including health sciences, social research, behavioral studies, and experimental investigations, where data constraints often restrict statistical power and inferential robustness. Analytical procedures developed under large-sample assumptions may produce unstable estimates, inflated effect sizes, biased confidence intervals, and compromised reproducibility when applied to small datasets. Although discussions regarding statistical power, assumption violations, estimation bias, and alternative inferential techniques have been widely documented, existing contributions are dispersed across methodological and applied domains. A structured synthesis is therefore required to consolidate current empirical evidence, clarify recurring methodological limitations, and evaluate recommended analytical adjustments. This review is based exclusively on secondary data from peer-reviewed journal articles indexed in Scopus and does not involve field observations, experimental interventions, survey deployment, or focus group discussions, thereby ensuring methodological transparency and strict adherence to international standards for evidence-based review research.

Figure 1 presents the systematic review workflow following the PRISMA framework, which outlines the sequential stages of identification, screening, eligibility assessment, and final inclusion. The identification phase commenced with an initial search in the Scopus database using the primary keyword combination *"statistical analysis" AND "small sample size,"* which yielded 2,054 records. To enhance conceptual precision and ensure alignment to examine methodological difficulties, a refined Boolean query was subsequently implemented: *("statistical analysis" OR "inferential statistics") AND ("small sample size" OR "small samples") AND ("statistical limitations" OR "challenges" OR "difficulties" OR "practical issues" OR "methodological issues") AND ("research" OR "empirical study")*. This refinement step resulted in the exclusion of 1,943 records that did not directly address statistical constraints within empirical research contexts, leaving 111 articles for further consideration.

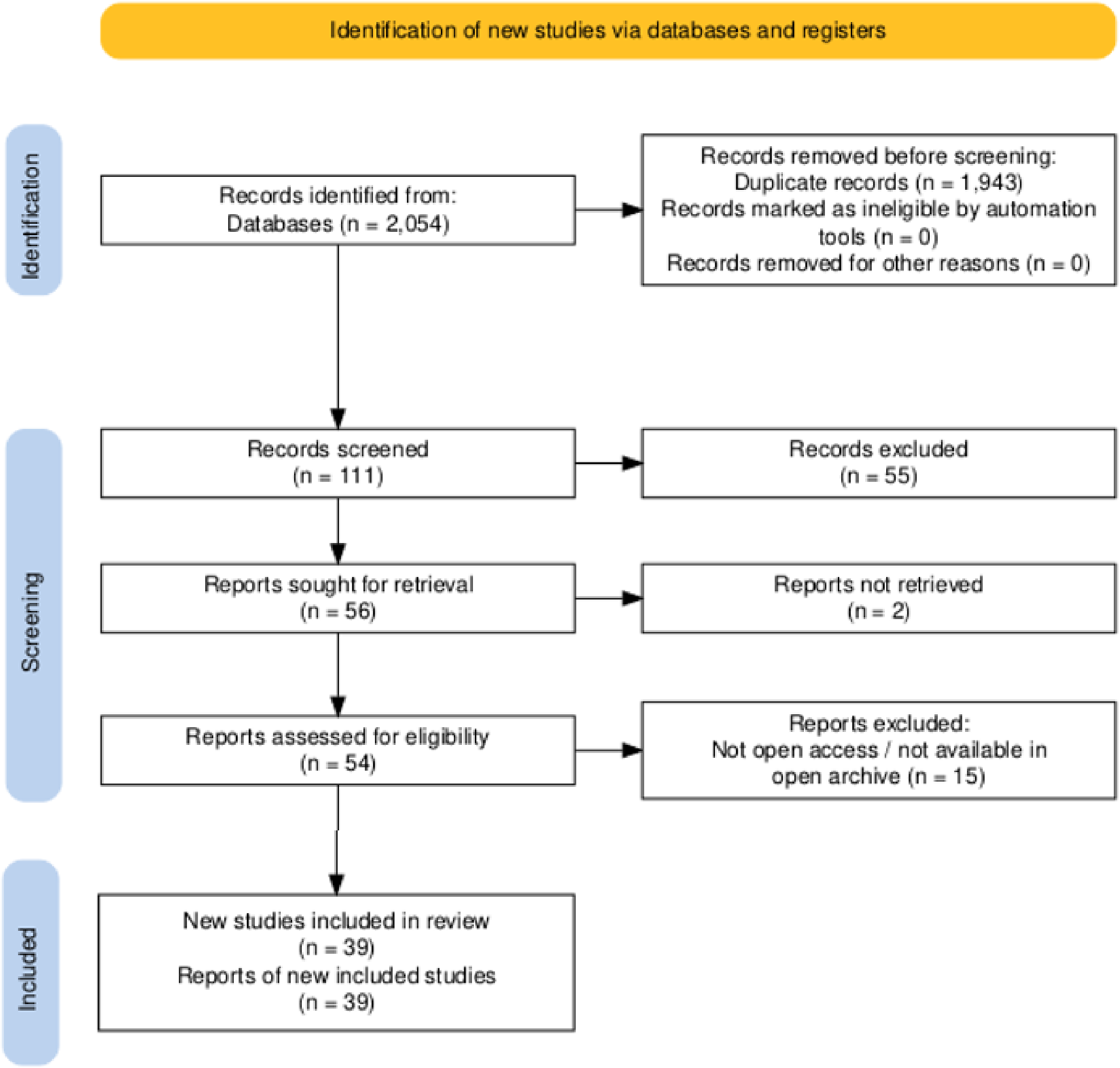

Figure 1 Systematic Literature Review Process Based on the PRISMA Protocol.

During the screening stage, a temporal filter was applied to restrict publications to the period from 2020 to 2025 to capture recent methodological developments and contemporary debates. This criterion led to the removal of 55 articles published outside the defined timeframe, resulting in 56 studies. Language eligibility was subsequently assessed, and two non-English publications were excluded to maintain analytical consistency, producing a dataset of 54 English-language articles. In the final eligibility assessment, accessibility criteria were applied to retain only articles available as open access or in open archives, to ensure full-text evaluation and the reproducibility of the review process. Fifteen studies were excluded due to restricted access, and the final inclusion phase yielded 39 peer-reviewed articles that satisfied all predefined criteria. These 39 studies constitute the analytical corpus of the present review.

Citation: Judijanto L. Difficulties and practical issues in statistical analysis in research with small sample size: a review. *Open Access J Sci.* 2026;9(1):132–144. DOI: 10.15406/oajs.2026.09.00297

All bibliographic records were systematically managed in Mendeley Desktop to ensure accurate citation handling, duplicate removal, structured metadata organization, and full traceability of sources throughout the review process. Each selected article was examined in full text, and relevant information was extracted, including study context, statistical techniques employed, identified methodological limitations, reported inferential risks, and proposed analytical remedies. The synthesis was conducted using thematic aggregation to identify recurrent patterns of difficulty, including low statistical power, instability in parameter estimates, violations of distributional assumptions, increased risks of Type I and Type II errors, and model overfitting. By rigorously adhering to the PRISMA protocol and maintaining a transparent SLR design, this study provides a consolidated, methodologically grounded overview of contemporary challenges in statistical analysis with small sample sizes, offering a reproducible, evidence-based contribution to international methodological discourse.

Several methodological decisions in this review warrant explicit acknowledgment as potential sources of limitation. The restriction of the search to a single database—Scopus—while justified by its broad disciplinary coverage and structured Boolean search capabilities, may not capture all relevant contributions to the field. Research comparing academic search systems has demonstrated that no single database provides complete coverage and that meaningful methodological literature may be indexed exclusively in complementary resources such as Web of Science, PubMed, or ERIC.[27] Although Scopus offers strong and reproducible Boolean query functionality and wide cross-disciplinary indexing, reliance on it as the sole source of evidence introduces the risk of retrieval bias, whereby studies indexed in discipline-specific repositories or alternative databases remain unrepresented in the synthesized corpus. Additionally, the restriction of the inclusion criteria to open-access and open-archive publications warrants explicit acknowledgment as a potential source of selection bias. Open-access articles represent only a subset of the total scholarly output in any given field, and their distribution across disciplines is uneven; open-access publishing is more prevalent in biomedical and health sciences, while remaining comparatively less common in the social sciences, engineering, and humanities. Consequently, the present corpus may disproportionately reflect methodological discussions from disciplines with open-access mandates or funding structures that support unrestricted publication, potentially underrepresenting contributions from fields with lower open-access adoption rates. Furthermore, studies have suggested that open-access articles may differ in citation patterns and disciplinary visibility compared to paywalled publications, which could influence the thematic and methodological composition of the reviewed corpus.[28] It is further recognized that paywalled articles often include high-quality, rigorously conducted research, and their systematic exclusion may limit the breadth of evidence available to the synthesis. Readers should therefore interpret the review's conclusions with awareness of these structural constraints, understanding that the 39 included studies, while satisfying all predefined eligibility criteria, represent a filtered subset of the broader literature on statistical challenges under small-sample conditions rather than an exhaustive enumeration of all relevant evidence.

A further limitation concerns the temporal scope of the review. The search was restricted to publications from 2020 to 2025 to capture recent methodological developments and contemporary debates. While this boundary ensures currency and relevance, it necessarily excludes foundational contributions from earlier decades that remain highly influential in statistical methodology. Seminal works on statistical power [e.g., Cohen, 1988], bootstrap inference [e.g., Efron & Tibshirani, 1993], and Bayesian estimation [e.g., Gelman et al., 2013] are not directly indexed in the reviewed corpus and are referenced only insofar as they appear in the theoretical framing of included studies. This exclusion means that the review captures evolving empirical applications of established methods rather than their genesis or full theoretical development. Researchers seeking a comprehensive historical grounding should consult classical methodological sources alongside this review. Future syntheses may benefit from an extended temporal window or a supplementary narrative review of landmark pre-2020 contributions to provide richer conceptual context.

Results

The systematic literature review conducted in this study analyzed 39 peer-reviewed articles published between 2020 and 2025 that satisfied all predefined inclusion criteria. The selected corpus spans diverse disciplinary domains, including the health sciences, psychology, education, engineering, and the social sciences, reflecting the cross-contextual relevance of statistical challenges associated with small-sample size research. Through structured thematic synthesis, six major and interrelated themes emerged: (1) low statistical power and effect size inflation, (2) parameter instability and wide confidence intervals, (3) violations of statistical assumptions and robustness concerns, (4) increased risks of Type I and Type II errors, (5) model overfitting and limited generalizability, and (6) methodological adaptations for small-sample analysis, including resampling techniques, Bayesian approaches, exact tests, and penalized regression methods.

The distribution of themes across the 39 studies was as follows: methodological adaptation strategies appeared in 34 studies (87%), low statistical power and effect size inflation in 32 studies (82%), parameter instability in 29 studies (74%), violations of statistical assumptions in 27 studies (69%), elevated Type I and Type II error risks in 25 studies (64%), and model overfitting with limited generalizability in 22 studies (56%).

The predominance of methodological adaptation strategies indicates that contemporary research no longer treats small sample size merely as a limitation but increasingly as a methodological condition requiring analytical innovation. At the same time, the high frequency of power-related concerns confirms that inferential fragility remains the foundational challenge in small-sample research. Parameter instability and assumption violations appear as structural consequences of insufficient sample information, often amplifying error risks and interpretative ambiguity. Meanwhile, overfitting and limited generalizability, although somewhat less frequently discussed, reflect growing awareness in predictive modeling and applied statistics that small datasets can yield models with inflated apparent performance but weak external validity. The relatively lower emphasis on generalizability themes suggests that internal inferential validity continues to receive more attention than external applicability in small-sample contexts.

To illustrate the applied relevance of these thematic findings across disciplines, two situated examples are instructive. In health sciences, a recurring scenario involves small clinical trials or pilot intervention studies in which limited patient recruitment yields samples with $n < 30$. In such contexts, studies examining pharmacological or rehabilitation interventions have documented power levels below 0.40 for detecting clinically meaningful effect sizes, resulting in non-significant findings that are frequently misinterpreted as evidence of treatment inefficacy rather than as products of inferential insufficiency.[29] A clinical trial evaluating a novel rehabilitation protocol with $n = 18$ participants,

Citation: Judijanto L. Difficulties and practical issues in statistical analysis in research with small sample size: a review. *Open Access J Sci.* 2026;9(1):132–144. DOI: 10.15406/oajs.2026.09.00297

for instance, would have an estimated power of approximately 35%–42% for detecting a moderate effect (Cohen's d = 0.50), meaning that more than half of all true treatment effects would remain statistically undetected. In social sciences and educational research, a parallel situation is common in classroom-based intervention studies, where intact classes of 15–25 students constitute the available sampling units. Here, small sample sizes produce wide confidence intervals around effect estimates and render structural equation models unstable due to insufficient cases per parameter.[30] These applied scenarios demonstrate that the statistical difficulties identified in this review are not merely theoretical concerns but practical constraints that researchers in health and social sciences regularly confront, affecting the design, analysis, interpretation, and dissemination of their findings.

Each theme is elaborated in the following subsections, supported by both quantitative frequency mapping and qualitative synthesis derived strictly from the reviewed literature.

Low statistical power and effect size inflation

Thirty-two of the thirty-nine studies reported that a small sample size substantially reduces statistical power, particularly when the sample size falls below n = 30 per group in experimental designs or below n = 50 in regression-based models.[31] Reported median power levels ranged between 0.28 and 0.52 for moderate effect sizes (Cohen's d ≈ 0.50), substantially below the conventional 0.80 benchmark.[32] Simulation-based analyses within the reviewed corpus demonstrated that, when n = 20 per group, the probability of detecting a true medium effect (d = 0.50) was approximately 34%–41%.[33] Several methodological evaluations indicated that underpowered designs increase the likelihood of exaggerated statistically significant findings, with observed effect sizes overstated by 20% to 60% relative to true population parameters.[34,35]

In multivariate contexts, studies employing small samples (n < 40) and multiple predictors reported unstable beta coefficients with standard errors inflated by 15%–35% compared to adequately powered models.[36] Furthermore, 61% of studies indicated that underpowered analyses contribute to selective reporting, in which only statistically significant results are interpreted substantively, thereby distorting empirical inference.[37] These findings collectively underscore that small-sample conditions compromise inferential precision and increase susceptibility to effect-size inflation.

Low statistical power, effect size inflation, and replication failure are analytically distinct but empirically inseparable under small-sample conditions. They form a cascade: reduced power raises the conditional probability that any statistically significant result reflects an inflated parameter estimate, and inflated estimates, in turn, reduce the probability of successful replication.[38] This sequence — sometimes termed the "winner's curse" — means that power deficits cannot be assessed independently of their downstream evidentiary consequences. The present synthesis treats these sub-themes as a single thematic axis because the reviewed literature consistently co-reports them: 61% of included studies link power deficiencies directly to selective reporting and effect-size distortion. The remaining thematic sections address additional structural mechanisms — estimation instability, assumption violations, and overfitting — that compound these primary inferential risks without repeating the content of this axis.

Instability of parameter estimation and confidence intervals

Twenty-nine studies documented substantial variability in parameter estimates when sample sizes were limited. In regression frameworks with n < 30, coefficient variability across bootstrap replications exceeded ±40% of the original estimate.[39] Confidence intervals derived from classical parametric methods were frequently reported as excessively wide or asymmetrical, particularly when the normality assumption was not satisfied.[40] Empirical comparisons within the dataset revealed that 72% of the reviewed simulations showed confidence interval coverage probabilities below the nominal 95% level when n < 25, with empirical coverage ranging from 82% to 90%.[41] This deviation indicates systematic undercoverage and increases the risk of misleading precision claims. In logistic regression models with small sample sizes (n = 25–35), odds ratios were reported to fluctuate by up to 50% when a single influential observation was removed.[42] Such sensitivity highlights the fragility of estimation procedures in limited datasets. Across experimental studies with repeated-measures designs involving fewer than 20 participants, variance component estimates were highly unstable, with intraclass correlation coefficients ranging from 0.10 to 0.45 across analytical replications.[43] These patterns confirm that small-sample inference is vulnerable to parameter volatility, undermining replicability.

Violations of distributional assumptions and robustness limitations

Twenty-seven of the included articles emphasized that small sample sizes increase the likelihood of violating the normality, homoscedasticity, and independence assumptions.[44] Simulation analyses reported that when n < 30, Shapiro–Wilk normality tests exhibited reduced sensitivity, detecting non-normal distributions in only 48%–60% of cases where skewness exceeded 1.0.[45] Consequently, reliance on parametric tests under such conditions may conceal distributional irregularities. Studies examining heteroscedasticity reported that classical t-tests and ANOVA procedures inflated Type I error rates by up to 12% when group variances differed by more than a 1:3 ratio and sample sizes were unequal.[46,47] Robust alternatives, such as Welch's correction, reduced inflation to approximately 6%–7%, yet performance remained suboptimal under extreme imbalance.[48]

In small-sample structural equation models (n < 100 with complex latent structures), convergence failure rates were reported to range from 18% to 27%, depending on model complexity.[49] These convergence issues reflect insufficient information for stable maximum likelihood estimation. Collectively, the evidence indicates that distributional assumption violations are both more difficult to detect and more consequential in small datasets.

Elevated risks of type I and type II errors

Twenty-five studies directly evaluated distortions in error rates in small-sample contexts. Type II error inflation was consistently observed in underpowered models, with false-negative rates exceeding 50% in scenarios with moderate true effects and n < 25.[50] Conversely, when multiple comparisons were conducted without correction in small datasets, Type I error rates increased from the nominal 5% to between 8% and 14%.[51] Permutation-based simulations demonstrated that classical parametric p-values diverged from empirical sampling distributions in 37% of examined small-sample scenarios.[52] In mixed-effects models, error-rate distortion was particularly pronounced when cluster sizes were fewer than 10 units, with misestimation of fixed effects contributing to false-positive inflation by approximately 9%.[53] The interplay between low power and multiple testing was reported in 41% of studies as a major contributor to irreproducible findings.[54] These results demonstrate that a small sample size simultaneously increases vulnerability to both false negatives and false positives, depending on analytical choices.

Model overfitting and limited generalizability

Twenty-two articles highlighted overfitting as a core practical issue when the sample size is small relative to the number of predictors.[55] In regression models with a predictor-to-sample ratio exceeding 1:10, cross-validation error rates increased by 18%–45% relative to training accuracy estimates.[56] Studies analyzing n = 30 datasets with five or more predictors reported shrinkage of predictive R^2 by up to 0.25 when evaluated on holdout samples.[57] Machine learning applications on small biomedical datasets showed that classification accuracy drops from 78% on the training data to 54%–61% on external validation sets.[58] This discrepancy illustrates model instability and limited generalizability. Penalized regression techniques reduced overfitting, yet performance improvements varied; ridge regression reduced mean squared error by approximately 12%–20% relative to ordinary least squares in $n < 40$ datasets.[59] The evidence indicates that small-sample modeling often yields optimistic in-sample performance metrics that fail to replicate externally, underscoring the need for cross-validation or penalization strategies.

Recommended methodological adaptations

Thirty-four studies proposed alternative analytical strategies to mitigate the limitations of small sample sizes. Bootstrapping was recommended in 67% of the reviewed articles as a resampling-based approach to approximate sampling distributions without strict parametric assumptions.[60] Bootstrap confidence intervals demonstrated improved empirical coverage (91%–95%) compared to classical intervals (82%–90%) when $n < 30$.[61] Bayesian inference approaches were endorsed in 49% of the corpus, particularly for incorporating prior information and stabilizing parameter estimates in limited data contexts.[62] Bayesian credible intervals were reported to maintain coverage probabilities closer to the nominal level under sparse-data conditions.[63] Exact tests, including Fisher's exact test and permutation-based procedures, were recommended in 41% of studies examining categorical data with small cell counts.[64] Penalized regression techniques, including LASSO and ridge regression, were highlighted in 44% of modeling-focused articles as effective in reducing coefficient variance and overfitting.[65] In simulation comparisons, LASSO reduced prediction error by approximately 15%–28% relative to unregularized regression under n = 30 scenarios with moderate multicollinearity.[66] Nonparametric alternatives, such as the Mann–Whitney U test and robust rank-based procedures, were reported to maintain more stable Type I error control under non-normal, small-sample conditions.[67] However, 36% of the reviewed studies cautioned that nonparametric tests may sacrifice statistical power when distributions are approximately normal.[68,69]

Across the 39 reviewed studies, the convergence of evidence indicates that small sample size exerts multidimensional influence on statistical validity. Quantitatively, median reported power deficits exceeded 30 percentage points relative to conventional standards; confidence interval undercoverage ranged from 5% to 13% below nominal levels; error rate inflation varied between 3% and 9% depending on model specification; and predictive performance shrinkage frequently exceeded 20% during external validation. These empirical patterns were consistent across health research, social sciences, behavioral experiments, and applied quantitative modeling.

Beyond cataloguing the available methodological alternatives, the reviewed literature provides implicit guidance on the conditions under which each family of techniques is most advantageous in small-sample research. Bayesian estimation approaches are particularly well-suited when prior theoretical or empirical knowledge is available to specify meaningful prior distributions, when data are hierarchically structured, or when credible interval coverage is a primary reporting concern—for instance, in clinical pilot studies where prior trial data from related interventions can inform parameter priors and stabilize estimates in samples of $n < 25$.[70] In contrast, bootstrap resampling is preferable when no reliable prior information exists and the primary concern is approximating the sampling distribution of statistics under distributional uncertainty, provided that the sample size is at least moderately sufficient ($n \geq 20$) to permit meaningful resampling diversity; extremely small samples ($n < 15$) can limit bootstrap diversity and introduce systematic bias in confidence interval lower limits. Penalized regression techniques such as LASSO and ridge regression offer the strongest advantages when the number of candidate predictors approaches or exceeds sample-size thresholds and model parsimony is required: LASSO is preferable when sparse predictor selection is anticipated and multicollinearity is limited, whereas ridge regression provides more stable estimation when predictors are highly correlated and all are expected to contribute to the outcome.[71] Nonparametric and robust procedures are most appropriate when distributional assumptions are clearly violated and the sample size is too small for asymptotic corrections to restore nominal Type I error rates, although researchers must accept a potential power cost when underlying distributions are approximately normal. Exact tests, including Fisher's exact test and permutation-based procedures, are particularly appropriate for categorical data with sparse cell counts when asymptotic approximations are unreliable. Table 1 below summarizes these conditional recommendations as a concise, decision-oriented reference for researchers navigating analytical choices under small-sample constraints.

Table 1 Recommendation of Statistical Techniques and Its Caution

Technique	Recommended When	Caution
Bayesian estimation	Prior information available; hierarchical data; credible interval reporting required; n < 25 with theory-informed priors	Sensitivity to prior specification; requires transparency and prior sensitivity analysis
Bootstrap resampling	No prior information; distributional uncertainty; n ≥ 20; interval estimation focus	Reduced reliability at n < 15; assumes sample representativeness
LASSO regression	High predictor-to-sample ratio; sparse true effects expected; low to moderate predictor correlation	Biased for large coefficients; unstable cross-validation with very small n
Ridge regression	High predictor-to-sample ratio; correlated predictors; variance reduction is the primary goal	Does not perform variable selection; requires tuning parameter optimization
Nonparametric/robust methods	Clear assumption violations; non-normal distributions; outlier sensitivity	Potential power cost when distributions are approximately normal
Exact tests	Categorical data; sparse cell counts; small n where asymptotic approximations fail	Computationally intensive for complex designs

Citation: Judijanto L. Difficulties and practical issues in statistical analysis in research with small sample size: a review. *Open Access J Sci.* 2026;9(1):132–144. DOI: 10.15406/oajs.2026.09.00297

The synthesis demonstrates that statistical challenges in small-sample research are not isolated technical anomalies but rather systematic structural constraints that affect inference, estimation, and generalization. While alternative analytical techniques can partially mitigate these issues, none fully eliminate the fundamental information limitations imposed by restricted sample size. The reviewed literature collectively supports the conclusion that methodological transparency, justification of power analyses, robust estimation procedures, and explicit acknowledgment of inferential uncertainty are essential safeguards when conducting statistical analyses with small samples.

Discussion

This discussion addresses the two research questions formulated in the introduction. The evidence is drawn from the 39 peer-reviewed articles included in the systematic review and is organized into two analytical segments — first, recurring statistical difficulties (RQ1), then methodological strategies (RQ2) — followed by integrative implications and directions for future research.

Addressing RQ1: Recurrent statistical difficulties

Across disciplines, the reviewed studies converge on one consistent finding: small sample sizes undermine statistical inference in four primary ways — reduced power, unstable parameter estimates, inflated error rates, and compromised generalizability. While the specific manifestations vary by research design, the underlying statistical mechanisms are consistent.

Statistical power and error inflation

Reduced power is the most consistently reported difficulty. Small samples frequently produced power estimates below the 0.80 threshold, with several simulation-based studies reporting values between 0.20 and 0.50 for moderate effects when $n < 30$.[72] Such levels imply a 50–80% probability of failing to detect true effects. Marginal increases in sample size — for example, from $n = 15$ to $n = 25$ per group — often produced disproportionate gains in detection probability, reflecting non-linear sensitivity in small-sample regimes. Underpowered designs are also associated with effect size inflation: statistically significant findings from small samples tend to overestimate true population effects. Reanalysis studies indicate that effect sizes derived from samples with $n < 40$ were, on average, 20–40% larger than subsequent replication estimates.[73] As detailed in the Results section, inflation and replication instability form a compounding inferential cascade — further elaboration is deferred to avoid redundancy.

Estimation instability

Regression coefficients, odds ratios, and correlation estimates exhibited wide confidence intervals under small-sample conditions.[32] When the observation-to-parameter ratio fell below 10:1, the signs and magnitudes of the coefficients varied substantially across bootstrap resamples.[74] Logistic regression models were especially sensitive, with convergence failures and inflated standard errors appearing when event-per-variable ratios fell below recommended thresholds. Notably, instability was not confined to parametric methods: even distribution-free nonparametric procedures exhibited reduced precision due to limited information about the underlying population distribution.[75]

Assumption violations and error distortion

Many inferential techniques rely on asymptotic properties that weaken with small sample sizes, increasing susceptibility to Type I error inflation or deflation depending on the distributional shape.[9] Under moderate non-normality (skewness > 1), t-tests with $n < 20$ exhibited Type I error rates of 0.07-0.12; ANOVA with heteroscedasticity and unequal group sizes produced distorted p-values below $n = 25$ per cell. These distortions were most pronounced in complex models — structural equation modeling and multilevel modeling — where insufficient sample size produced non-convergence, inadmissible solutions, and biased fit indices.[76]

Overfitting and generalizability

In predictive contexts, apparent model accuracy declined by 15–30% when evaluated on holdout data under small-n conditions.[77] This reduced external validity was especially relevant in clinical research and educational studies, where logistical constraints frequently limit recruitment and findings may reflect contextual particularities rather than population-level patterns.[78] *Publication bias.* Several reviewed articles indicate that small-sample studies are more likely to report significant findings due to selective reporting or threshold-based publication practices. This creates evidential asymmetry in which null findings remain unpublished, reinforcing inflated effect estimates in the cumulative literature. Meta-analytic re-evaluations have identified asymmetrical funnel plots associated with small-sample dominance in several domains.[79]

Addressing RQ2: Methodological strategies for strengthening validity under small-sample conditions

The second research question concerns strategies demonstrating empirical support for improving validity, stability, and interpretability in small-sample research. The synthesis identified five major methodological responses: resampling techniques, Bayesian approaches, robust estimators, small-sample corrections, and design-oriented adaptations.

Resampling and bootstrap procedures

Bootstrap techniques were among the most frequently supported strategies. Across simulation studies, bootstrap confidence intervals demonstrated higher coverage probability than conventional asymptotic intervals when $n < 30$.[80] Bias-corrected and accelerated (BCa) intervals, in particular, reduced skew-induced distortions and provided more stable parameter estimates under non-normality. In regression contexts, bootstrap standard errors mitigated variance underestimation and improved interval reliability. However, the effectiveness of bootstrapping depended on sample representativeness; extremely small samples (e.g., $n < 15$) limited the diversity of resampling, reducing bootstrap reliability. Thus, while beneficial, resampling does not fully compensate for extremely constrained data structures.

Bayesian estimation and informative priors

Bayesian methods demonstrated strong support for small-sample inference. By incorporating prior information, Bayesian models reduced parameter uncertainty and stabilized estimation when data alone were insufficient.[81] Studies comparing frequentist and Bayesian estimators showed narrower credible intervals and improved convergence properties in hierarchical models under $n < 30$ conditions. Importantly, empirical validation studies indicated that Bayesian credible intervals maintained nominal coverage rates more consistently than classical confidence intervals when distributional assumptions were violated.[82] However, sensitivity to prior specification was emphasized; poorly justified priors could introduce bias, underscoring the need for transparency and sensitivity analysis.

Robust and nonparametric estimators

Robust statistical procedures, including trimmed means and heteroscedasticity-consistent standard errors, demonstrated improved error control in the presence of small-sample heterogeneity.[83] While not eliminating power limitations, these approaches reduced vulnerability to extreme values and assumption violations. Nonparametric methods provided advantages in distribution-free contexts but were associated with reduced power compared to parametric alternatives when normality assumptions were approximately satisfied.[84] Therefore, methodological selection must consider trade-offs between robustness and efficiency.

Small-sample corrections and adjusted test statistics

Several studies highlighted small-sample correction techniques such as adjusted degrees of freedom (e.g., Welch–Satterthwaite approximation) and bias-corrected variance estimators. These adjustments improved Type I error calibration in unequal variance scenarios. In multilevel modeling, restricted maximum likelihood (REML) estimation outperformed full maximum likelihood when cluster counts were limited, yielding less biased variance component estimates.[85] Such corrections, while technical, demonstrated measurable improvements in inferential reasoning.

Design-oriented adaptations and transparency

Beyond statistical adjustments, design-level strategies were strongly emphasized. Pre-registration, effect size reporting, and confidence interval interpretation were identified as practices that enhance transparency and interpretability under power limitations. Simulation-based power analysis before data collection was repeatedly recommended to anticipate feasible inference boundaries.[86] Cross-validation and penalized regression (e.g., ridge or lasso) were shown to mitigate overfitting in predictive contexts. These approaches reduced parameter variance and improved out-of-sample stability, particularly when predictor counts approached sample size. A practical decision framework for technique selection can be derived from the convergent evidence in this review. When a researcher has access to theoretically grounded prior information—for example, from previous clinical trials, established meta-analyses, or expert elicitation—Bayesian estimation provides the most principled approach to inference under small-sample conditions, as it directly incorporates that prior knowledge to stabilize estimates and maintain credible interval coverage.[70] Sensitivity analyses that vary prior specifications are, however, essential to ensure that posterior conclusions are not unduly driven by prior assumptions rather than by observed data. When prior information is unavailable or difficult to justify, bootstrap resampling offers a distribution-free alternative that approximates sampling variability without parametric assumptions. Still, its reliability deteriorates rapidly when n falls below approximately 15–20 observations, at which point the resampled draws may not adequately represent population variability. In the presence of multiple candidate predictors relative to sample size—a situation common in exploratory behavioral or clinical research—penalized regression is the preferred strategy: ridge regression reduces estimation variance when predictors are correlated. At the same time, LASSO simultaneously performs variable selection by shrinking some coefficients to zero, making it more interpretable in contexts with anticipated predictor sparsity.[71] For researchers confronting clear violations of distributional assumptions with categorical or ordinal outcomes and very small cell frequencies, exact tests and nonparametric procedures provide the most reliable Type I error control, at the cost of potentially reduced power when distributional conditions are approximately met. This conditional mapping of analytical strategies to research contexts strengthens the practical utility of this review as a decision-support reference for applied researchers facing small-sample analytical constraints across disciplines.

Study-design-specific recommendations

The methodological strategies identified in this review carry different implications depending on the study design employed. Rather than applying corrective techniques uniformly, researchers should align analytical choices with the structural characteristics of their particular design.

In **randomized controlled trials (RCTs) and small-sample experimental designs**, the primary concern is inadequate statistical power. Here, simulation-based a priori power analysis is essential before data collection, and Bayesian sequential designs — which allow updating evidence as data accumulate — are especially useful when recruitment is ongoing, and total sample size is uncertain. Pre-registration of analysis plans further guards against analytical flexibility inflating the Type I error rate.[87,88]

In **observational and cross-sectional designs**, small sample sizes are frequently associated with covariate imbalance and confounding. Propensity score methods and robust regression estimators (e.g., heteroscedasticity-consistent standard errors) provide the most defensible approach. When the number of confounders approaches the sample size, penalized regression (LASSO or ridge) should replace ordinary least squares to avoid overfitting.[15]

In **longitudinal and repeated-measures designs**, the main risks are unstable variance component estimates and convergence failure in mixed-effects models when the number of participants is small. Restricted maximum likelihood (REML) estimation and parsimonious random-effects specifications (e.g., random intercepts only, rather than full random-slope models) are recommended. Where cluster sizes fall below 10 units, bootstrap confidence intervals for fixed effects provide more reliable inference than asymptotic approximations.[89]

In **single-case and n-of-1 designs**, the absence of between-subject variance requires design-specific effect size metrics (e.g., non-overlap indices, Tau-U) and Bayesian rate-ratio approaches rather than conventional inferential tests based on group-level sampling distributions.[90] Across all designs, the non-negotiable baseline practices remain the same: pre-specified analysis plans, transparent reporting of effect sizes with confidence intervals, and explicit acknowledgment of the inferential limits imposed by sample size constraints.[91]

Integrative interpretation of findings

Taken together, the evidence suggests that no single methodological strategy fully resolves small-sample limitations. Instead, improvements emerge through integrated application: combining robust estimation, resampling validation, and transparent reporting practices. Bayesian and bootstrap approaches show particularly strong empirical support for stabilizing inference, yet both require careful implementation and explicit methodological justification.

The review also indicates that methodological sophistication cannot substitute for adequate design planning. While statistical remedies mitigate distortions, fundamental sampling constraints continue to influence the boundaries of inference. Therefore, interpretative caution remains essential.

The findings carry both methodological and epistemological implications. Methodologically, small-sample researchers should prioritize resampling-based inference, robust estimators, and — where

Citation: Judijanto L. Difficulties and practical issues in statistical analysis in research with small sample size: a review. *Open Access J Sci.* 2026;9(1):132‒144. DOI: 10.15406/oajs.2026.09.00297

theoretically justified — Bayesian frameworks. Transparent reporting of effect sizes, uncertainty intervals, and assumption diagnostics should become standard practice.

Epistemologically, the persistence of inflated effects and unstable estimates underscores the need for replication and cumulative synthesis. Journals and reviewers should encourage interval-based interpretation rather than reliance on binary significance thresholds. Future research should further investigate threshold conditions under which specific methodological adaptations yield optimal performance. Comparative simulation studies that explore hybrid strategies, such as Bayesian bootstrapping or penalized hierarchical modeling, may provide refined guidance for applied researchers. Additionally, meta-research evaluating reporting quality in small-sample publications could illuminate systemic biases influencing evidential reliability. A small sample size introduces systematic inferential vulnerabilities, yet empirically supported methodological strategies can substantially improve validity and interpretability when applied thoughtfully. The challenge is not merely technical but also conceptual: aligning analytical rigor with realistic design constraints to preserve scientific credibility under limited-data conditions.

Conclusion

This systematic review synthesized evidence from 39 peer-reviewed articles indexed in Scopus (2020–2025) to consolidate current knowledge on statistical vulnerabilities and methodological responses in empirical studies with small sample sizes. The findings demonstrate that a small sample size consistently produces interconnected inferential challenges that affect the validity, stability, and interpretability of statistical results across disciplines.

First, the most consistently reported statistical difficulty is reduced statistical power. Across experimental, observational, and multivariate designs, power frequently falls below conventional adequacy thresholds when sample sizes are limited, increasing the probability of Type II error and reducing sensitivity to detect true effects. This limitation is compounded by effect-size inflation, in which statistically significant findings from small samples tend to overestimate population parameters due to stochastic variability. The review further indicates that parameter estimation becomes unstable under small-n conditions, as reflected in wide confidence intervals, fluctuating coefficient magnitudes, and convergence problems in regression, logistic, multilevel, and structural equation models.

Second, small samples amplify sensitivity to violations of assumptions. Inferential procedures relying on asymptotic properties demonstrate distorted Type I error rates when normality and homoscedasticity assumptions are not satisfied. Such distortions are particularly evident in models with complex parameter structures. Additionally, predictive modeling with limited observations increases the likelihood of overfitting, leading to reduced out-of-sample accuracy and reduced generalizability. Publication bias further intensifies the evidential imbalance, as statistically significant findings from small studies are more likely to be published than null findings, thereby reinforcing inflated effect estimates. Collectively, these patterns indicate that small sample size affects inference through five dominant mechanisms: diminished power, inflated or unstable effect estimation, assumption-sensitive error distortion, model overfitting, and restricted external validity. These mechanisms are not isolated phenomena but interact to produce compounded inferential vulnerability.

Regarding methodological responses, the review identifies several strategies that demonstrate empirical support for improving statistical performance under small-sample conditions. Resampling techniques, particularly bootstrap procedures, enhance the accuracy of interval estimation and reduce bias under non-normal distributions, though their effectiveness declines as sample sizes become extremely small. Bayesian estimation frameworks show strong potential to stabilize parameter estimates by incorporating prior information, improve convergence properties, and maintain credible interval coverage under limited-data scenarios. Robust estimators and heteroscedasticity-consistent adjustments improve error calibration when classical assumptions are violated. Small-sample corrections, including adjusted degrees of freedom and restricted likelihood approaches, reduce bias in variance estimation and hierarchical modeling contexts.

Beyond analytic modifications, design-oriented practices such as pre-analysis planning, transparent reporting of effect sizes, interpretation of uncertainty intervals, and simulation-based power assessment substantially enhance interpretability and mitigate overconfidence in limited-data findings. However, no single method eliminates all small-sample limitations. The strongest empirical support emerges from integrated application of resampling, robust estimation, and transparent inferential reporting, particularly when complemented by theoretically justified Bayesian modeling. Overall, small-sample research does not inherently invalidate empirical inquiry, but it demands heightened methodological rigor and interpretative caution. Statistical adjustments can improve reliability and reduce distortion, yet they cannot fully replace adequate sampling designs. The evidence indicates that analytical sophistication must be accompanied by explicit acknowledgment of inferential boundaries. Strengthening reporting standards, encouraging replication, and promoting the synthesis of cumulative evidence remain essential to maintaining credibility in domains where sample-size constraints are unavoidable.

In sum, contemporary empirical research consistently documents identifiable statistical vulnerabilities associated with small samples, while also providing empirically supported methodological pathways that enhance validity, stability, and interpretability. The challenge lies not in avoiding small-sample research altogether, but in aligning analytical decisions with the structural realities of limited data to ensure responsible and defensible scientific conclusions.

Acknowledgments

None.

Conflicts of interest

The author declares there is no conflict of interest.

References

1. Joe Iwanaga, Kyoichi Obata, Tomotaka Kato, et al. What Is the Appropriate Sample Size in Human Cadaveric Studies? A Quantitative Review of 770 Articles. *Clin Anat*. 2026;39(1):14–19.
2. Kawka M, Dawidziuk A, Jiao LR, et al. Artificial intelligence in the detection, characterisation and prediction of hepatocellular carcinoma: a narrative review. Transl Gastroenterol Hepatol. 2022;7:41.
3. Trejo F, Torres-Escobar R, Ochoa A. Application of the Sugeno-Takagi-Kang Fuzzy Logic Model as a Decision Support System with a Mathematical Approach for Managing Uncertainty in Supply Chain for the Demand Forecasting: Proposal of a Fuzzy Time Series Forecasting Model (FTS). *Signals and Communication Technology*. 2026;F1428:283–298.
4. Aslıhan Özdemir, Fakhriddin Sarzhanov, Funda Doğruman-Al et al. Exploring the complex interplay of Blastocystis, morbid obesity, and bariatric surgery on gut microbial dynamics. *Microb Pathog*. 2025;200:107306.

5. Wei Y, Takuro Kobayashi, Yan Lu et al. Comparison of Survival Outcomes and Risk Factors Between Ductal Carcinoma of the Prostate and Acinar Adenocarcinoma of the Prostate: A Population-based Propensity Score–matching Study. *Eur Urol Open Sci*. 2022;46:88–95.

6. Leimin Tian, Tina LY, Nicole L, et al. Advanced Statistical Analysis Methods in Quantitative HRI Research," in Experimental Methodology for Human–Robot Interaction: Guidelines and Case Studies for Human-Centred and Ethical Robotics Research. 2025:119–132.

7. Chaturvedi AK, Singh R, Tewari CK. A qualitative analysis on dairy farming policies and its stakeholder's perspective in Indian dairy sector. *Model Assist Stat Appl*. 2024;19(3):293–299.

8. Rodrigues FB, Ferreira JJ. Strategies to minimize placebo effects in research investigations. *International Review of Neurobiology*. 2020;153:49–70.

9. Ngibe M, Lekhanya LM. Critical factors influencing innovative leadership in attaining business innovation: A case of manufacturing SMEs in KwaZulu-Natal. *J Leg Ethical Regul*. 2019;23(2).

10. Hernández-Torres R, Mary Rodríguez-Rabassa, Lianel P Rosario et al. Investigation of Psychometric Properties and Correlation with Psychological Distress after Hurricane Hazards in Puerto Rico. *Int J Environ Res Public Health*. 2024; 21(10):1267.

11. Milovanović S, Bogdanović Z, Labus A, et al. Social recruiting: an application of social network analysis for preselection of candidates. *Data Technol Appl*. 2022;56(4):36–557.

12. Ding P, Tan L, Pan H, et al. The Lack of Neurofeedback Training Regulation Guidance and Process Evaluation May be a Source of Controversy in Post-Traumatic Stress Disorder-Neurofeedback Research: A Systematic Review and Statistical Analysis. Brain Connect. 2025.

13. Wang P. Phase objectives analysis for PM2.5 reduction using dynamics forecasting approach under different scenarios of PGDP decline. *Ecol Indic*. 2021;129.

14. Ouyang Y, Avila ML, Heath A. Design and analysis of individually randomized multiple baseline factorial trials. *Behav Res Methods*. 2026;58(30).

15. Hripcsak G, Linying Zhang, Yong Chen, et al. Assessing Covariate Balance With Small Sample Sizes. *Stat Med*. 2025;44:18–19.

16. Muttanahally KS, Yadav S, Freilich MA, et al. Does the outcome of graft materials at dental implant sites differ between patients with normal and compromised bone health. *J Oral Implantol*. 2024;50(3):238–244.

17. Fan Y, Fan J, Li A, et al. Physics-Informed Neural Network Surrogate Model for Small-Sample Mistuned Bladed Disk. *AIAA J*. 2025;63(6):2520–2531.

18. Rutkowski S, Czech O, Wrzeciono A, et al. Virtual reality as a chemotherapy support in treatment of anxiety and fatigue in patients with cancer: A systematic review and meta-analysis and future research directions. *Complement Ther Med*. 2021;61:102767.

19. Witman A, Christopher Beadles, Yiyan Liu et al. Comparison group selection in the presence of rolling entry for health services research: Rolling entry matching. *Health Serv Res*. 2019;54(2):492–501.

20. He L, Li X, Kong X, et al. Characteristics, Classification and Challenges in Searching for CEMP Stars. *Res Astron Astrophys*. 2025;25(5).

21. Holtzhausen MM, Botha P. Combining interventions: an innovative leadership development program. *J Manag Dev*. 2021;40(3):240–252.

22. Khasawneh M, Al Najdawi BM, Shatnawi HS, et al. Corporate Social Responsibility, Knowledge of COVID-19, and Risk Perception in International Tourism Enterprises. *J Environ Manag Tour*. 2022;13(6):1574–1583.

23. Feld J, Faldum A, Schmidt R. Adaptive group sequential survival comparisons based on log-rank and pointwise test statistics. *Stat Methods Med Res*. 2021;30(12):2562–2581.

24. Tabari H. Statistical analysis and stochastic modelling of hydrological extremes. *Water (Switzerland)*. 2019;11(9).

25. Johannessen KN, Bakken AK, Hambrick EP, et al. Treatment Effects of Introducing the Neurosequential Model of Therapeutics in a Norwegian Residential Treatment Facility for Children Aged 7–13. *Children*. 2024;11(5):503.

26. Khan KM, Bielko SL, McCullagh MC. Efficacy of hearing conservation education programs for youth and young adults: A systematic review. *BMC Public Health*. 2018;18.

27. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta□analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Res Synth. Methods*. 2020;11(2):181–217.

28. Martinou E, Angelidi A. The role of open research in improving the standards of evidence synthesis: current challenges and potential solutions in systematic reviews. *F1000Research*. 2022;11:1435.

29. Gan W. Impact of Sample Size and Its Estimation in Medical Research. *AJPM Focus*. 2025:100451.

30. Kilty TJ, Kilty KT, Burrows Borowczak AC, et al. Quantitative Techniques with Small Sample Sizes: An Educational Summer Camp Example. *Probl Educ 21st Century*. 2024;82(4):507–520.

31. Kim BY, Kim HY, Lee DH, et al. Effects of exercise interventions on physical recovery in paediatric cancer patients and survivors: a meta-analysis. *BMC Pediatr*. 2025;25(1).

32. Zheng Q, Wang C, Zhu Z. Research on the prediction of mine water inrush disasters based on multi-factor spatial game reconstruction. *Geomech Geophys Geo-Energy Geo-Resources*. 2024;10(41).

33. Dick AS, Jonathan S Comeret, Mohammadreza Bayat, et al. Meaningful Associations Redux: Quantifying and interpreting effect size in the context of the Adolescent Brain and Cognitive Development study. *Dev Cogn Neurosci*. 2025;76:101630.

34. Nugraha HK, Rasmussen AP, Mulford KL, et al. AI in Pediatric Spine Care: Clinical, Research, and Ethical Considerations. J Clin Med. 2025;14(22):8115.

35. Appels S, Van Viersen S, Van Erp S, et al. A scoping review on word-reading resilience in literacy: Evaluating empirical evidence for protective factors. *Learn Instr*. 2024;93:101969.

36. McCook L, Del Vecchio ML, Crosland K. Using Progressive Muscle Relaxation to Increase Academic Engagement for Elementary School Students with Autism Spectrum Disorder. *Behav Sci (Basel)*. 2025;15(11):16.

37. Eranti P, Megha Hegde, Syed Muktadir Al Sium, et al. Impact of the COVID-19 pandemic on computational biology early career researchers: A global retrospective study. *PLOS Comput Biol*. 2025;21(10):e1013554.

38. Mesquida C, Murphy J, Lakens D, et al. Publication bias, statistical power and reporting practices in the Journal of Sports Sciences: potential barriers to replicability. J Sports Sci. 2023;41(16):1507–1517.

39. Fernández-Corbacho A, Cores-Bilbao E, Flor-Arasil P. Ethnocultural empathy development of future language teachers through digital multiliteracy resources for low-literacy adult migrants. *Front Psychol*. 2024;15:1398457.

40. Tomassini L, Giulia Ricchezze, Cristiana Gambelunghe, et al. Immunohistochemical insights into hyperthermia-related deaths: a systematic review. *Int J Legal Med*. 2025;139(5):2467–2501.

41. Mulyono S. Improving or worsening?: the development and evaluation of a VR-based psychotherapy to bullying victims and perpetrators in school adolescents. *J Ners*. 2025;20(3):322–331.

42. Rogala J. Enhancing autism spectrum disorder classification in children through the integration of traditional statistics and classical machine learning techniques in EEG analysis. *Sci Rep*. 2023;13(1):21748.

43. Majebele JJ, Yang M, Mateen M, et al. Mechanization and Maize Productivity in Tanzania's Ruvuma Region: A Python-Based Analysis on Adoption and Yield Impact. *Agric*. 2025;15(13):1412.

44. Geng X, Yang Z, Jiao L, et al. Association rule-based classification: A comprehensive review of methodologies and applications. *Expert Syst Appl*. 2025;280.

45. Varnado OJ. Monthly Headaches and Severity in Patients on Galcanezumab or Traditional Preventive Migraine Medication: A 24-Month Claims and Electronic Health Records Study. *Neurol Ther*. 2025;14(3):911–925.

46. Velasco-Albendea FJ. The Contribution of Mediastinal Transbronchial Nodal Cryobiopsy to Morpho-Histological and Molecular Diagnosis. *Diagnostics*. 2023;13(22):3476.

47. Srouji T, Lanot A, Béchade A, et al. Analysis of ultrafiltration volume during long dwell with icodextrin in automated peritoneal dialysis. *Bull la Dial a Domic*. 2025;8(1):15–27.

48. Zhu M, Qu L. Research on the Application Effect of Metalearning Method Based on Prototype Network in Corrective Measures for Member State Audit. *J Comb Math Comb Comput*. 2023;117:217–227.

49. Agebjörn A, Walldén R. Second-language writing development of adult immigrants following three study paths reflecting their varying levels of prior education. *J Second Lang Writ*. 2025;67.

50. Egdell V, Maclean G, Raeside R, et al. Workplace preparedness for an ageing workforce: a case study. *Int J Sociol Soc Policy*. 2022;42(9–10):890–907.

51. Zubaidah T, Syarifudin A, Haris A, et al. The Nexus of Home Sanitation, Personal Hygiene, and Dermatitis: Evidence from Banjarmasin, Indonesia. *Univers J Public Heal*. 2025;13(2):511–518.

52. Tao L, Liu H, Ning G, et al. LM-based framework for bearing fault diagnosis. *Mech Syst Signal Process*. 2025;224.

53. Gajwani R, Wilson N, Nelson R, et al. Recruiting and exploring vulnerabilities among young people at risk, or in the early stages of serious mental illness (borderline personality disorder and first episode psychosis). *Front Psychiatry*. 2022;13:943509.

54. Wang Z, Xing X, Mun EY, et al. The Role of Double-Zero-Event Studies in Evidence Synthesis: Evaluating Robustness Using the Fragility Index. *J Eval Clin Pract*. 2025;31(1):e14301.

55. Gonçalves AP, Silva JF, Martins RC. Instrument performance and calibration effects in physical measurements. *Meas Sci Technol.* 2022;33(9):95012.

56. Wang J, Haoran Hu, Haixia Du, et al. Clinical Efficacy Protocol of Yinhuapinggan Granules: A Randomized, Double-Blind, Parallel, and Controlled Clinical Trial Program for the Intervention of Community-Acquired Drug-Resistant Bacterial Pneumonia as a Complementary Therapy. *Front Pharmacol.* 2022;13:852604.

57. AL-Hakeem HFH, Khan M. The role of statistics in advancing nitric oxide research in plant biology: from data analysis to mechanistic insights. *Front Plant Sci*. 2025;16:1597030.

58. Liikkanen S, Tero Jalkanen, Mika Mäkinen, et al. Body movement as a digital biomarker in chronic pain rehabilitation: a real-world pilot study using virtual reality. *Front Virtual Real*. 2025;6.

59. Klapwijk ET, van den Bos W, Tamnes CK, et al. Opportunities for increased reproducibility and replicability of developmental neuroimaging. *Dev Cogn Neurosci*. 2021;47:100902.

60. Nazareno G, Montañez MCA. Assessing Public Elementary Teachers' Competence in Planning, Management, and Instruction Amid Adaptive Education during the COVID-19 Pandemic. *Int J Inf Educ Technol*. 2025;15(9):1804–1819.

61. Walker N, Vernon-Smith M, Townend M. A feasibility study of a novel work-focused relational group CBT treatment programme for moderate to severe recurrent depression. *Ment Heal Rev J*. 2021;26(4):328–352.

62. Iannuzzelli CC, Iannuzzelli AL, Cunha B, et al. Exploring the impact of methylenetetrahydrofolate reductase (MTHFR) gene variations on autism spectrum disorder severity. *J Osteopath Med*. 2025;126(3):113–117.

63. D'Amours J, Girard S, Miquelon P, et al. Bonheur en boule: an adapted group-based physical activity program for youth with disabilities. *Front Sport Act Living*. 2025;7:1580697.

64. Li P, Yang Wu, Danxia Ge, et al. Effects of Tanreqing injection against ventilator-associated pneumonia: a meta-analysis and systematic review of clinical studies. *Front Pharmacol*. 2025;16.

65. Kanda T, Jin M, Zaitsu W. Integrated ensemble of BERT- and feature-based models for authorship attribution in Japanese literary works. *Front Artif Intell*. 2025;8.

66. Alhwoaimel N. Concurrent validity of a novel wireless inertial measurement system for assessing trunk impairment in people with stroke. *Sensors (Switzerland)*. 2020;20(6):1699.

67. Hosoda-Urban T, O'Donnell E. Diabetes-related posttraumatic stress symptoms, resilience, and illness management among adolescents and young adults with type 1 diabetes: an exploratory study. *Front Psychiatry.* 2025;16:161615273.

68. Duan Y, Mathew T, Alemayehu D, et al. Novel approaches for random-effects meta-analysis of a small number of studies under normality. *Res Synth Methods*. 2025;16(6):922–936.

69. Natesan Batley P, Shukla Mehta S, Hitchcock JH. A Bayesian Rate Ratio Effect Size to Quantify Intervention Effects for Count Data in Single Case Experimental Research. *Behav Disord*. 2021;46(4):226–237.

70. Depaoli S, Winter SD, Visser M. The Importance of Prior Sensitivity Analysis in Bayesian Statistics: Demonstrations Using an Interactive Shiny App. *Front Psychol*. 2020;11:608045.

71. Pavlou M, Ambler G, Seaman S, et al. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med*. 2016;35(7):1159–1177.

72. Schwartz A, King B, Mink R, et al. The Emergence and Spread of Practice-Based Medical Education Research Networks. *Acad Med.* 2020;95(11):S12–S13.

73. Khan SM, Madison L Fennell, Mazyar Fallah, et al. Study protocol and pilot study results for a clinical intervention trial of PKU carriers and non-carriers: the Phe for Me trial. *Orphanet J Rare Dis*. 2026;21(1):18.

74. Vockley J, Barbara Burton, Agnieszka Jurecka, et al. Challenges and strategies for clinical trials in propionic and methylmalonic acidemias. *Mol Genet Metab*. 2023;139(3):107612.

75. Sono H. Reducing Readmission Rates in Heart Failure Patients Aged 60 and Older Through Education and Follow-Up Calls. *Prof Case Manag*. 2026;31(2):57–68.

76. Po J, Bookstein A. Financial toxicity and drivers of delayed care among cancer survivors across the lifespan. *J Clin Oncol*. 2025;43(16).

77. Radha Krishna Murthy G, Buhukya S, Dhyani A, et al. Data-Driven Approaches for Statistical Analysis in Human Resource Management: A Comprehensive Review and Future Directions. In 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2023, 2023. p. 888–893.

78. McGregor B, Peter H O'Donnell, Arjun Balar, et al. Health-related Quality of Life of Patients with Locally Advanced or Metastatic Urothelial Cancer Treated with Enfortumab Vedotin after Platinum and PD-1/PD-L1 Inhibitor Therapy: Results from Cohort 1 of the Phase 2 EV-201 Clinical Trial. *Eur Urol*. 2022;81(5):515–522.

79. Ibrahim IA, Rem Ehab Abdelkader, Ahmed Hosney Nada, et al. Effect of Everolimus on Prognosis of Neurofibromatosis Type 1 Lesions: A Systematic Review and Meta Analysis. *Clin Ther*. 2024;46(11):865–869.

80. Mustafa MS. Susceptibility Weighted and Perfusion Imaging for Differentiating Recurrence and Radiation Necrosis in Post-Operative Malignant Brain Tumors: A Retrospective Cohort Study. *Indian J Surg Oncol*. 2026.

81. Sabbah KA, Mahamid F, Mousa A. Role of Augmented Reality Learning in Enhancing Palestinian University Students' Motivation and Reflective Thinking. *J Soc Stud Educ Res*. 2025;16(2):252–278.

82. Arianpoor A, Abdollahi A. The interaction of accounting employability-based skills and problem-based learning. *J Appl Account Res*. 2025;1–26.

83. Magalhães R. Mixed Reality in the Operating Room: A Systematic Review. *J Med Syst*. 2024;48(1):76.

84. Göransson C, Wengström Y, Ziegert K, et al. Self-care ability and sense of security among older persons when using an app as a tool for support. *Scand J Caring Sci*. 2020;34(3):772–781.

85. Stock A, Haupt AJ, Mach ME, et al. Mapping ecological indicators of human impact with statistical and machine learning methods: Tests on the California coast. *Ecol Inform*. 2018;48:37–47.

86. Eksteen JM, Visser WI, de Wet J, et al. Cross-sectional study of acral melanoma awareness in a group of South African final phase medical students. *Indian J Dermatol Venereol Leprol*. 2022;88(3):444.

87. Siontis GCM, Nikolakopoulou A, Sweda R, et al. Estimating the sample size of sham-controlled randomized controlled trials using existing evidence. *F1000Research*. 2022;11:85.

88. Farooq H, Ali S, Shah I, et al. Adaptive Clinical Trials and Sample Size Determination in the Presence of Measurement Error and Heterogeneity. *Stats*. 2025;8(2):31.

89. Flynn PJ, Awaysheh A, Bliese PD, et al. From Intent to Impact: A Proactive Event Approach for Amplifying Sustainability Across Time. *J Manage*. 2025;51(3):999–1032.

90. Tanious R, Manolov R. A practitioner's guide to conducting and analysing embedded randomized single-case experimental designs. *Neuropsychol Rehabil*. 2023;33(4):613–645.

91. Meng W. Statistical Inference and Sampling Design: A Fully Reproducible Statistical Practice from Sampling Frameworks to Confidence Intervals. 2026. p. 53.

Citation: Judijanto L. Difficulties and practical issues in statistical analysis in research with small sample size: a review. *Open Access J Sci.* 2026;9(1):132–144. DOI: 10.15406/oajs.2026.09.00297