Revista Española
de Educación Médica

Application of biostatistical methods in health sciences research: a review of the scope of classical and contemporary scientific and methodological literature.

Aplicación de métodos bioestadísticos en investigaciones en ciencias de la salud: revisión de alcance de literatura científica y metodológica clásica y contemporánea.

Lissette Beatriz Villavicencio Cedeño [1,] liso_villa@hotmail.com, https://orcid.org/0000-0001-8475-3947 ; **Marieta del Jesús Azúa Menéndez** [2,] marieta.azua@unesum.edu.ec, https://orcid.org/0000-0002-5601-6621 ; **Caleb Isaac Chilán Santana** [2,] caleb.chilan@unesum.edu.ec, https://orcid.org/0000-0002-2832-8759 ; **Yoiler Batista Garcet** [3,] ybatista@sangregorio.edu.ec, https://orcid.org/0000-0002-7851-5763

[1] National University of Tumbes, Peru, Cristo Rey Educational Unit, Portoviejo. Manabí, Ecuador.
[2] Southern State University of Manabí, professor of the Clinical Laboratory program, Jipijapa. Manabí, Ecuador.
[3] San Gregorio University of Portoviejo, Department of Medicine. Manabí, Ecuador.
* Correspondence: liso_villa@hotmail.com

Received: 28/4/26; Accepted: 19/5/26; Published: 21/5/26

Summary

Introduction: Despite the importance of biostatistics in health sciences research, numerous shortcomings persist in the biomedical literature. **Objective** : To analyze the application of biostatistical methods in health sciences research through a comprehensive, systematic review of recent scientific literature, with an emphasis on inferential foundations, application problems, and contemporary methodological alternatives. **Methods:** A comprehensive, systematic review was conducted, guided by PRISMA-ScR. The search was primarily performed in Scopus and Web of Science, with supplementary searches in PubMed, ERIC, CINAHL, and Google Scholar. Search terms related to biostatistics, statistical methods, medical research, epidemiology, public health, and clinical research were used. 245 records were identified; after removing duplicates and applying eligibility criteria, 32 studies were included. The information was organized into three categories: theoretical foundations of statistical inference, problems and limitations in the application of biostatistical methods, and contemporary methodological approaches. **Results:** The reviewed studies showed a predominance of the frequentist approach, especially regarding the use of hypothesis tests and the p-value. Recurring errors were identified in the interpretation of statistical significance, inappropriate use of tests, poor reporting of assumptions, inadequate handling of multiple comparisons, limited consideration of confounding variables, and insufficient integration of clinically relevant data. Emerging proposals were also observed, focusing on the use of confidence intervals, effect sizes, multivariate models, Bayesian analyses, metascience, open science practices, and improvements in statistical reporting. **Conclusions:** Biostatistics remains a crucial element for the quality of health sciences research; however, its practical application presents significant gaps. Overcoming these limitations requires strengthening statistical training, promoting a more critical inferential interpretation, abandoning the exclusive dependence on the p-value, integrating measures of magnitude and precision, improving the transparency of reporting, and fostering complementary analytical approaches oriented towards the reproducibility of evidence.

Keywords: biostatistics; health sciences; statistical inference; p-value; scope review; biomedical research; Bayesian inference; reproducibility.

Resumen

Introducción: A pesar de la importancia de la bioestadística en la investigación en ciencias de la salud, en la literatura biomédica persisten múltiples deficiencias. **Objetivo**: Analizar la aplicación de métodos bioestadísticos en investigaciones en ciencias de la salud mediante una revisión de alcance con enfoque sistemático de la literatura científica reciente, con énfasis en fundamentos inferenciales, problemas de aplicación y alternativas metodológicas contemporáneas. **Métodos:** Se realizó una revisión de alcance con enfoque sistemático, orientada por PRISMA-ScR. La búsqueda se efectuó principalmente en Scopus y Web of Science, con exploración complementaria en PubMed, ERIC, CINAHL y Google Scholar. Se utilizaron ecuaciones de búsqueda relacionadas con bioestadística, métodos estadísticos, investigación médica, epidemiología, salud pública e investigación clínica. Se identificaron 245 registros; luego de eliminar duplicados y aplicar criterios de elegibilidad, se incluyeron 32 estudios. La información fue organizada en tres categorías: fundamentos teóricos de la inferencia estadística, problemas y limitaciones en la aplicación de métodos bioestadísticos, y enfoques metodológicos contemporáneos. **Resultados:** Los estudios revisados evidenciaron predominio del enfoque frecuentista, especialmente del uso de pruebas de hipótesis y del valor p. Se identificaron errores recurrentes en la interpretación de la significación estadística, uso inapropiado de pruebas, escaso reporte de supuestos, deficiente manejo de comparaciones múltiples, limitada consideración de variables de confusión e insuficiente integración de relevancia clínica. Asimismo, se observaron propuestas emergentes orientadas al uso de intervalos de confianza, tamaños del efecto, modelos multivariados, análisis bayesianos, metaciencia, prácticas de ciencia abierta y mejora del reporte estadístico. **Conclusiones:** La bioestadística sigue siendo un eje decisivo para la calidad de la investigación en ciencias de la salud; no obstante, su aplicación práctica presenta brechas importantes. Superar estas limitaciones exige fortalecer la formación estadística, promover una interpretación inferencial más crítica, abandonar la dependencia exclusiva del valor p, integrar medidas de magnitud y precisión, mejorar la transparencia del reporte y fomentar enfoques analíticos complementarios orientados a la reproducibilidad de la evidencia.

Palabras clave: bioestadística; ciencias de la salud; inferencia estadística; valor p; revisión de alcance; investigación biomédica; inferencia bayesiana; reproducibilidad.

---

1. Introduction

Health sciences research unfolds within a context marked by biological complexity, individual variability, diagnostic uncertainty, population heterogeneity, and the constant need for evidence-based decision-making. In this context, biostatistics should not be viewed as an accessory or a mechanical stage at the end of the study, but rather as a cross-cutting component of scientific reasoning. Its function begins with the formulation of the research question, continues in the definition of the methodological design, guides the selection of variables and measurement procedures, allows for the calculation of sample sizes, organizes data collection, structures the analysis, and plays a decisive role in the interpretation of the results (1).

However, the use of biostatistics in research within the biomedical scientific literature has its limitations. One of the most debated elements in health research is the use of the p-value. Historically, frequentist statistical inference established the p-value and hypothesis testing as central instruments for evaluating the compatibility of data with a null hypothesis (2). This tradition has its roots in the contributions of Fisher and the Neyman-Pearson approach, which structured decision-making by controlling for Type I and Type II errors (3). However, the p-value is often misinterpreted as the probability that the null hypothesis is true, as the probability that the result is

due to chance, or even as a direct measure of clinical significance (4). These interpretations have been challenged by methodological statements that recommend using the p-value cautiously and always within the context of the design, the data, and the available evidence (5). Furthermore, it has been noted that p-values, confidence intervals, and statistical power are frequently misinterpreted in biomedical research (6).

Over-reliance on the $p < 0.05$ threshold has fostered a culture of dichotomous decision-making: significant/not significant, useful/not useful, publishable/not publishable. This logic can impoverish scientific interpretation by reducing complex phenomena to an arbitrary boundary. Furthermore, it can encourage practices such as repeating analyses until significance is achieved, selectively choosing results, and underreporting negative analyses. In health sciences, where decisions can impact diagnoses, treatments, and health policies, this simplification presents a methodological and ethical challenge (7).

Another common problem is the omission of complementary measures such as confidence intervals and effect sizes. While the p-value indicates the compatibility between the data and a hypothesis under specific assumptions, confidence intervals provide information about the precision of the estimate, and effect sizes allow us to assess the magnitude of the observed difference or association (8-9). In clinical research, a difference can be statistically significant and, at the same time, clinically irrelevant. The opposite can also occur: a non-significant result may be potentially important if the study lacks sufficient statistical power. Therefore, rigorous interpretation requires integrating statistical significance, effect size, precision, biological plausibility, consistency with previous evidence, and clinical relevance (10).

Inappropriate selection of statistical tests is another persistent limitation. In health research, errors such as applying parametric tests without verifying assumptions, using tests for independent samples when the data are paired, analyzing ordinal variables as if they were continuous without justification, employing multiple comparisons without adjustments, ignoring the distribution of the data, failing to control for confounding variables, not evaluating the goodness of fit of the proposed statistical models, or interpreting bivariate associations as if they were causal relationships are observed. These errors are not always visible to the non-specialist reader, but they can substantially alter the conclusions (1).

Scientific reproducibility has also become a critical issue. Lack of transparency in analysis, limited data availability, absence of protocols, flexible use of analytical decisions, and publication bias have all contributed to the inability to replicate many results (11-13). In this context, biostatistics is directly related to metascience, open science, and reporting standards. It is not simply a matter of applying more advanced methods, but of reporting them clearly, justifying analytical decisions, and enabling other researchers to understand, evaluate, and, where possible, reproduce the analysis (14-15).

In response to these limitations, methodological proposals have emerged aimed at improving data interpretation. These include the increased use of confidence intervals, effect sizes, sensitivity analysis, multivariate models, Bayesian approaches, likelihood-based estimation, multiple comparison corrections, study pre-registration, publication of protocols, availability of open databases and analytical codes, and specific reporting guidelines for different designs. Bayesian inference, in particular, has gained interest because it allows for the integration of prior knowledge with observed data and the expression of evidence in a probabilistic manner (16); the Bayesian factor has been proposed as an alternative or complement to the p-value, allowing for the direct comparison of relative evidence between hypotheses (17). Bayesian models offer tools for modeling uncertainty and analyzing complex data in biomedical and clinical research (18-19).

Consequently, it is necessary to review the available evidence on the application of biostatistical methods in health sciences research, identifying patterns of use, recurring limitations, and contemporary methodological alternatives to recognize methodological deficiencies, guide training processes, strengthen reporting standards, and contribute to more rigorous, transparent, and useful research for decision-making. This study aimed to analyze the application of biostatistical methods in health sciences research through a comprehensive review guided by systematic procedures for searching, selecting, and synthesizing classic and contemporary scientific and methodological literature, with the purpose of contributing to strengthening methodological rigor, the quality of scientific evidence, and evidence-based decision-making.

2. Methods

2.1 Study design

A scope review was conducted guided by systematic procedures of searching, selecting, extracting and synthesizing classical and contemporary scientific and methodological literature (20-21). The methodological process was guided by the PRISMA extension for scoping reviews, known as PRISMA-ScR (22). This framework was used because it provides criteria for transparently reporting the identification, selection, extraction, and synthesis of evidence sources in scoping reviews. Although the study employed systematic search, screening, and synthesis procedures, it was not presented as a classic systematic intervention review because the objective was mapping and descriptive rather than evaluative of effectiveness (22).

2.2 Review Question

The guiding question for this review was: How are biostatistical methods applied, interpreted, and reported in health sciences research according to both classic and contemporary scientific and methodological literature? From this general question, three specific questions were derived: What are the main theoretical and inferential foundations that support the application of biostatistics in health research? What problems, errors, or limitations are most frequently reported in the use of biostatistical methods? What contemporary methodological approaches are proposed to improve the interpretation, transparency, and reproducibility of statistical analyses?

2.3 Sources of information

The literature search was primarily conducted in Scopus and Web of Science, as these are multidisciplinary databases with broad coverage of peer-reviewed scientific literature. To broaden the search's sensitivity, PubMed, ERIC, CINAHL, and Google Scholar were also explored in April 2026. These complementary sources allowed for the verification of relevant publications in public health, health sciences education, clinical research, epidemiology, and scientific methodology.

2.4 Search Strategy

The search strategy combined terms related to biostatistics, statistical methods, and health research. The main equation was:

ALL = (“biostatistics” OR “statistical methods”) AND (“health sciences” OR “medical research” OR “epidemiology”).

Subsequently, a second iteration was performed with expanded terms:

(“biostatistics” OR “statistical analysis” OR “epidemiological methods”) AND (“health sciences” OR “clinical research” OR “public health”).

The combination of terms was adjusted according to the specific characteristics of each database. Publications in Spanish and English were considered, with an emphasis on relevant scientific and methodological literature, both classic and contemporary, and methodological documents of high conceptual relevance. The search was not limited exclusively to empirical studies, as the objective of the review included theoretical foundations, inferential debates, and methodological proposals.

2.5 Inclusion criteria

Studies and scientific documents that met the following criteria were included:

- Methodological publications related to the application, interpretation or justification of biostatistical methods in health sciences.
- Empirical studies, reviews, methodological essays, theoretical articles or consensus documents related to statistical inference, p-value, hypothesis testing, statistical models, reproducibility, Bayesian methods or quality of statistical reporting.
- Works published in scientific journals, recognized academic documents, or widely accepted methodological sources.
- Studies with direct relevance to clinical, epidemiological, biomedical, public health research or health sciences education.
- Publications in Spanish or English.

2.6 Exclusion criteria

The following were excluded:

- Publications with no explicit relation to biostatistical methods or statistical inference.
- Studies focused exclusively on mathematical developments with no applicability in health sciences.
- Opinion documents without sufficient methodological support.
- Duplicate jobs.
- Publications whose full text was not available.
- Studies with insufficient information to categorize their contributions.
- Letters to the editor, editorials without methodological development, or documents without relevant conceptual contribution to the objectives of the review.

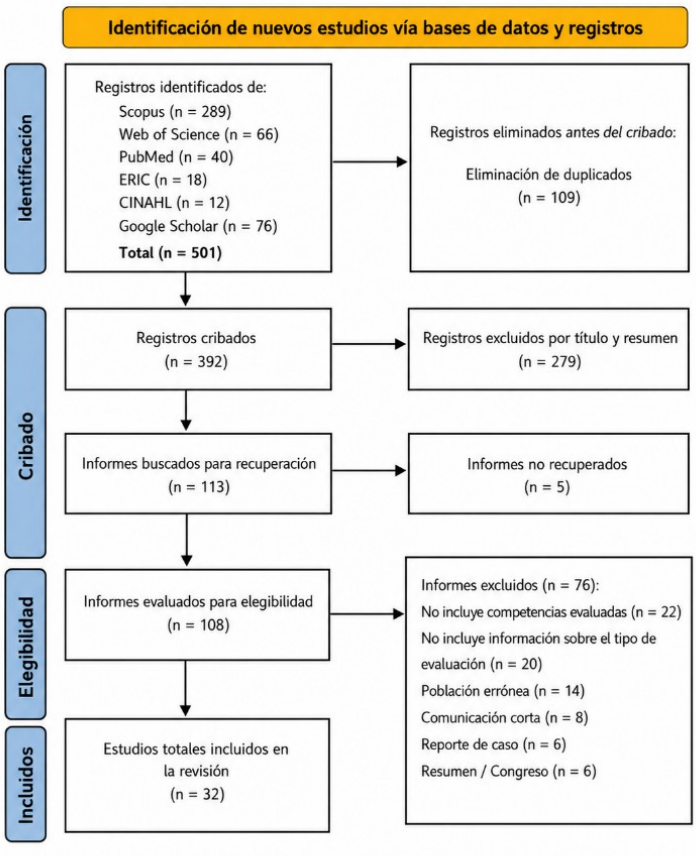

Figura 1. Diagrama de flujo del proceso de selección de estudios según PRISMA-ScR.

The initial search identified 245 records. After removing 35 duplicates, 210 records remained for title and abstract review. In this phase, 130 studies were excluded for not meeting the inclusion criteria, primarily due to a lack of relevance to health sciences, absence of biostatistical discussion,

or an exclusively technical focus without biomedical application. Subsequently, 80 full-text articles were evaluated. Of these, 48 were excluded due to methodological deficiencies for the purposes of the review, lack of substantive biostatistical analysis, or lack of alignment with the study objective. Finally, 32 studies were included in the synthesis (figure 1).

2.7 Extraction and organization of information

Information from the included studies was extracted using a matrix designed to record: author, year, type of publication, objective, area of application, statistical method or problem analyzed, main findings, and contribution to the review. Subsequently, the studies were grouped into three thematic categories:

- Theoretical foundations of statistical inference.
- Problems and limitations in the application of biostatistical methods.
- Contemporary methodological approaches and analytical alternatives.

This categorization allowed for the interpretive organization of the evidence and facilitated the integration of heterogeneous findings. A meta-analysis was not performed due to the conceptual, methodological, and documentary diversity of the sources included using Zotero.

2.8 Synthesis of the evidence

The synthesis was conducted in a narrative and thematic manner. Recurring patterns, conceptual convergences, methodological tensions, and emerging proposals were identified. The information was presented in summary tables that group the studies according to their main contribution. This procedure allowed for the integration of historical, methodological, and applied literature without forcing a homogeneity that did not exist among documents of different natures.

2.9 Ethical considerations

Because this was a literature review based on published sources, the study did not require informed consent or approval from an ethics committee. No individual data from human participants were used. The principles of academic integrity, proper citation of sources, and responsible use of scientific evidence were respected.

3. Results

3.1 General characteristics of the included studies

The 32 included studies addressed the application of biostatistical methods in health sciences from theoretical, methodological, and critical perspectives. Based on the main content of each document, the studies were organized into three thematic categories: theoretical foundations of statistical inference, problems and limitations in the application of biostatistical methods, and contemporary methodological approaches or analytical alternatives. The first category grouped 11 studies related to the theoretical foundations of statistical inference. The second category included 13 studies focused on errors, limitations, or common problems in the application of biostatistical methods. The third category comprised 8 studies oriented toward contemporary methodological proposals, including Bayesian inference, Bayes' factor, open science, reproducibility, and improvement of statistical reporting. All selected articles are shown in Table 1.

3.2 Category 1. Theoretical foundations of statistical inference

This category included 11 documents that provided conceptual, historical, and epistemological foundations for understanding statistical inference applied to health research. The selected studies addressed the development of frequentist inference, significance testing, hypothesis testing, control

of Type I and Type II errors, probability, the logic of scientific research, modern epidemiology, and interpretation based on estimates. The classic works of Fisher and Neyman-Pearson were related to the historical development of the frequentist approach and hypothesis testing (2-3). Other texts, such as those by Jeffreys, Popper, and Hacking, provided a probabilistic and epistemological perspective on the construction of scientific knowledge (16, 23-24). Likewise, the texts by Altman, Rothman, Greenland, and Lash, and Szklo and Nieto contributed from the fields of applied medical statistics and epidemiology (1, 25-26). Finally, the contributions of Cumming, Cohen and Gigerenzer made it possible to identify criticisms of the ritualized use of statistical significance and proposals for interpretation based on estimates (8, 27-28).

3.3 Category 2. Problems and limitations in the application of biostatistical methods

This category included 13 studies that addressed errors and limitations in the use of statistical methods in biomedical, clinical, epidemiological, and public health research. The problems identified were mainly related to the misinterpretation of the p-value, dependence on the $p < 0.05$ threshold, low statistical power, p-hacking, false positives, multiple comparisons, the difference between statistical and clinical significance, and the inappropriate use of measurement or analysis procedures. The reviewed studies described problems linked to the use of significance tests as an isolated decision criterion (29), the misinterpretation of the p-value of other inferential indicators (4-6), the increased risk of false-positive findings and low reproducibility (11, 13, 30-31), analytical multiplicity and multiple outcomes (32-33), errors in studies comparing measurement methods (34), and the need to differentiate between statistically significant and clinically relevant results (10).

3.3 Category 3. Contemporary methodological approaches and analytical alternatives

This category included eight studies that addressed analytical alternatives and methodological proposals aimed at improving the interpretation, transparency, and reproducibility of results in health sciences. The approaches identified included the Bayesian factor, Bayesian inference, Bayesian data analysis, moving beyond the dichotomous use of the p-value, effect size reporting, open research culture, reproducible science, and the PRISMA-ScR guidelines. These studies contributed strategies to complement or overcome interpretation based solely on the p-value (7). They also highlighted the importance of reporting effect magnitudes (9), promoting open science practices, making analytical procedures transparent, and strengthening reproducibility (14-15), as well as using methodological reporting guidelines when conducting scoping reviews or evidence synthesis studies (22).

4. Discussion

This design was selected because the primary purpose was not to estimate a pooled effect or answer a narrow clinical question through meta-analysis, but rather to map the scientific output related to the application of biostatistical methods in health sciences research, identify thematic categories, describe methodological trends, and recognize knowledge gaps. Scope reviews are useful when the area of study is broad, heterogeneous, or conceptually diverse, as is the case with applied biostatistics, where theoretical foundations, analytical practices, interpretation problems, reporting standards, and emerging approaches converge (20-21). The results of this review show that biostatistics occupies a central position in health sciences research, but its practical application continues to be marked by significant tensions between technical availability, conceptual understanding, and interpretive quality (1, 6). While methodological documents and previous reviews exist on statistical inference, p-value, reproducibility, and Bayesian approaches, this review integrates theoretical foundations, recurring application problems, and contemporary methodological alternatives aimed at strengthening methodological rigor. In this sense, the objective of analyzing the application of biostatistical methods in health science research was met,

by identifying patterns of use, persistent limitations and proposals aimed at improving the interpretation, transparency and reproducibility of scientific evidence.

The studies included in this review showed that the p-value continues to occupy a central place as a criterion for statistical interpretation in health sciences research (5). However, several authors warn that its isolated use can favor reductionist interpretations, especially when it is assumed to be a direct indicator of the magnitude of the effect, the probability of the hypothesis being true, or the clinical relevance of the finding (4, 6). Consequently, the reviewed literature recommends complementing statistical significance with confidence intervals, measures of effect size, and a contextualized interpretation according to the study design, biological plausibility, and clinical relevance of the results (8-10).

The findings align with international debates that have questioned the simplistic use of statistical significance. The movement toward a richer interpretation of results proposes considering estimates, confidence intervals, plausibility, design quality, biases, and practical consequences (8). This perspective is particularly relevant in health sciences, where a statistically significant difference may lack clinical value, while a non-significant result may be relevant if the study has limitations in power or sample size (10).

The inappropriate selection of statistical tests observed in the reviewed literature reveals another critical gap. Many errors arise from applying tests without considering the type of variable, the distribution of the data, the independence of the observations, the sample size, or the study design. This situation is exacerbated by the uncritical use of statistical software. Computer programs facilitate analysis, but they do not guarantee methodological rigor. The software executes commands; the scientific judgment rests with the researcher (1).

The handling of multiple comparisons requires particular attention, according to the literature reviewed. In studies with numerous outcomes or multiple subgroups, the probability of finding seemingly significant results increases by sheer chance. If adjustments are not made or a predefined analytical strategy is not established, false positives can multiply. This problem is especially relevant in exploratory studies, secondary analyses, and high-dimensional contexts. The solution is not to prohibit exploration, but to clearly distinguish between exploratory and confirmatory analyses, apply appropriate controls, and report procedures transparently (32-33).

The confusion between association and causation is one of the most delicate limitations in health sciences. Observational studies are indispensable, but their results must be interpreted with caution. A statistical association may be due to causation, confounding, bias, chance, or a combination of these. To approach a causal inference requires more than a significant test: it requires robust designs, clear timescales, adequate measurement of variables, control of confounders, biological plausibility, and, when relevant, explicit causal models. Biostatistics provides tools, but causality demands epidemiological thinking (25, 35).

The review also shows that incomplete statistical reporting remains a barrier to critical appraisal. Lack of information on assumptions, handling of missing data, exclusion criteria, test selection, software used, or sensitivity analysis limits the critical appraisal of clinical and observational studies (36-37). In evidence reviews, PRISMA 2020 contributes to making the processes of searching, selecting, and synthesizing literature more transparent (38). In randomized controlled trials, CONSORT 2010 updates the criteria for fully communicating procedures and results (39). The original PRISMA standard also serves as a baseline guide for reporting systematic reviews (40). In an era where science demands transparency, insufficient reporting is no longer a minor detail; it is a structural flaw in reproducibility (15). A reader must be able to understand what was done, why it was done, and how the conclusions were reached.

The contemporary approaches identified offer a path for improvement. Bayesian inference, for example, allows for the integration of prior information with new data and the expression of conclusions in probabilistic terms (16, 18-19). This can be particularly useful in contexts where there is accumulated evidence or when knowledge needs to be updated progressively. The Bayesian factor allows for the comparison of relative evidence between hypotheses and can complement interpretation based on p-values (17). However, its use should not be idealized: Bayesian methods also require assumptions, decisions, and transparency, especially in the choice of prior distributions (18-19).

The adoption of Bayesian analyses in health sciences remains limited, according to the reviewed methodological literature on statistical inference, Bayesian analysis, and critiques of the exclusive use of p-values. The work of Jeffreys, Gelman et al., Goodman, and Spiegelhalter et al. indicates that Bayesian approaches allow for the integration of prior information, updating of available evidence, and expression of results in probabilistic terms (16-19); however, they also caution that their application requires greater conceptual training, a clear definition of prior distributions, and an understanding of probabilistic reasoning (18-19). Likewise, the recommendations of Wasserstein, Schirm, and Lazar, along with the arguments of Cumming, demonstrate the need to move beyond exclusive reliance on traditional frequentist statistics and toward more integrative interpretations of evidence (7-8). In this regard, the predominance of teaching focused on hypothesis testing and p-values may limit the incorporation of Bayesian approaches and other contemporary statistical models in health research.

Scientific reproducibility and open science emerge as key dimensions. Pre-registration of studies, publication of protocols, availability of anonymized databases, and access to analytical codes contribute to reducing bias and strengthening reproducibility (15). Likewise, open science practices promote methodological transparency and verifiable access to research outputs (14). These practices are not mere editorial embellishments; they are mechanisms for scientific quality control. In health research, where results can inform clinical decisions and public policy, methodological transparency is an ethical responsibility.

Strengthening biostatistics training should be a priority. The goal is not to turn all healthcare professionals into pure statisticians, but rather to train researchers capable of critically engaging with data (1). This involves understanding study designs, recognizing types of variables, selecting relevant analyses, interpreting estimates, identifying biases, critically reading results, and communicating findings judiciously (8). In medical and postgraduate training, biostatistics should be taught through real-world problems, not as a collection of formulas disconnected from research practice.

An important implication of this review is the need to integrate biostatistics, epidemiology, and research methodology (1, 25). These areas should not be taught or applied as isolated compartments. The research question determines the design; the design conditions the analysis; the analysis delimits the interpretation; and the interpretation must return to the clinical or health problem (25, 37). When this chain is broken, weak conclusions emerge, even if the analysis appears sophisticated.

In the Latin American and Ecuadorian context, these reflections are especially relevant. Scientific production in health has grown, but it still faces challenges related to methodological training, access to statistical advice, a culture of publication, funding, and the availability of robust databases. Strengthening applied biostatistics can improve the quality of theses, articles, institutional projects, and health policy decisions. Improvement depends not only on using more complex methods, but also on correctly applying the appropriate methods for each research question.

The results of this review allow us to propose several practical recommendations. First, studies should explicitly justify the selection of statistical tests. Second, all inferential analyses should be accompanied by measures of effect size and precision. Third, researchers should avoid interpreting the p-value as an absolute criterion of truth. Fourth, observational studies should control for confounding factors and avoid causal language when the design does not permit it. Fifth, articles should clearly report the handling of missing data, assumptions about the applicability of the analysis performed, control for confounding factors, measurement of exposure, potential biases, software, analytical criteria, and sensitivity analysis. Sixth, academic institutions should strengthen biostatistical consultation from the design phase, not only at the end of the study.

This review has limitations. First, as it was a scoping review with a systematic approach, a formal risk of bias assessment, as in systematic intervention reviews, was not performed. Second, the heterogeneity of the included documents prevented quantitative synthesis. Third, although several databases were explored, EMBASE was not consulted, so some relevant studies may not have been identified. Nevertheless, the approach used allowed for mapping a broad field and providing an integrative view of the application of biostatistical methods in health sciences.

As a strength, the study compiles and integrates theoretical foundations, practical problems, and contemporary alternatives, allowing for an understanding of biostatistics not only as an analytical technique but also as part of scientific reasoning. This perspective is useful for researchers, professors, graduate students, journal reviewers, and academic committees interested in improving the methodological quality of scientific output.

In summary, biostatistics will remain indispensable for health research, but its value depends on how it is used. Simply reporting a p-value or applying a test out of habit is insufficient. Biomedical research requires a more robust statistical culture, capable of interpreting uncertainty, recognizing limitations, integrating evidence, and communicating results responsibly. These considerations should also be applied to medical education. The teaching of biostatistics should not be limited to the rote memorization of formulas or isolated statistical tests, but rather should focus on developing critical thinking, the clinical interpretation of data, and evidence-based decision-making. In the training of students and healthcare professionals, it is essential to strengthen competencies in analyzing scientific results, identifying methodological errors, understanding the inherent uncertainty of biomedical research, and using evidence ethically and responsibly in clinical practice, public health, and research.

5. Conclusions

- This review identified three major areas of evidence: the theoretical foundations of statistical inference, recurring problems in the application of biostatistical methods, and contemporary methodological alternatives aimed at improving the interpretation and reproducibility of results.
- The findings show that the frequentist approach remains predominant, especially through the use of hypothesis testing and p-values. While these tools are still useful, their isolated and dichotomous application limits the interpretive quality of research. The p-value should not be used as the sole measure of evidence or as a substitute for clinical, epidemiological, and methodological reasoning.
- Among the main deficiencies identified are the incorrect interpretation of statistical significance, the omission of confidence intervals and effect sizes, the inappropriate use of tests, the lack of control for confounding variables, multiple comparisons without adjustment, incomplete reporting of analytical procedures, and confusion between

association and causality. These limitations can compromise the validity of the results and reduce the usefulness of the scientific evidence.

- Contemporary approaches, such as Bayesian inference, the use of Bayesian factors, sensitivity analysis, estimation-based interpretation, open science, and reporting guidelines, offer opportunities to strengthen health research. Furthermore, the expansion of new fields such as artificial intelligence and machine learning in medical sciences demands a solid conceptual and methodological foundation in biostatistics, since the development, validation, and interpretation of predictive models require an understanding of uncertainty, model fit, bias, reproducibility, and data quality.
- Improving the application of biostatistics in health sciences requires action at every stage, from training and study design to analysis, reporting, and editorial review. Statistical quality should not be corrected at the end of the manuscript; it must be built in from the very beginning of the project. Only in this way can scientific evidence be more valid, reproducible, and useful for clinical and healthcare decision-making.

6. References

1. Altman DG. Practical statistics for medical research. London: Chapman and Hall; **1991.** https://www.routledge.com/Practical-Statistics-for-Medical-Research/Altman/p/book/9780412276309.

2. Fisher RA. Statistical methods for research workers. Edinburgh: Oliver and Boyd; **1925.** https://archive.org/details/statisticalmetho00fish.

3. Neyman J, Pearson ES. On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc London A.* **1933,** 231, 289-337. https://doi.org/10.1098/rsta.1933.0009.

4. Goodman SN. Toward evidence-based medical statistics. 1: The p value fallacy. *Ann Intern Med.* **1999,** 130(12), 995-1004. https://pubmed.ncbi.nlm.nih.gov/10383371/.

5. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat.* **2016,** 70(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108.

6. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol.* **2016,** 31(4), 337-350. https://doi.org/10.1007/s10654-016-0149-3.

7. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond p < 0.05. *Am Stat.* **2019,** 73(sup1), 1-19. https://doi.org/10.1080/00031305.2019.1583913.

8. Cumming G. The new statistics: why and how. *Psychol Sci.* **2014,** 25(1), 7-29. https://doi.org/10.1177/0956797613504966.

9. Sullivan GM, Feinn R. Using effect size - or why the p value is not enough. *J Grad Med Educ.* **2012,** 4(3), 279-282. https://doi.org/10.4300/JGME-D-12-00156.1.

10. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: clinical versus statistical significance. *Perspect Clin Res.* **2015,** 6(3), 169-170. https://doi.org/10.4103/2229-3485.159943.

11. Ioannidis JPA. Why most published research findings are false. *PLoS Med.* **2005,** 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124.

12. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. *PLoS Biol.* **2015,** 13(3), e1002106. https://doi.org/10.1371/journal.pbio.1002106.

13. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci.* **2013,** 14(5), 365-376. https://doi.org/10.1038/nrn3475.

14. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open research culture. *Science.* **2015,** 348(6242), 1422-1425. https://doi.org/10.1126/science.aab2374.

15. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. *Nat Hum Behav.* **2017,** 1, 0021. https://doi.org/10.1038/s41562-016-0021.

16. Jeffreys H. Theory of probability. 3rd ed. Oxford: Oxford University Press; **1961.** https://archive.org/details/theoryofprobabil00jeff.
17. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med.* **1999,** 130(12), 1005-1013. https://pubmed.ncbi.nlm.nih.gov/10383350/.
18. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 3rd ed. Boca Raton: CRC Press; **2013.** https://www.routledge.com/Bayesian-Data-Analysis/Gelman-Carlin-Stern-Dunson-Vehtari-Rubin/p/book/9781439840955.
19. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester: Wiley; **2004.** https://doi.org/10.1002/0470092602.
20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol.* **2005,** 8(1), 19-32. https://doi.org/10.1080/1364557032000119616.
21. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci.* **2010,** 5, 69. https://doi.org/10.1186/1748-5908-5-69.
22. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* **2018,** 169(7), 467-473. https://doi.org/10.7326/M18-0850.
23. Popper KR. The logic of scientific discovery. London: Routledge; **1959.** https://www.routledge.com/The-Logic-of-Scientific-Discovery/Popper/p/book/9780415278447.
24. Hacking I. The emergence of probability. Cambridge: Cambridge University Press; **1975.** https://doi.org/10.1017/CBO9780511817557.
25. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; **2008.** https://shop.lww.com/Modern-Epidemiology/p/9781451190052.
26. Szklo M, Nieto FJ. Epidemiology: beyond the basics. 4th ed. Burlington: Jones & Bartlett Learning; **2019.** https://www.jblearning.com/catalog/productdetails/9781284116595.
27. Cohen J. The earth is round (p <.05). *Am Psychol.* **1994,** 49(12), 997-1003. https://doi.org/10.1037/0003-066X.49.12.997.
28. Gigerenzer G. Mindless statistics. *J Econ Partner.* **2004,** 33(5), 587-606. https://doi.org/10.1016/j.socec.2004.09.033.
29. Sterne JAC, Smith GD. Sifting the evidence - what's wrong with significance tests? *BMJ.* **2001,** 322(7280), 226-231. https://doi.org/10.1136/bmj.322.7280.226.
30. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. *R Soc Open Sci.* **2014,** 1(3), 140216. https://doi.org/10.1098/rsos.140216.
31. Nuzzo R. Scientific method: statistical errors. *Nature.* **2014,** 506(7487), 150-152. https://doi.org/10.1038/506150a.
32. Streiner DL. Best but oft-forgotten practices: the multiple problems of multiplicity - whether and how to correct for many statistical tests. *Am J Clin Nutr.* **2015,** 102(4), 721-728. https://doi.org/10.3945/ajcn.115.113548.
33. Feise RJ. Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol.* **2002,** 2, 8. https://doi.org/10.1186/1471-2288-2-8.
34. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* **1986,** 327(8476), 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8.
35. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology: explanation and elaboration. *PLoS Med.* **2007,** 4(10), e297. https://doi.org/10.1371/journal.pmed.0040297.
36. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet.* **2001,** 357(9263), 1191-1194. https://doi.org/10.1016/S0140-6736(00)04337-3.
37. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The STROBE statement: guidelines for reporting observational studies. *Lancet.* **2007,** 370(9596), 1453-1457. https://doi.org/10.1016/S0140-6736(07)61602-X.

38. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* **2021,** 372, n71. https://doi.org/10.1136/bmj.n71.
39. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *BMJ.* **2010,** 340, c332. https://doi.org/10.1136/bmj.c332.
40. Moher D, Liberati A, Tetzlaff A, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLoS Med.* **2009,** 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Table 1. Extraction and analysis matrix of the included studies.

Author, year (reference)	Document type	Thematic category	Main objective or focus	Biostatistical method/problem analyzed	Main findings	Contribution to the review
Fisher, 1925 (2)	classic methodological book	Theoretical foundations of statistical inference	Establishing a basis for significance testing and frequentist inference.	p-value, significance tests, and frequentist inference.	It consolidated the use of statistical tests to assess evidence against the null hypothesis.	It establishes historical bases for the use of the p-value as an indicator of evidence against the null hypothesis.
Neyman and Pearson, 1933 (3)	Theoretical article	Theoretical foundations of statistical inference	Develop a formal framework for hypothesis testing.	Hypothesis testing, type I and II errors.	He proposed a decision-making approach based on rules, power, and long-term error control.	It provides the foundation for the decisional approach to statistical inference and error control.
Jeffreys, 1961 (16)	Theoretical book	Theoretical foundations of statistical inference	Develop probabilistic foundations for Bayesian inference.	Probability, hypothesis testing, and Bayesian inference.	He proposed probability as a tool for updating and comparing hypotheses.	It introduces fundamentals for comparing hypotheses from a Bayesian perspective.
Popper, 1959 (23)	Epistemological book	Theoretical foundations of statistical inference	Explain the logic of scientific research and falsification.	Falsification and empirical testing of hypotheses.	He emphasized that scientific knowledge advances through critical testing and the possibility of refutation.	It provides an epistemological framework for understanding the empirical testing of hypotheses.
Hacking, 1975 (24)	Historical-epistemological book	Theoretical foundations of statistical inference	Analyze the historical emergence of probabilistic thinking.	Probability as the basis of modern scientific reasoning.	He explained the conceptual evolution of probability and its role in modern science.	Explain the evolution of probability as the basis of scientific reasoning.

Altman, 1991 (1)	Applied Methodological Book	Theoretical foundations of statistical inference	Systematize statistical procedures applied to medical research.	Applied medical statistics.	He organized essential statistical methods for the design, analysis, and interpretation of clinical studies.	Systematizes essential statistical procedures for clinical and biomedical research.
Rothman, Greenland and Lash, 2008 (25)	Methodological book	Theoretical foundations of statistical inference	Integrating modern epidemiology, biases, confounding, and causal inference.	Causal inference, bias, confounding, and epidemiological analysis.	It demonstrated the need to articulate design, statistical analysis, and causal reasoning in observational studies.	It integrates statistical reasoning, biases, confounding, and causality.
Szklo and Nieto, 2019 (26)	Methodological book	Theoretical foundations of statistical inference	Develop fundamentals of applied epidemiology and analysis of population studies.	Epidemiological design, statistical analysis and interpretation.	It strengthens the relationship between study design, analysis, and validity of the epidemiological interpretation.	It strengthens the connection between epidemiological design, statistical analysis, and interpretation.
Cohen, 1994 (27)	Conceptual article	Theoretical foundations of statistical inference	Questioning the ritualism of statistical significance.	Mechanical use of $p < 0.05$ and effect size.	He argued that dependence on the p-value limits scientific interpretation and should be complemented by effect size.	It questions the mechanical use of $p < 0.05$ and promotes interpretation based on effect size.
Gigerenzer, 2004 (28)	Theoretical-critical article	Theoretical foundations of statistical inference	Criticize the uncritical use of statistical tests.	Automatic application of statistical tests.	He warned against the ritualized use of tests without sufficient statistical reasoning.	It warns against the automatic application of tests without statistical reasoning.
Cumming, 2014 (8)	Methodological article	Theoretical foundations of statistical inference	Propose a statistical interpretation based on estimates.	New statistics, confidence intervals, and effect size.	He proposed prioritizing estimates, confidence intervals, and visualization	It proposes prioritizing confidence intervals, effect sizes, and

					over dichotomous decisions.	visual statistical thinking.
Sterne and Smith, 2001 (29)	Methodological article	Problems and limitations in the application of biostatistical methods	Examine limitations of significance testing in biomedical research.	Limitations of significance tests.	It showed that statistical significance used in isolation can lead to erroneous interpretations.	Evidence of interpretive risks when statistical significance is used in isolation.
Ioannidis, 2005 (11)	Metascientific article	Problems and limitations in the application of biostatistical methods	Analyze causes of false positives and low reproducibility in biomedical research.	False positives, biases, and reproducibility.	He argued that multiple methodological factors increase the likelihood of false findings.	It supports the need for methodological rigor and analytical transparency.
Goodman, 1999a (4)	Methodological article	Problems and limitations in the application of biostatistical methods	Explain common fallacies in the interpretation of the p-value.	p-value fallacy.	He showed that the p-value is often misinterpreted as the probability of the hypothesis being true or of chance.	Explain the consequences of misinterpreting the p-value in evidence-based medicine.
Greenland et al., 2016 (6)	Methodological article	Problems and limitations in the application of biostatistical methods	Systematize misinterpretations of statistical tests, p, intervals, and power.	P, confidence intervals, power, and statistical inference.	He identified common errors in the reading of statistical significance, precision, and power.	It guides a more rigorous interpretation of tests, intervals, and power.
Wasserstein and Lazar, 2016 (5)	Methodological statement	Problems and limitations in the application of biostatistical methods	Present ASA recommendations on the use of the p-value.	Use and interpretation of the p-value.	He cautioned that the p-value should not be used as the sole measure of evidence or as an absolute decision criterion.	It presents an institutional position on the responsible use of the p-value.

Head et al., 2015 (12)	Metascientific study	Problems and limitations in the application of biostatistical methods	Evaluate the extent and consequences of p-hacking in science.	P-hacking and analytical bias.	He identified flexible analytical practices associated with the search for statistical significance.	It provides evidence on biases arising from flexible analytical decisions.
Nuzzo, 2014 (31)	Scientific analysis article	Problems and limitations in the application of biostatistical methods	Describe common statistical errors in scientific publications.	Statistical errors in research.	He pointed out frequent failures in the interpretation and reporting of statistical results.	It reinforces the need for statistical training and transparent reporting.
Colquhoun, 2014 (30)	Methodological article	Problems and limitations in the application of biostatistical methods	Analyze the false discovery rate and the interpretation of the p-value.	False discoveries and misinterpretation of the p-value.	It showed that improper interpretation of the p-value can inflate apparent evidence.	It supports caution in the face of isolated significant results.
Button et al., 2013 (13)	Methodological article	Problems and limitations in the application of biostatistical methods	Analyze the impact of small sample size on the reliability of findings.	Low statistical power.	It showed that small samples reduce the reliability and reproducibility of the results.	It supports the importance of statistical power and adequate sample size.
Ranganathan, Pramesh and Buyse, 2015 (10)	Educational article	Problems and limitations in the application of biostatistical methods	Differentiate between clinical significance and statistical significance.	Clinical significance versus statistical significance.	He explained that a significant finding may lack clinical relevance and vice versa.	It supports the integration of clinical relevance into statistical interpretation.
Streiner, 2015 (32)	Methodological article	Problems and limitations in the application of biostatistical methods	Explain problems arising from multiple comparisons.	Multiplicity and adjustment by multiple comparisons.	It showed when and how to apply corrections to reduce false positives.	It supports the need to control multiplicity in inferential analysis.

Feise, 2002 (33)	Methodological article	Problems and limitations in the application of biostatistical methods	Discuss the need to adjust p-values for multiple outcomes.	Adjustment for multiple outcomes.	He analyzed scenarios in which multiple outcome measures require statistical adjustment.	It provides criteria for handling multiple comparisons.
Bland and Altman, 1986 (34)	Methodological article	Problems and limitations in the application of biostatistical methods	To propose appropriate methods for comparing clinical measurements.	Comparison between measurement methods.	He proposed more appropriate concordance procedures than simple correlation.	It helps to avoid inadequate analysis in measurement studies.
Goodman, 1999b (17)	Methodological article	Contemporary methodological approaches and analytical alternatives	Propose the Bayes factor as an alternative for interpreting evidence.	Bayes factor.	He showed how to compare relative evidence between competing hypotheses.	It proposes the Bayes factor as a complement or alternative to the p-value.
Gelman et al., 2013 (18)	Methodological book	Contemporary methodological approaches and analytical alternatives	Develop fundamentals and applications of Bayesian data analysis.	Bayesian models and complex data analysis.	He presented Bayesian methods for modeling uncertainty and integrating prior information.	It provides a basis for Bayesian approaches applied to health.
Spiegelhalter, Abrams and Myles, 2004 (19)	Methodological book	Contemporary methodological approaches and analytical alternatives	Applying Bayesian approaches to clinical trials and health assessment.	Bayesian inference in clinical research.	It demonstrated the usefulness of Bayesian models in the evaluation of health interventions and services.	Evidence of clinical and health applicability of Bayesian inference.
Wasserstein, Schirm and Lazar, 2019 (7)	Consensus article	Contemporary methodological approaches and analytical alternatives	Promote a scientific interpretation beyond the $p < 0.05$ threshold.	Overcoming the dichotomous use of the p-value.	He recommended abandoning rigid decisions based on statistical significance.	It supports comprehensive interpretations of scientific evidence.

Sullivan and Feinn, 2012 (9)	Educational article	Contemporary methodological approaches and analytical alternatives	Highlighting the importance of effect size in medical education and research.	Effect size.	He explained why the p-value is not sufficient to interpret practical importance.	It reinforces the need to report the magnitude of the effect.
Nosek et al., 2015 (14)	Science policy article	Contemporary methodological approaches and analytical alternatives	Promote open science practices to improve reproducibility.	Open research culture.	He proposed opening up data, materials, and protocols as transparency strategies.	It supports open practices to improve quality and reproducibility.
Munafò et al., 2017 (15)	Methodological Manifesto	Contemporary methodological approaches and analytical alternatives	Propose actions to strengthen reproducible science.	Reproducible science, pre-registration, and bias reduction.	He proposed methodological reforms to increase robustness, transparency, and reproducibility.	It provides recommendations to improve the robustness of the research.
Tricco et al., 2018 (22)	Methodological guide	Contemporary methodological approaches and analytical alternatives	Establish reporting criteria for scope reviews.	PRISMA-ScR.	He defined elements for reporting identification, selection, extraction, and synthesis in scoping reviews.	It strengthens the methodological transparency of this article.