Sapiens in Health Science

Sapiens in Health Sciences. 2026;4.1. 006
ISNN: 3072-7960
DOI: https://doi.org/10.71068/hs000106

Article: Review

Concordance, efficiency and regional application of epidemiological methodologies versus randomized trials: a systematic review and meta-epidemiological meta-analysis

Concordancia, eficiencia y aplicación regional de metodologías epidemiológicas frente al ensayo aleatorizado: revisión sistemática y meta-análisis meta-epidemiológico

[1]Marco Antonio Ramos Espinal.

[1]National Autonomous University of Honduras (UNAH), University City, Tegucigalpa, Honduras

***Corresponding author:** Marco Antonio Ramos Espinal

Submitted: 04-08-2026 **Revised:** 23-08-2026 **Accepted:** 04-09-2026 **Published:** 11-10-2026

How to cite:

Ramos Espinal, M. A. (2026). Concordance, efficiency, and regional application of epidemiological methodologies versus randomized trials: systematic review and meta-epidemiological meta-analysis. *Sapiens in Health Sciences*, *4*(1), 1-33. https://doi.org/10.71068/hs000106

ABSTRACT

Introduction: the choice of study design conditions the validity, timeliness and usefulness of epidemiological evidence. We aimed to estimate the concordance of non-randomized and emerging designs with the randomized controlled trial (RCT) and to describe their efficiency and regional application.
Method: systematic review and meta-epidemiological meta-analysis following PRISMA 2020. We searched MEDLINE, Embase, Scopus, Web of Science and Cochrane (January 2020 to May 2026). The primary comparison was restricted to five non-randomized versus randomized-trial estimates; a six-estimate mixed-design pool was retained as exploratory. Random-effects analyses included REML, Hartung-Knapp-Sidik-Jonkman and a joint exclusion of potentially overlapping sources. Risk of bias, adapted GRADE certainty and ordinal regional and efficiency matrices were assessed.
Results: Twenty-four studies were included. Five non-randomized versus randomized-trial comparisons yielded a RoR of 1.00 (95% CI 0.92–1.10); the exploratory six-estimate pool yielded 0.98 (0.89–1.06; $I^2 = 77.9\%$). Joint exclusion of Toews and Kimachi yielded approximately 0.95 (0.88–1.02) among the three remaining non-randomized comparisons, reconstructed from published rounded estimates. Subgroup RoRs were 1.05 for observational designs and 0.91 for trial emulation/real-world evidence. Certainty was low and very low, respectively. Of 28 highest regional ratings, 16 relied on indirect evidence.
Conclusions: A pooled RoR close to 1 does not establish equivalence with randomized trials. Between-design comparisons and regional scoring require cautious, family-specific interpretation; the proposed combination of methods remains a hypothesis for future evaluation.

Keywords: epidemiology; research methodology; meta-epidemiology; target trial emulation; real-world evidence; ratio of ratios; health equity

RESUMEN

Introducción: la elección del diseño condiciona la validez, la oportunidad y la utilidad de la evidencia epidemiológica. Estimamos la concordancia de los diseños no aleatorizados y emergentes con el ensayo controlado aleatorizado (ECA) y describimos su eficiencia y aplicación regional.
Método: revisión sistemática y meta-análisis meta-epidemiológico según PRISMA 2020. Consultamos MEDLINE, Embase, Scopus, Web of Science y Cochrane (enero de 2020 a mayo de 2026). Se priorizó la

©2026 by the Authors. This article is open access and distributed under the terms and conditions under a Creative Commons Attribution 4.0 International License. (CC BY 4.0) https://creativecommons.org/licenses/by/4.0/

comparación de cinco estimadores no aleatorizados frente a ECA; el agrupamiento heterogéneo de seis se mantuvo como exploratorio. Se aplicaron efectos aleatorios, REML, Hartung-Knapp-Sidik-Jonkman y exclusión conjunta de posibles fuentes solapadas. Se evaluaron riesgo de sesgo, certeza GRADE adaptada y matrices ordinales de distribución regional y eficiencia.
Resultados: se incluyeron 24 estudios. Los cinco contrastes entre diseños no aleatorizados y ECA produjeron una RoR de 1.00 (IC 95 % 0.92–1.10); el agrupamiento exploratorio de seis estimadores fue 0.98 (0.89–1.06; I^2 = 77.9 %). Al excluir conjuntamente Toews y Kimachi, los tres estimadores no aleatorizados restantes dieron aproximadamente 0.95 (0.88–1.02), reconstruido desde valores publicados redondeados. Las RoR por familias fueron 1.05 para observacionales y 0.91 para emulación/evidencia de mundo real, con certeza baja y muy baja, respectivamente. Dieciséis de 28 puntuaciones regionales máximas se basaron en evidencia indirecta.
Conclusiones: una RoR próxima a 1 no demuestra equivalencia con los ECA. Los contrastes entre diseños y las puntuaciones regionales requieren interpretación prudente y por familias; la combinación de métodos sigue siendo una hipótesis que debe evaluarse.

Palabras clave: epidemiología; metodología de la investigación; meta-epidemiología; emulación de ensayo diana; evidencia de mundo real; razón de razones; equidad en salud

INTRODUCTION

Epidemiology produces evidence through a wide repertoire of designs that differ in the way exposure is assigned, confounding controlled, and population selection. In experimental designs, the researcher assigns the intervention. The individual randomized controlled trial (RCT) does so randomly, thereby balancing measured and unmeasured confounders in expectation, which is why it occupies the top of the evidence hierarchies (1)(2). Its cluster, staggered, pragmatic, and platform variants retain randomization, but adapt it to population-based interventions, to the conditions of routine practice, or to the simultaneous evaluation of several treatments (3)(4).

In observational designs, exposure does not depend on the researcher. Cohorts follow exposed and non-exposed groups, case-control studies are based on the outcome, and cross-sectional studies describe prevalences at a given time. They gain in feasibility, sample size, and proximity to practice, and lose out due to residual confusion, selection bias, and measurement error (5). In addition, there are designs that seek to approximate causal inference without randomization, such as target trial emulation, real-world evidence analysis (RWE), test-negative design, and Mendelian randomization (2)(6-8), and schemes aimed at population surveillance, such as standardized seroepidemiology, genomic surveillance, and wastewater epidemiology (9-11).

A third group answers questions other than effectiveness. Mixed methods and hybrid effectiveness-implementation designs assess both the effect of an intervention and the conditions of its adoption (12)(13), and community-based participatory research incorporates the population as a co-producer of knowledge (14). Each family therefore has its own field of application: estimating causal effects, monitoring and early warning, or transferring interventions to specific contexts.

Since 2020, several studies have empirically compared these designs. The update of the Cochrane methodological review found small differences, on average, between the effects estimated by observational studies and RCTs, although with wide variability between comparisons (5). Meta-epidemiological studies in nephrology, nutrition, and general medical research described similar

patterns and showed that agreement increases when the population, exposure, comparator, and outcome of both bodies of evidence are similar (15-18). The RCT-DUPLICATE initiative emulated trials with insurance claims databases and documented that emulation fidelity explains a large part of the discrepancies (19-21). In accuracy, small trials of COVID-19 treatments yielded more extreme effects than large ones (22). In applicability, platform trials generated actionable evidence in weeks (4)(23), while predictive machine learning models showed high risk of bias and poor external validation (24).

This background leaves three gaps. The first is that each study compares a single pair of designs or a single clinical domain, so it is not known whether the deviations observed are specific to each methodological family or common to all. The second is that the quality of a design is usually judged only by its agreement with the RCT, without considering the speed, coverage, or sustainability of the evidence it produces. The third is that the available reviews are based almost entirely on data from high-income countries and do not examine where each design is applied, despite the fact that the COVID-19 pandemic showed marked asymmetries in sequencing, surveillance, and clinical research capacity between regions (9)(10). Comparing methodologies in the same framework is necessary because funding agencies, regulators, and universities choose designs with limited resources, and this choice determines which questions are answered and for which populations (6)(12).

The general objective of this study was to estimate the agreement of non-randomized and emergent designs with the RCT and to describe their efficiency and application by region in the period 2020-2026. We posed three research questions. (Q1) To what extent do the effect estimators of the alternative methodologies differ from those of the reference RCT, and does this deviation differ between methodological families? (Q2) What efficiency profile does each scheme show in internal validity, external validity, speed, cost and coverage, concordance, and equity and sustainability? (Q3) How is the documented application of each scheme distributed among seven regions of the world? P1 was answered with a meta-epidemiological meta-analysis, and P2 and P3, with a semi-quantitative synthesis based on explicit rubrics.

LITERATURE REVIEW

Three lines of work support the question of the study: meta-epidemiology, causal inference with observational data, and designs oriented to implementation and community participation. The section closes with a conceptual distinction between agreement, validity, and efficiency.

Meta-epidemiology and hierarchy of evidence

Meta-epidemiology studies research itself. It compares, within equivalent clinical questions, the effects estimated by different designs and expresses the difference as a ratio of ratios (RoR), the quotient between the estimator of a design and that of the reference design (25). An RoR of 1 indicates the absence of systematic difference, and values far from 1 indicate that a design produces on average higher or lower estimators on the ratio scale. Reading this direction as overestimation or underestimation of the benefit depends on how the effect is coded in each source, an aspect that any second-order synthesis must harmonize.

The Cochrane review by Toews et al. brought together meta-epidemiological studies from multiple clinical areas and found, on average, small differences between observational studies and RCTs, with low certainty (5). Specialty-specific studies in nephrology (15) and nutrition (16)(17), and analysis of bodies of evidence from Cochrane reviews (18), reached similar conclusions and pointed out that discrepancies are concentrated in pairs with less similarity in population, exposure, or outcome. They share two limitations: they are restricted to the contrast between observational design and RCTs, and they rarely distinguish the geographical origin of the data. None contrasted the emerging methodological families in a single synthesis.

Target Assay Emulation and Real-World Evidence

The target trial emulation framework proposes to specify the hypothetical RCT protocol to be performed (eligibility, treatment strategies, allocation, initiation of follow-up, outcomes, and analysis) and reproduce it with observational data (2). Its main contribution is to eliminate avoidable design biases, such as immortal time bias or misalignment between eligibility, allocation, and initiation of follow-up. The RCT-DUPLICATE initiative put this to the test by emulating published trials with health insurance claims databases. Its early results showed that agreement depended on whether the trial design could be replicated with available data (21), and analysis of 32 trials confirmed this (19)(20). Regulatory agencies debate under what conditions RWE can support decisions (6).

Two more designs expand this field. The test-negative estimates vaccine effectiveness by comparing vaccination between people with positive and negative tests, which allows continuous monitoring at low cost (7). Mendelian randomization uses genetic variants as instruments to anticipate the effect of pharmacological targets (8).

Pragmatic, participatory and implementation designs

Cluster and step-wedge trials allow population-based interventions to be evaluated when individual randomization is not feasible (3). The updated Medical Research Council guideline extended the evaluation of complex interventions to questions of context, implementation, and sustainability (12), and hybrid effectiveness-implementation designs, whose typology was consolidated in the last decade, measure the effect and implementation outcomes in the same study (13). The consolidated framework for implementation research (CFIR) has been adapted to low- and middle-income countries (26).

In parallel, community-based participatory research and Latin American critical epidemiology situate communities as co-producers of knowledge and link health with their social determination (14)(27). In Central America, intercultural dialogue methodologies with indigenous peoples illustrate this orientation (28). Accumulating evidence indicates that these approaches improve the quality of the process and community capacity, although evidence on health outcomes is still limited (14).

Concordance, Validity, and Efficiency: Three Different Criteria

This study distinguishes three criteria that the literature usually uses as equivalents. Concordance is an empirical property of estimators; it indicates how close the result of a design is to that of an RCT that answers the same question, and is measured with RoR. Methodological validity is a property of

design and its execution. Internal validity depends on the control of confounding, selection and measurement, and external validity depends on the representativeness of the population and the context (12). Finally, efficiency relates the evidence produced to the resources and time spent, and includes speed, cost, coverage and sustainability.

The three criteria do not imply each other. An observational study may agree with an RCT because biases of opposite sign are compensated, without being valid; a valid RCT may be inefficient if it arrives late or at a cost that prevents replication; and a participatory design may be efficient in sustainability without providing a comparable causal estimate. Therefore, in this article, agreement is analyzed quantitatively (P1), validity is assessed with risk of bias and certainty of evidence tools, and efficiency is described in separate dimensions (P2), without aggregating them in a single index.

METHOD

Design and reporting

We conducted a systematic review with meta-epidemiological quantitative synthesis, reported according to PRISMA 2020 (1)(29) and the reporting recommendations of meta-epidemiological research (25). The unit of analysis was the study and not the patient. We compared effect estimators produced by different designs on equivalent questions, with an adapted PECOS framework: bodies of evidence on health interventions or exposures (P); alternative methodology (E) versus the reference RCT (C); quantitative discrepancy between estimators (O); and meta-epidemiological, emulation or design comparison studies (S). The protocol was not prospectively recorded.

Eligibility Criteria

For the meta-analysis, we included studies published between January 2020 and May 2026 that quantitatively compared the effect estimators of at least two methodological schemes, one of them the RCT or its reference emulation, and that reported a measure of discordance with its uncertainty. We excluded studies prior to 2020, comparisons without a quantifiable estimator, and duplicate cohorts. For the semi-quantitative synthesis of regional application and efficiency, we also admitted, in the same time window, reference trials, systematic reviews, surveillance initiatives, implementation studies, and methodological or normative documents that provided information on the speed, coverage, cost, equity, or regional distribution of a scheme.

Sources of information and search strategy

MEDLINE/PubMed, Embase, Scopus, Web of Science (Core Collection) and Cochrane Library were consulted, with manual reference review (last update declared: May 2026). Table S1 preserves the formulations and filters reported for each platform. The main strategy combined meta-epidemiology, reference assay, alternative design and concordance expressions, and the complementary strategy identified surveillance, implementation and participation designs. To reconstruct the original 1,287 records, in addition to the equations, the interface used, the date and time of each execution, the activated limits, the number retrieved per base and the exported identifiers are needed. These details do not appear in the source file and are not presented as if they had been retrospectively checked.

Study selection and data extraction

After debugging duplicates, titles and abstracts were screened and 111 records were reviewed in full text. Author, year, setting, comparison, metric, confidence interval, clinical domain, and use in synthesis were extracted; Table 2 shows the 24 studies finally included and Table 3, the six published estimators, also available in a complementary tabular set. The 87 excluded texts were reported for general reasons in PRISMA, but a list of individual records with a specific reason is not available in this file. This pending traceability is expressly recognized, without reconstructing undocumented exclusions.

Estimator selection and overlay control

For quantitative synthesis, RoR or equivalent measures on the ratio with uncertainty scale were selected, harmonized as an alternative design divided by the reference design. Matches from primary sources were examined to avoid reusing the same comparisons. Independence, however, cannot be absolutely asserted when an aggregated study incorporates previous reviews without publishing all of its identifiers. Table S3 presents pairs with known or potential overlap and the decisions made. As an additional response to this risk, a sensitivity was calculated that simultaneously excludes the two potentially related syntheses, rather than relying exclusively on one-by-one exclusions.

The RCT-DUPLICATE reports of Franklin and Wang (19)(21) were represented by Heyard et al. (20); Schwingshackl (17), by Stadelmaier's pair analysis (16); and Bröckelmann (18), by Toews' Cochrane review (5). Potential overlap between Toews (5) and Kimachi (15) persists among the six retained sources, as a review of reviews can reuse thematic synthesis comparisons. The additional sensitivity test omits both sources simultaneously, both in the exploratory set of six and in the main subset of five. The absence of verifiable shared primary identifiers limits the definitive independence check.

Risk of bias and certainty of the evidence

There is no specific validated tool for meta-epidemiological studies, so we adapted the ROBIS domains (30) and the meta-epidemiological reporting recommendations (25) into four domains: D1, representativeness of comparison sampling (systematic search vs. selection for convenience or feasibility); D2, matching of the compared questions (population similarity, exposure, comparator, and outcome); D3, harmonization of metrics and direction of effect; and D4, treatment of dependence between comparisons and heterogeneity, and analytical transparency. Each domain was rated as low, with some doubts, or high, and the overall judgment corresponded to the worst-rated domain. The tool was applied to all ten quantitative studies; the semi-quantitative synthesis documents were characterized by their type, with no risk of bias assessment.

The GRADE assessment was used as an explanatory adaptation for the certainty of methodological comparisons, not as a certification of the efficacy of a clinical intervention (31). To keep the original logic visible, a high conventional starting point was set and levels were discounted for risk of bias, inconsistency, indirectness, imprecision, and possible publication bias. A decrease was added for potential dependence in the observational subgroup and for indirectity and imprecision in emulation/RWE, where there are only two estimators and one is at high risk. Table 6 details the

decisions, whose judgment remains the responsibility of a single evaluator, with no documented independent review.

Quantitative synthesis

The primary measure was RoR. To homogenize the direction, all estimators were expressed as the quotient between the alternative design and the reference design (the RCT, its reference emulation, or, in the sample size study, the large RCT). An RoR of 1 indicates no systematic difference, and values other than 1 consistently higher or lower estimators on the ratio scale. Since the benefit coding is not the same in all sources, we interpret RoR as relative deviation and reserve the terms overestimation and underestimation for cases where the source defines them.

Logarithmic random-effects models were used, with DerSimonian-Laird, REML, Hartung-Knapp-Sidik-Jonkman (HKSJ) intervals, individual exclusion, and family analysis (32)(33). The main substantive contrast includes five estimators of nonrandomized designs versus RCTs; the aggregate that also incorporates small versus large RCTs is retained as exploratory only. A joint exclusion of Toews and Kimachi, first on the six estimators and then on the five of the main analysis, was added to examine the influence of a possible dependence. The new results were reconstructed with RoR and rounded intervals from Table 3; therefore, they are approximations reproduced on summary data, not a reestimation of individual comparisons.

Heterogeneity was described with Q, I^2, τ^2, and prediction interval. The difference between families was maintained as an exploratory analysis, since the contrasts have different objectives and some groups only include one or two estimators. A RoR with an interval containing 1 does not demonstrate equivalence, as no clinically justified margin has been set (34). The pooled six-point analysis should not be interpreted as an overall average effect of alternative designs, since it includes a comparison between randomized trial sizes. The funnel and Egger tests with six points are presented only descriptively due to their limited power; the result does not rule out publication bias.

Semi-quantitative synthesis

Regional distribution and efficiency were described using two documentary appreciation matrices. The first used seven regions with an ordinal score of 0 to 3 and the second six efficiency domains, on a scale of 1 to 5, applied using the criteria in Table 1. Their values are not empirical rates of method use, nor do they allow conclusively classifying regions or health systems. Table S2 distinguishes scores based directly on the corpus from those estimated from indirect sources; the latter are kept as exploratory guidance and are not interpreted as evidence of proven regional implementation.

Table 1.

Scoring criteria for regional application and efficiency matrices

https://doi.org/10.71068/hs000106

Matrix or dimension	Assignment rule
Regional application (0-3)	0: no documented application in the region in 2020-2026. 1, emerging: isolated studies, pilots or minority participation in global initiatives. 2, established: recurrent application in several countries in the region or substantive participation in multinational networks. 3, extensive or leading: national programs, data infrastructure or networks led from the region.
Internal Validity (1-5)	5: individual randomization with allocation concealment. 4: cluster or adaptive randomization, with specific risks controlled. 3: control for confounding as measured by design or analysis. 2: no formal control for confounding. 1: descriptive design.
External or translational validity (1-5)	5: evidence generated in real conditions with the target population and oriented towards implementation. 3: large population in usual care settings. 1: highly selected population and controlled conditions.
Time Efficiency (1-5)	5: actionable evidence in days or weeks. 4: weeks to months. 3: months to one year. 2: one to three years. 1: more than three years.
Cost & Coverage (1-5)	5: population coverage with low marginal cost. 3: moderate cost per participant and coverage of hundreds or thousands of people. 1: high cost per participant and limited coverage.
Concordance with RCT (1-5)	5: referent. 4: quantitatively demonstrated agreement (RoR close to 1 with low heterogeneity or correlation ≥ 0.9). 3: average agreement with heterogeneity. 2: frequent discordance or poor validation. 1: not assessable.
Equity and Sustainability (1-5)	5: community co-production and documented capacity transfer. 3: local participation in implementation. 1: dependence on external infrastructure without local participation.

Note: Scores 2 and 4 of the efficiency dimensions correspond to intermediate situations between the levels described. Prepared by the authors.

To score each cell, the countries of origin of the studies, regional data, networks, national programs, time to results, coverage, and community participation were examined. The ordinal rules of Table 1 were then applied and the support was classified as direct or indirect. Among the 28 maximum scores, twelve have direct support and sixteen are based on indirect indications (Table S2); this predominance prevents attributing census accuracy or demonstrated regional differences to the matrix. The

assignment was made by a single evaluator, so it is presented as a structured documentary judgment without calculated inter-evaluator agreement.

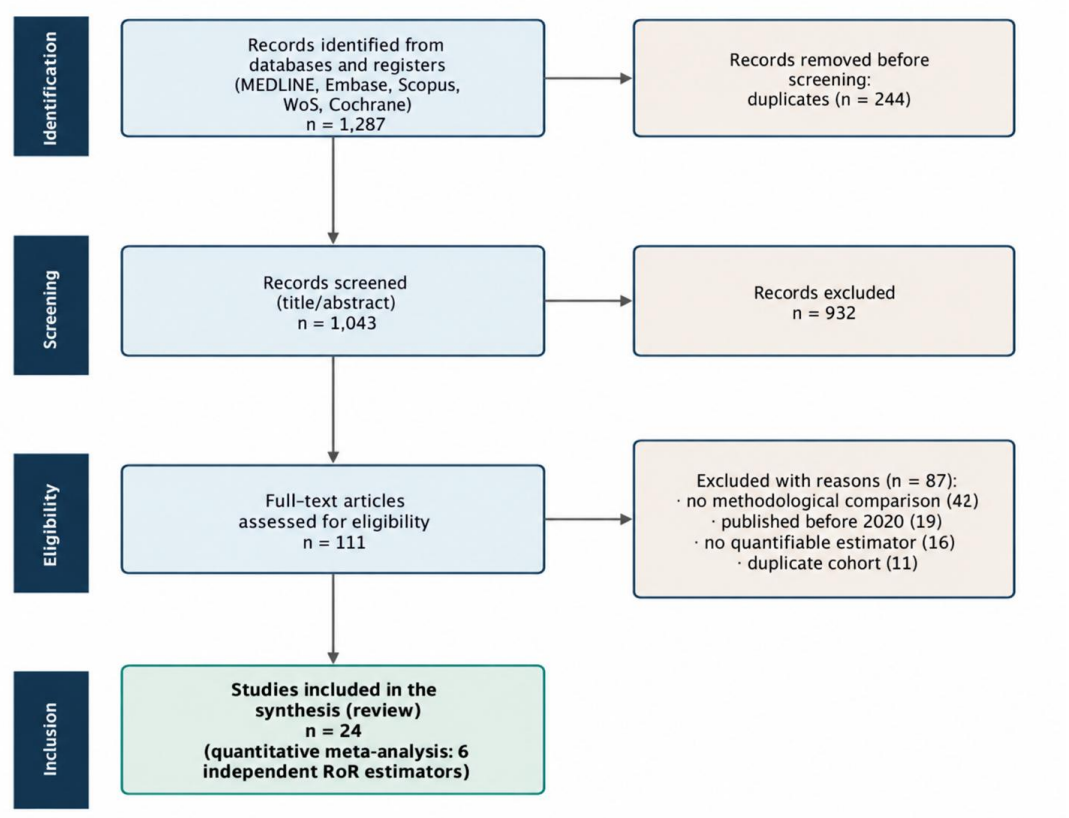
RESULTS

Study selection

The search reported 1,287 records, 244 duplicates, and 1,043 screened titles or abstracts. A total of 111 full-texts were assessed and 87 were excluded, with 24 studies being included in the final (Figure 1). Ten provided quantitative information; six provided selected estimators, and four were represented by sources with overlapping comparisons. A further fourteen were used in the semi-quantitative synthesis. Figure 1 reports aggregate counts of exclusion, not the specific reason for each of the 87 discarded texts. The individual list of decisions is not part of the documentation available for this version and would be necessary for a full audit of the process.

Figure 1.

Flowchart of the selection of studies (PRISMA 2020).



Note: Prepared by the author based on systematic research.

Study characteristics

Table 2 presents 24 studies: ten quantitative comparison between designs and fourteen semi-quantitative synthesis. Quantitative studies include five meta-epidemiological studies, three RCT-DUPLICATE publications, one RWE comparison, and one RCT size analysis. Teams from Europe and North America predominate (9)(10)(26)(28).

Table 2.

Matrix of the 24 studies included in the review

No.	Author (year) [ref.]	Country / Scope	Design	Objective	Main results
1	Toews (2024) (5)	Germany; international	Cochrane methodological review of meta-epidemiological studies	To compare estimated health effects with observational designs and RCTs	RoR 1.08 (1.01-1.15); small differences on average, low certainty
2	Kimachi (2021) (15)	Japan; international	Meta-epidemiological study	To compare estimators from observational studies and RCTs in nephrology meta-analyses	RoR 1.05 (0.90-1.23); no systematic difference
3	Stadelmaier (2025) (16)	Germany; PAI	Meta-epidemiological single-pair RCT-cohort	To assess concordance between RCTs and individual cohorts in nutrition	RRR 1.00 (0.91-1.10)
4	Schwingshackl (2021) (17)	Germany; PAI	Meta-epidemiological bodies of evidence	Comparing bodies of evidence from RCTs and cohorts in nutrition	Greater agreement when PI/ECO similarity was high
5	Bröckelmann (2022) (18)	Germany; international	Meta-epidemiological bodies of evidence from Cochrane Reviews	Comparing bodies of evidence from RCTs and cohorts in medical research	Average agreement; discrepancies associated with PI/ECO differences
6	Franklin (2021) (21)	United States	Claim Basis Testing Emulation (RCT-DUPLICATE)	Emulating the first ten RCTs with RWE studies	Concordance depended on the emulability of the design

No.	Author (year) [ref.]	Country / Scope	Design	Objective	Main results
7	Wang (2023) (19)	United States	32 RCT Emulation (RCT-DUPLICATE)	Assess whether database analyses replicate RCTs	Correlation 0.82; 0.93 in high-fidelity emulations
8	Heyard (2024) (20)	Switzerland, United States	Assay-emulation pair meta-analysis	Quantify the variation between RCTs and emulations and their design determinants	RoR 0.91 (0.86-0.96); greater deviation in cardiometabolic outcomes
9	Huang Bartlett (2020) (35)	United States	Comparison of RWE to a Phase 3 Trial	Compare Real-World PFS and Endocrine Therapy Trials in Metastatic Breast Cancer	HR Ratio 1.04 (0.69-1.56)
10	Kim (2026) (22)	South Korea; international	Meta-epidemiological study	Comparing effects of small and large RCTs on treatments for COVID-19	RoR 0.85 (0.76-0.95)
11	RECOVERY (2021) (4)	United Kingdom	Adaptive Platform ECA	Evaluating treatments for hospitalized patients with COVID-19	Dexamethasone reduced mortality with mechanical ventilation by up to a third
12	WHO Solidarity (2021) (23)	30 countries	multinational adaptive ECA	Evaluate repurposed antivirals for COVID-19	More than 11,000 adults in 405 hospitals; no relevant effect on mortality

No.	Author (year) [ref.]	Country / Scope	Design	Objective	Main results
13	Bergeri (2022) (9)	Global (OMS)	Systematic review and meta-analysis	Estimating the global seroprevalence of SARS-CoV-2 with standardized studies	Infection ratio: case close to 2:1 in PAI in the Americas and Europe and greater than 100:1 in Africa
14	Diamond (2022) (11)	United States; International	Health Policy Analysis	Assess Wastewater Pathogen Surveillance for Public Health	Early warning and population coverage at low cost
15	Brito (2022) (10)	Global	Sequencing Data Analysis	Describe disparities in SARS-CoV-2 genomic surveillance	78% of PAIs vs. 42% of LMICs sequenced more than 0.5% of cases
16	Means (2020) (26)	LMICs	Systematic review	Evaluate the use of CFIR in LMICs	CFIR requires adapting constructs to the context of LMICs
17	Ortiz (2020) (14)	United States; International	Meta-Scope Review	Synthesize processes and results of community-engaged research	Improved community capacity; limited evidence on health outcomes
18	Andaur Navarro (2021) (24)	Netherlands; international	Systematic review	Assess the risk of bias of predictive models with supervised learning	High risk of bias in most; poor external validation
19	MSPAS and PAHO (2024) (28)	Guatemala	Methodological guide	Guiding intercultural health dialogues with indigenous peoples	Participatory methodology applied to health services

No.	Author (year) [ref.]	Country / Scope	Design	Objective	Main results
20	Dean (2021) (7)	United States	Methodological analysis	Examine the test-negative design to estimate vaccine effectiveness	Efficient design for continuous surveillance; sensitive to attention-seeking bias
21	Sanderson (2022) (8)	United Kingdom; international	Methodological review	Describe the fundamentals and applications of Mendelian randomization	Useful for anticipating the effects of drug targets
22	Concato (2022) (6)	United States	Regulatory analysis	Describe the use of RWE in regulatory decisions	Conditions for accepting RWE as regulatory evidence
23	Hemming (2020) (3)	United Kingdom, Canada	Methodological review	Define when a step wedge test is appropriate	Useful for staggered population-based interventions
24	Skivington (2021) (12)	United Kingdom	Methodological Guide (MRC)	Update the framework for developing and evaluating complex interventions	Extends assessment to context, implementation, and sustainability

Note: MA: included in meta-analysis; SP: excluded from meta-analysis due to overlapping of primary comparisons (study representing it in parentheses); SC: semi-quantitative synthesis; RCT: randomized controlled trial; RWE: real-world evidence; RoR: ratio of ratios; RRR: ratio of risk ratios; HR: ratio of hazards; PFS: progression-free survival; PAI: high-income countries; LMICs: low- and middle-income countries; PI/CEE: population, intervention or exposure, comparator and outcome. Authors.

The six estimators in Table 3 cover nephrology, nutrition, oncology, various clinical domains, and COVID-19. Three contrast observational studies with RCTs (5)(15)(16), two emulation or RWE with RCTs (20)(35), and one compares small and large RCTs (22). The latter answers a question about effects of trial size, not the relative validity of non-randomized methods. Therefore, the primary analysis is restricted to the first five relevant contrasts and the set of six is preserved to illustrate conceptual heterogeneity. Most sources represent populations in high-income countries.

Table 3.

RoR estimators selected for synthesis (with possible residual dependence)

Source (year) [ref.]	Comparison	Metrics	RoR (95% CI)	Region	Domain
Toews 2024 (5)	Observational vs. RCT	RoR	1.08 (1.01-1.15)	Global	All areas
Kimachi 2021 (15)	Observational vs. RCT	RoR	1.05 (0.90-1.23)	Global	Nephrology
Stadelmaier 2025 (16)	Cohort vs. RCT	RRR	1.00 (0.91-1.10)	Global (PAI)	Nutrition
Heyard 2024 / RCT-DUPLICATE (19)(20)	Emulation / RWE vs. ECA	RoR	0.91 (0.86-0.96)	Global (PAI)	Multi-domain
Huang Bartlett 2020 (35)	RWE (emulation) vs. RCT	HR Ratio	1.04 (0.69-1.56)	North America	Oncology (breast)
Kim 2026 (22)	Small vs. Large RCT	RoR	0.85 (0.76-0.95)	Global	COVID-19

Note: All RoRs are expressed as alternative design/reference design (in (1), small RCTs/large RCTs). MYP: high-income countries; RoR: ratio of ratios; RRR: ratio of risk ratios; HR: hazard ratios. Prepared by the authors based on the included studies.

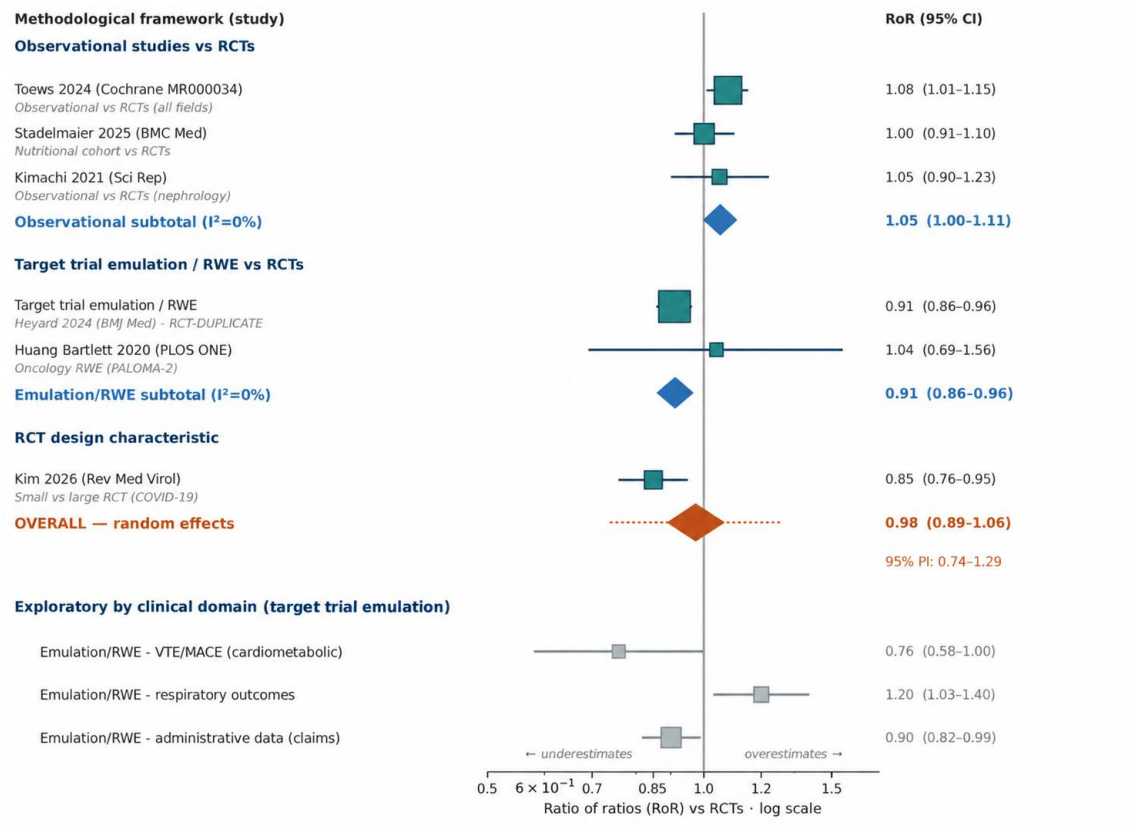
Concordance with RCT

The comparison of interest is the discrepancy between estimators obtained with non-randomized designs and RCTs. The restricted aggregate of five RoRs is therefore considered the main outcome. The sixth comparison, small versus large RCTs, addresses a problem of size and bias of randomized trials and is reported separately. The exploratory set of six provides context on heterogeneity between families, but does not constitute a homogeneous estimate of non-randomized agreement. In addition, the possible dependence between Cochrane review and nephrology makes it necessary to examine the result when removing both sources, as shown in Table 4.

Primary analysis restricted to five nonrandomized estimators versus RCTs yielded a RoR of 1.00 (95% CI 0.92–1.10; I^2 = 75.8%). The exploratory combination of the six comparisons, represented in Figure 2, yielded a RoR of 0.98 (95% CI 0.89–1.06; p = 0.57), heterogeneity I^2 = 77.9%, and prediction interval of 0.74–1.29. Neither mean demonstrates methodological equivalence. The additional comparison between RCT sizes is maintained as a separate stratum and is not part of the main inference about nonrandomized designs.

Figure 2.

Forest plot of the meta-analysis of random effects of ratio ratio (RoR) versus baseline RCT, with subtotals by methodological family.



Note: the graph preserves the six sources to show the three methodological families. The result of six is exploratory: RoR 0.98 (95% CI 0.89–1.06), I^2 = 77.9%. The main comparison excludes small RCTs versus large RCTs (k = 5; RoR 1.00, 95% CI 0.92–1.10). Authors' elaboration with the estimators of the manuscript.

The between-subgroup test was significant (Q = 20.5; df = 2; p < 0.001). Observational studies produced estimators on average 5% higher than RCTs on the ratio scale (RoR 1.05; 95% CI 1.00-1.11; I^2 = 0%), in line with the Cochrane review (5) (RoR 1.08) and with the agreement described in nutrition (16) and nephrology (15). Emulation of target trials and RWE produced 9% lower estimators (RoR 0.91; 95% CI 0.86-0.96; I^2 = 0%), with a greater deviation in cardiometabolic outcomes (RoR 0.76) than in respiratory outcomes (RoR 1.20) (20). Small trials showed more extreme effects than large trials (RoR 0.85; 95% CI 0.76-0.95) (22). Families deviate from the referent in opposite directions, so that the overall value close to 1 results, in part, from the compensation between deviations of the opposite sign.

The restricted primary analysis remained close to unity (k = 5; RoR 1.00; 95% CI 0.92–1.10), although with considerable heterogeneity. The results of six estimators, presented for exploratory purposes only, varied according to REML, HKSJ, and individual exclusion (Table 4). In an additional sensitivity that Toews and Kimachi jointly withdrew, the remaining four estimators in the exploratory set produced approximately RoR 0.92 (95% CI 0.86–0.99), while the remaining three nonrandomized ones yielded RoR 0.95 (0.88–1.02). These figures were reconstructed from the published and rounded RoRs and CIs, so they do not amount to a check of independence of the primary comparisons. The variation forces us to avoid conclusions of universal stability.

Table 4.

Restricted Main Analysis and Sensitivities of Exploratory Clustering

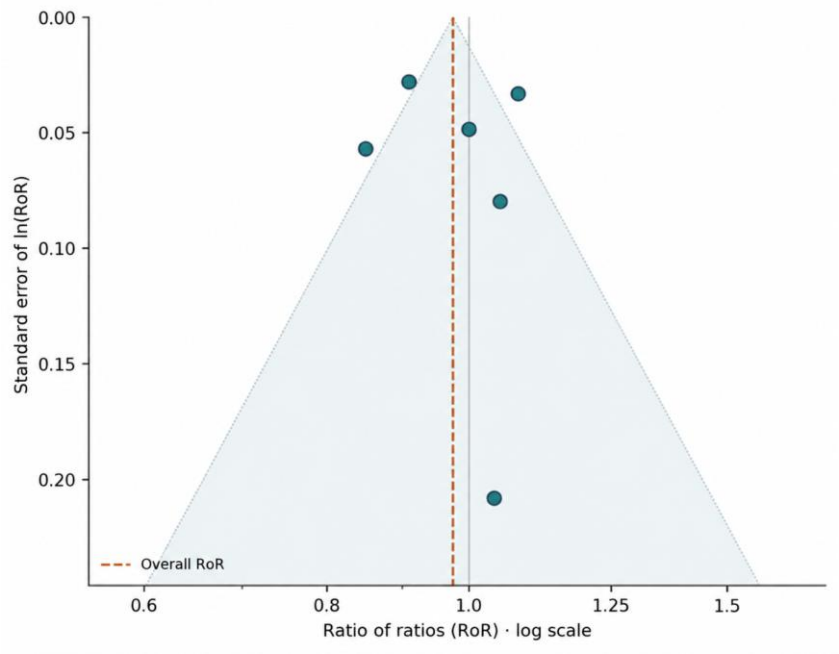
Non-randomized designs only vs. RCTs (no Kim 2026)	**5**	**1.00 (0.92-1.10)**	**75.8**	**0.007**	**0.74-1.36**
Analysis	k	RoR (IC 95 %)	I^2 (%)	t^2	IP 95 %
Analysis	k	RoR (IC 95 %)	I^2 (%)	t^2	IP 95 %
No Toews and Kimachi (non-randomized only vs. RCT)	3	0.95 (0.88–1.02)	35.9	0.002	Not calculated
Sin Toews and Kimachi (six exploratory sources)	4	0.92 (0.86–0.99)	44.5	0.002	Not calculated
τ^2 by REML	6	0.97 (0.90-1.06)	77.9	0.007	0.75-1.27
DerSimonian-Laird con HKSJ	6	0.98 (0.88-1.08)	77.9	0.008	0.74-1.29
REML with HKSJ	6	0.97 (0.88-1.08)	77.9	0.007	0.75-1.27
Without the HR-based estimator (Huang Bartlett 2020)	5	0.97 (0.89-1.07)	82.2	0.009	0.70-1.35
Sin Toews 2024	5	0.94 (0.88-1.01)	49.7	0.003	0.76-1.16
Without Kimachi 2021	5	0.96 (0.87-1.06)	81.4	0.009	0.69-1.35
Sin Stadelmaier 2025	5	0.97 (0.87-1.08)	81.9	0.011	0.67-1.41
Sin Heyard 2024	5	1.00 (0.90-1.10)	70.4	0.008	0.72-1.38

Note: HKSJ: Hartung-Knapp-Sidik-Jonkman; REML: maximum restricted likelihood; PI: prediction interval; k: number of estimators. The joint exclusion rows were calculated from estimates and rounded limits in Table 3 by DerSimonian-Laird; their accuracy is approximate. The other values are from the previous version of the manuscript. Authors.

RCT-DUPLICATE shows what determines concordance. When emulating 32 RCTs, the overall correlation was 0.82 and rose to 0.93 when the design could be emulated with high fidelity (19). Mendelian randomization also showed remarkable qualitative agreement with drug trials (8). The funnel plot did not show manifest asymmetry and the Egger test was not significant (intercept = 0.39; p = 0.86), although with six estimators its power is limited (Figure 3).

Figure 3.

Funnel chart of the six selected estimators and exploratory assessment of publication bias.



Note: Non-significant Egger's test (intercept = 0.39; k = 6). Authors.

Risk of bias and certainty of the evidence

Of the ten quantitative studies, one was at low risk of bias, eight had some doubts, and one was at high risk, the latter because it was based on a single comparison (Table 5). The doubts were concentrated in the D1 domain, due to the selection of trials according to their feasibility of emulation in RCT-DUPLICATE, and in D2, due to the imperfect similarity between the questions compared in the body of evidence studies.

Table 5.

Risk of bias of the included quantitative studies

Study	D1	D2	D3	D4	Global
Toews 2024	Low	Doubts	Low	Low	Doubts
Kimachi 2021	Low	Doubts	Low	Doubts	Doubts
Stadelmaier 2025	Low	Low	Low	Low	Low
Schwingshackl 2021	Low	Doubts	Low	Low	Doubts
Bröckelmann 2022	Low	Doubts	Low	Low	Doubts
Franklin 2021	Doubts	Low	Low	Doubts	Doubts
Wang 2023	Doubts	Low	Low	Low	Doubts
Heyard 2024	Doubts	Low	Low	Low	Doubts
Huang Bartlett 2020	High	Low	Low	Doubts	High
Kim 2026	Low	Low	Doubts	Doubts	Doubts

Note: D1: representativeness of the comparison sampling; D2: matching of the questions compared; D3: harmonization of the metric and the direction of the effect; D4: dependence, heterogeneity and analytical transparency. Low: low risk; Doubts: some doubts; High: high risk. Adapted from ROBIS (30) and Murad and Wang (25). Authors.

The adapted certainty of the evidence remained very low for the exploratory mean of six comparisons, due to heterogeneity, indirectity and imprecision. For observational versus RCTs, we estimated it to be low, when considering the risk of bias in pairing and the possibility of reusing comparisons between Toews and Kimachi. For emulation/RWE versus RCTs, certainty was assessed as very low: only two estimators, emulability-dependent selection, one high-risk, and very different study intervals. For small versus large RCTs, it remained low for a single estimator and its limitation to the COVID-19 context. Table 6 shows the reductions applied.

Table 6.

Certainty of the evidence (GRADE) by comparison

Comparison	**k**	**RoR (95% CI)**	**Reasons for downgrading certainty**	**Certainty**
Global (all families)	6	0.98 (0.89-1.06)	High game; −1 serious inconsistency (I^2 = 77.9%), −1 indirectity (different families and questions), −1 imprecision (prediction interval 0.74–1.29).	Very low
Observational vs. RCT	3	1.05 (1.00-1.11)	High score; −1 risk of bias (D2 matching), −1 possible dependence between Toews and Kimachi not ruled out with primary identifiers. Not decreased by I^2 within the subgroup.	Low

Comparison	k	RoR (95% CI)	Reasons for downgrading certainty	Certainty
Emulation / RWE vs. ECA	2	0.91 (0.86-0.96)	High score; −1 bias (selection for emulability; one at high risk), −1 indirectity (distinct domains), −1 imprecision (only k = 2 and a wide individual CI).	Very low
Small vs. Large RCT	1	0.85 (0.76-0.95)	High score; −1 indirectity (COVID-19 treatments only), −1 imprecision (one estimator, no consistency between assessable studies).	Low

Note: Adaptation of GRADE to meta-epidemiological comparisons, with a high conventional starting point and explicit reductions. It was not downgraded for publication bias exclusively on the basis of Egger, whose power is insufficient with k = 6; nor can this bias be ruled out. PI: prediction interval; k: number of estimators. Evaluation carried out by a single investigator.

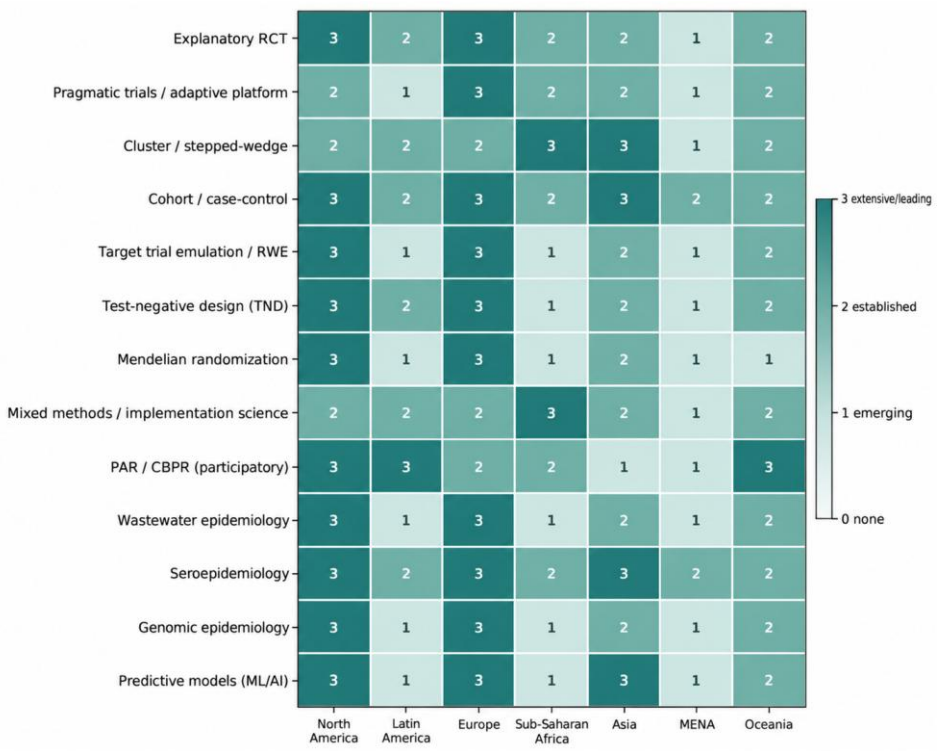
Regional distribution

Figure 4 represents the ordinal classification of the material located according to the rubric of Table 1, not a census measurement of regional use. Of the 28 maximum scores, twelve have direct support and sixteen depend on indirect documentation (Table S2). Therefore, the classification allows the identification of regions and methods that require additional research, but does not establish conclusive empirical hierarchies of application between continents.

The most well-supported differences concern genomic and serological surveillance capacity, for which data are available by income level and region (9)(10). These results point to inequalities in infrastructure and reporting, but do not allow for a complete classification of all methodologies to be extrapolated. The high scores of participatory and hybrid designs outside Europe or North America stem in part from reviews and guidance that do not provide comparable regional rates (14)(26)(28). Indirect cells and those for the Middle East and North Africa or Oceania should be interpreted only as documentary signals, without assigning them a higher frequency of proven application.

Figure 4.

Documented application intensity of each methodological scheme by region of the world, 2020-2026.



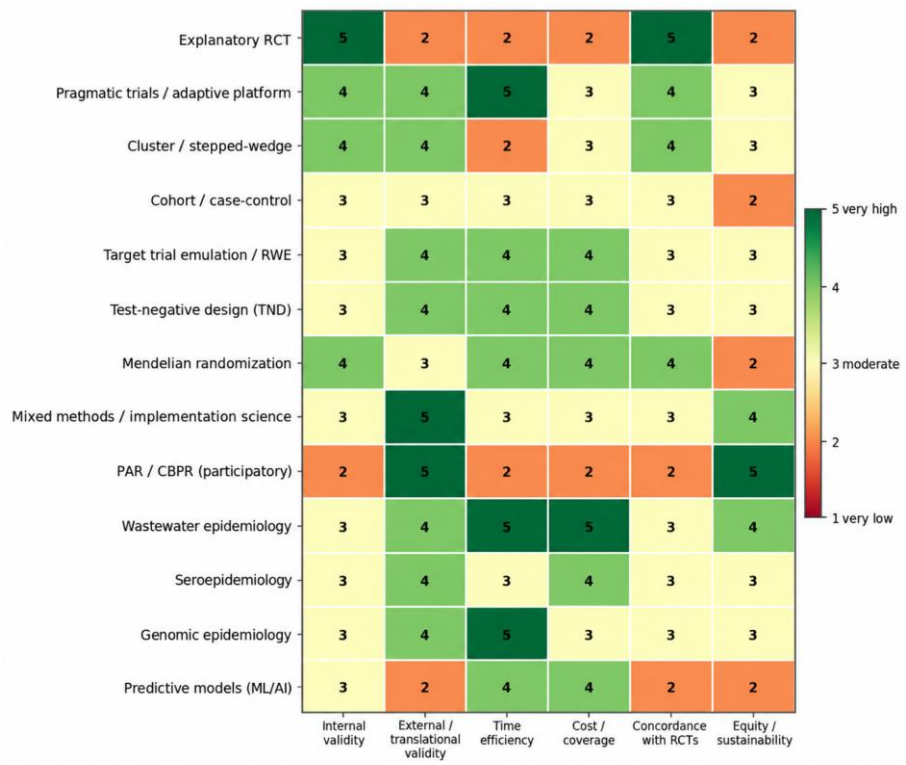
Note: exploratory classification on an ordinal scale of 0–3 according to Table 1. Of the 28 cells with a score of 3, sixteen have indirect support, so the map does not represent observed regional rates or confirmed differences. See Table S2. Source. Authors.

Efficiency by dimension

No single scheme dominated in the six dimensions evaluated (Figure 5; Table 7). The explanatory RCT retains maximum internal validity, with modest temporal efficiency. Adaptive platform trials attenuated this bottleneck. RECOVERY enrolled its first patient about nine days after writing the protocol and demonstrated that dexamethasone reduced mortality in mechanically ventilated patients by up to one-third (4), and SOLIDARITY randomized more than 11,000 adults in 405 hospitals in 30 countries (23). Wastewater surveillance stood out for its speed and low-cost coverage (11). Standardized seroepidemiology revealed massive underdetection, with a ratio between infections and reported cases close to 2:1 in high-income countries in the Americas and Europe and greater than 100:1 in Africa (9).

Figure 5.

Indicative profile of efficiency by dimension and methodological scheme, obtained by means of a documentary rubric.



Note: Ordinal judgment scale 1–5 based on Table 1. Scores do not equate to comparative measures of cost, performance, or effectiveness and are not aggregated into a single index. Single-author evaluation.

The profile in Figure 5 summarizes documentary appreciations, not direct experimental comparisons of efficiency. Participatory designs received higher scores in equity and sustainability; predictive models, in potential velocity, with risk of bias warnings and external validation (14)(24). Implementation methods were considered relevant for contextual transfer (3)(12)(26). Table 7 describes these characteristics and the regions mentioned in the corpus, whose direct or indirect support should be consulted in Table S2. None of these scores demonstrate relative superiority without homogeneous comparative data.

Table 7.

Key efficiency and regions of greatest application by methodological scheme

Methodological outline	**Efficiency profile according to available evidence**	**Regions indicated in the corpus (indicative)**	**Ref.**
Explanatory ECA	Maximum internal validity; concordance referent	North America, Europe	(2)
Pragmatic Testing / Adaptive Platform	Actionable evidence in weeks; high accuracy	Europe; 30 countries (SOLIDARITY)	(4) (23)
Conglomerates / stepped wedge	Population Assessment in Real Contexts	Sub-Saharan Africa, Asia, Latin America	(3)

Methodological outline	Efficiency profile according to available evidence	Regions indicated in the corpus (indicative)	Ref.
Target Assay Emulation / RWE	Approximates ECA when emulation is high-fidelity; fast	North America, Europe	(6) (19) (20)
Test-negative design	Continuous monitoring of vaccine effectiveness	Global (multicontinental)	(7)
Mendelian randomization	Anticipates effects of drug targets	North America, Europe	(8)
Mixed methods/implementation	Closes the effectiveness-practice gap; sustainability	Sub-Saharan Africa, Latin America	(12) (26)
IAP / CBPR (participatory)	Equity, ownership and sustainability	Latin America, Oceania, North America	(14) (27) (28)
Wastewater epidemiology	Early warning (days to weeks); low cost	North America, Europe, Asia	(11)
Standardized seroepidemiology	Measures actual infection; corrects underreporting	Global	(9)
Genomic epidemiology	Variant detection; transmission tracing	North America, Europe	(10)
Predictive Models (ML/AI)	Analytical speed; requires external validation	North America, Europe, Asia	(24)

Note: RWE: real-world evidence; PAR: participatory action research; CBPR: community-based participatory research; ML/AI: machine learning/artificial intelligence. The regions listed are indicative, with regional evidence of varying quality. Authors.

DISCUSSION

Main findings

The primary substantive result comes from five comparisons between nonrandomized designs and RCTs and yields RoR 1.00 (95% CI 0.92–1.10), with considerable heterogeneity. This average does not equate to demonstrating absence of bias or equivalence, as individual comparisons and harmonization of metrics differ. The exploratory inclusion of the RCT size study yields RoR 0.98 (0.89–1.06), but mixes different questions. The separation between the two sets preserves the research question and prevents a trial-size-related deviation from being interpreted as property of all nonrandomized designs.

Implications of heterogeneity and number of estimators

Interpretation is conditioned by few estimators, possible dependence between Toews and Kimachi, and high heterogeneity. By excluding both sources together, the rounded-information reanalysis in Table 3 yielded RoR of approximately 0.95 (95% CI 0.88–1.02) among the remaining three

nonrandomized contrasts, and 0.92 (0.86–0.99) when the study on RCT sizes is also preserved. The divergence between sets shows how much the definition of the aggregate question influences. Neither the joint exclusion proves independence from the remaining sources nor the result of a sensitivity allows correcting all conceptual or clinical differences (32)(33).

Interpretation by methodological family

RoR by families retains explanatory interest, but confidence in these estimates was reduced. The observational group provides three contrasts, with a risk of overlap between a general review and a thematic study; emulation/RWE, two heterogeneous comparisons, one of them classified as high risk. A RoR close to unity does not identify the absence of confusion, and a discrepancy does not allow attributing its origin to a specific bias without data from the study. RCT-DUPLICATE provides information on concordance conditioned by emulation fidelity (19)(20), but this observation does not guarantee extrapolation to oncology or other domains. The test-negative design (7) also responds to different inference conditions.

Regional dimension and equity

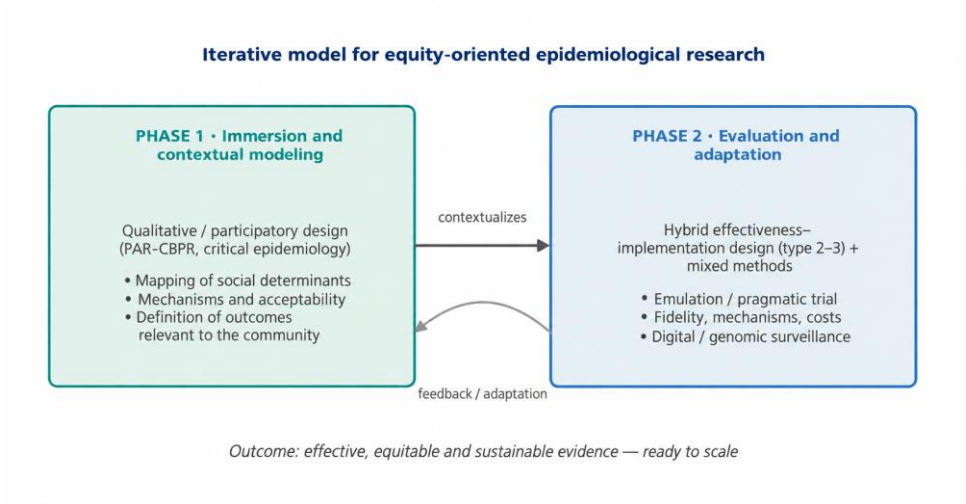
The regional matrices should be read as provisional documentary cartography. Of the 28 maximum scores, 16 come from indirect inferences, so it is not possible to conclude that certain families are systematically more used in Latin America, Africa, or Europe. There is direct evidence on inequalities in genomic sequencing and serological surveillance (9)(10), as well as contextualized examples of community participation (14)(28). Without comparable regional denominators, study counts by design, and independent auditing of the rubrics, the other differences in Figure 4 are distributional hypotheses and not consolidated empirical facts. Equity is a relevant normative consideration, distinct from the descriptive score.

Proposal for a two-phase framework

Based on this complementarity, we propose, as a working hypothesis and not as a result of the quantitative analysis, a two-phase operational framework (Figure 6). The first, immersion and contextual modeling, uses a qualitative or participatory design to map determinants and define relevant outcomes. The second, evaluation and adaptation, uses a hybrid effectiveness-implementation design with high-fidelity emulation or with a pragmatic trial (12)(13)(26). That this combination is more efficient in equity and sustainability than isolated designs is an interpretative proposition, derived from semi-quantitative synthesis, which should be contrasted with studies that prospectively compare methodological strategies.

Figure 6.

A two-phase synthetic methodological framework for equity-oriented epidemiological research.



Note: Interpretative proposal. Own elaboration.

Strengths and limitations

The contribution of the review is to define comparisons by family, to separate a restricted primary outcome from an exploratory one, and to present the factors that condition their interpretation. The sensitivity with simultaneous exclusion of Toews and Kimachi expands the assessment of possible overlap. The direct or indirect nature of regional support is also documented in a differentiated manner, and decisions to downgrade GRADE are made explicit. Table 2 and the complementary tabular set allow us to identify the studies and aggregate figures of the six comparisons used. These improvements favor transparency, although they do not replace access to primary screening and extraction registries.

Relevant limitations remain: there is no individual list of the 87 excluded full texts or execution logs that reproduce the count by database. The possible dependence between some syntheses cannot be definitively resolved without comparison of trials and source cohorts. The six estimators combine different comparisons and the restricted primary metric rests on only five. Certainty and regional matrices were assessed by a single person, with no second independent evaluation; arithmetic aggregate checks do not replace this validation. More than half of the regional maximum scores come from indirect evidence. The protocol was not prospectively enrolled, and the available evidence limits the generalizability of conclusions.

Implications for science policy

Funding agencies, journals, and universities could value each scheme for its adequacy to the question and context, rather than treating RCTs as the sole standard of quality, and requiring transparency about the fidelity of emulations and reporting of implementation (1)(2)(12). Reducing the capacity gap in the global South would encourage the most timely evidence to represent the real burden of disease.

CONCLUSIONS

No design showed universal superiority. The primary outcome, restricted to five comparisons between nonrandomized methods and RCTs, was RoR 1.00 (95% CI 0.92–1.10), with no equivalence

demonstrated. Sensitivity that excluded two potentially dependent sources together did not eliminate all uncertainty. The exploratory mean of six estimators, which includes the contrast between small and large RCTs, should be interpreted as heterogeneous and conceptually mixed. The adapted GRADE certainty of the subgroups was low or very low, so the aggregate differences do not allow us to propose a universal correction factor by design.

From the semi-quantitative synthesis we propose, as an interpretative hypothesis, that the combination of methods can improve the efficiency of epidemiological research: high-fidelity emulation and pragmatic or platform assays to estimate effects; genomic, serological and environmental surveillance to gain coverage and opportunity; and participatory and hybrid implementation designs to ensure relevance and sustainability. This proposition is not derived from the meta-analysis and requires empirical contrast.

The strongest regional evidence describes asymmetries in surveillance capacity, especially genomic and serological. The remaining regional classifications are based on a documentary matrix with a significant indirect component and do not allow us to affirm observed differences in adoption for all designs. These scores guide priorities for future comparative studies and audits with the participation of independent evaluators.

BIBLIOGRAPHIC REFERENCES

1. Kim DH, et al. The impact of study size on COVID-19 treatment outcomes: a meta-epidemiological study comparing large and small randomized controlled trials. Rev Med Virol. 2026; 36(1):e70125. https://doi.org/10.1002/rmv.70125.
2. 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.
3. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. JAMA. 2022; 328(24):2446-7. https://doi.org/10.1001/jama.2022.21383.
4. Hemming K, Taljaard M. Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice? Int J Epidemiol. 2020; 49(3):1043-52. https://doi.org/10.1093/ije/dyaa077
5. RECOVERY Collaborative Group; Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in hospitalized patients with Covid-19. N Engl J Med. 2021; 384(8):693-704. https://doi.org/10.1056/NEJMoa2021436.
6. Toews I, Anglemyer A, Nyirenda JLZ, Alsaid D, Balduzzi S, Grummich K, et al. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials: a meta-epidemiological study. Cochrane Database Syst Rev. 2024; 1(1):MR000034. https://doi.org/10.1002/14651858.MR000034.pub3.
7. Concato J, Corrigan-Curay J. Real-world evidence: where are we now? N Engl J Med. 2022; 386(18):1680-2. https://doi.org/10.1056/NEJMp2200089
8. Dean NE, Hogan JW, Schnitzer ME. Covid-19 vaccine effectiveness and the test-negative design. N Engl J Med. 2021; 385(15):1431-3. https://doi.org/10.1056/NEJMe2113151.

9. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. Nat Rev Methods Primers. 2022;2:6. https://doi.org/10.1038/s43586-021-00092-5.
10. Bergeri I, Whelan MG, Ware H, Subissi L, Nardone A, Lewis HC, et al. Global SARS-CoV-2 seroprevalence from January 2020 to April 2022: a systematic review and meta-analysis of standardized population-based studies. PLoS Med. 2022; 19(11):e1004107. https://doi.org/10.1371/journal.pmed.1004107.
11. Brito AF, Semenova E, Dudas G, Hassler GW, Kalinich CC, Kraemer MUG, et al. Global disparities in SARS-CoV-2 genomic surveillance. Nat Commun. 2022; 13(1):7003. https://doi.org/10.1038/s41467-022-33713-y.
12. Diamond MB, Keshaviah A, Bento AI, Conroy-Ben O, Driver EM, Halden RU, et al. Wastewater surveillance of pathogens can inform public health responses. Nat Med. 2022; 28(10):1992-5. https://doi.org/10.1038/s41591-022-01940-x.
13. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021; 374:n2061. https://doi.org/10.1136/bmj.n2061.
14. Curran GM, Landes SJ, McBain SA, Pyne JM, Smith JD, Fernandez ME, et al. Reflections on 10 years of effectiveness-implementation hybrid studies. Front Health Serv. 2022;2:1053496. https://doi.org/10.3389/frhs.2022.1053496.
15. Ortiz K, Nash J, Shea L, Oetzel J, Garoutte J, Sanchez-Youngman S, et al. Partnerships, processes, and outcomes: a health equity-focused scoping meta-review of community-engaged scholarship. Annu Rev Public Health. 2020;41:177-99. https://doi.org/10.1146/annurev-publhealth-040119-094220.
16. Kimachi M, Onishi A, Tajika A, Kimachi K, Furukawa TA. Systematic differences in effect estimates between observational studies and randomized control trials in meta-analyses in nephrology. Sci Rep. 2021; 11(1):6088. https://doi.org/10.1038/s41598-021-85519-5.
17. Stadelmaier J, Bantle G, Gorenflo L, Kiesswetter E, Nikolakopoulou A, Schwingshackl L. Evaluating agreement between individual nutrition randomised controlled trials and cohort studies: a meta-epidemiological study. BMC Med. 2025;23. https://doi.org/10.1186/s12916-025-03860-2
18. Schwingshackl L, Balduzzi S, Beyerbach J, Bröckelmann N, Werner SS, Zähringer J, et al. Evaluating agreement between bodies of evidence from randomised controlled trials and cohort studies in nutrition research: meta-epidemiological study. BMJ. 2021; 374:n1864. https://doi.org/10.1136/bmj.n1864.
19. Bröckelmann N, Balduzzi S, Harms L, Beyerbach J, Petropoulou M, Kubiak C, et al. Evaluating agreement between bodies of evidence from randomized controlled trials and cohort studies in medical research: a meta-epidemiological study. BMC Med. 2022; 20(1):174. https://doi.org/10.1186/s12916-022-02369-2.
20. Wang SV, Schneeweiss S; RCT-DUPLICATE Initiative. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. JAMA. 2023; 329(16):1376-85. https://doi.org/10.1001/jama.2023.4221.

21. Heyard R, Held L, Schneeweiss S, Wang SV. Design differences and variation in results between randomised trials and non-randomised emulations: meta-analysis of RCT-DUPLICATE data. BMJ Med. 2024; 3(1):e000709. https://doi.org/10.1136/bmjmed-2023-000709.
22. Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K, et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative. Circulation. 2021; 143(10):1002-13. https://doi.org/10.1161/CIRCULATIONAHA.120.051718.
23. WHO Solidarity Trial Consortium; Pan H, Peto R, Henao-Restrepo AM, Preziosi MP, Sathiyamoorthy V, et al. Repurposed antiviral drugs for Covid-19: interim WHO Solidarity trial results. N Engl J Med. 2021; 384(6):497-511. https://doi.org/10.1056/NEJMoa2023184.
24. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ. 2021; 375:n2281. https://doi.org/10.1136/bmj.n2281.
25. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med. 2017; 22(4):139-42. https://doi.org/10.1136/ebmed-2017-110713.
26. Means AR, Kemp CG, Gwayi-Chore MC, Gimbel S, Soi C, Sherr K, et al. Evaluating and optimizing the consolidated framework for implementation research (CFIR) for use in low- and middle-income countries: a systematic review. Implement Sci. 2020; 15(1):17. https://doi.org/10.1186/s13012-020-0977-0.
27. Breilh J. Critical epidemiology and the people's health. New York: Oxford University Press; 2021. ISBN: 9780190492793.
28. Ministry of Public Health and Social Assistance of Guatemala, Pan American Health Organization. Guide to Intercultural Dialogues in Health. Guatemala: MSPAS, PAHO; 2024. Available at: https://www.paho.org/es/documentos/guia-dialogos-interculturales-salud
29. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021; 372:n160. https://doi.org/10.1136/bmj.n160.
30. Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016;69:225-34. https://doi.org/10.1016/j.jclinepi.2015.06.005.
31. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011; 64(4):383-94. https://doi.org/10.1016/j.jclinepi.2010.04.026.
32. IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014;14:25. https://doi.org/10.1186/1471-2288-14-25.
33. Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. Res Synth Methods. 2019; 10(1):83-98. https://doi.org/10.1002/jrsm.1316.
34. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ. 1995; 311(7003):485. https://doi.org/10.1136/bmj.311.7003.485.
35. Huang Bartlett C, Mardekian J, Cotter MJ, Huang X, Zhang Z, Parrinello CM, et al. Concordance of real-world versus conventional progression-free survival from a phase 3 trial of endocrine

therapy as first-line treatment for metastatic breast cancer. PLoS One. 2020; 15(4):e0227256. https://doi.org/10.1371/journal.pone.0227256.

ETHICAL CONSIDERATIONS

The study was developed through a literature review based exclusively on previously published documentary sources. It did not involve the direct participation of human or animal beings, the performance of interventions, the use of biological samples, or access to personally identifiable data. Consequently, it did not require approval from an ethics committee or the obtaining of informed consent. During the research process, the principles of scientific integrity, adequate attribution of authorship, rigorous citation, and responsible use of the sources consulted were observed.

ARTIFICIAL INTELLIGENCE (AI) USAGE STATEMENT

The authors state that the AI tool was used exclusively to support the organization of the article and the translation of the text. Its use did not determine the scientific content, study design, selection of sources, data analysis, interpretation of results, or conclusions. Full responsibility for the content remains with the authors.

FUNDING

The authors state that this study did not receive any external funding from public, private, or non-profit organizations. All research, analysis, and development activities were carried out with their own resources.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest related to this study and that all the procedures followed comply with the ethical standards established by the journal. They also confirm that this work is unpublished and has not been published, either partially or totally, in any other publication.

AUTHOR'S CONTRIBUTION

Authors' names and initials: Marco Antonio Ramos Espinal (MARE)
1. Conceptualization: MARE
2. Data Curation: MARE
3. Formal analysis: MARE
4. Acquisition of financing: not applicable
5. Research: MARE
6. Methodology: MARE
7. Project Management: MARE
8. Remedies: Not applicable
9. Software: Not applicable
10. Supervision: MARE
11. Validation: MARE
12. View: SEA
13. Original Draft: MOTHER
14. Writing, Proofreading, and Editing: MARE

SUPPLEMENTARY MATERIAL

Table S1.

Search strategies by database

Database	Strategy (Boolean operators, descriptors and free terms)	Filters applied
MEDLINE/PubMed	("meta-epidemiolog*"[tiab] OR "meta-research"[tiab] OR "ratio of ratios"[tiab] OR "ratio of odds ratios"[tiab] OR "ratio of risk ratios"[tiab] OR "target trial emulation"[tiab] OR "trial emulation"[tiab] OR "real-world evidence"[tiab]) AND ("Randomized Controlled Trials as Topic"[Mesh] OR "randomized controlled trial*"[tiab] OR "randomised controlled trial*"[tiab] OR "randomized trial*"[tiab]) AND ("Observational Studies as Topic"[Mesh] OR observational[tiab] OR "non-randomized"[tiab] OR nonrandomized[tiab] OR cohort[tiab] OR "study design*"[tiab]) AND (agreement[tiab] OR concordance[tiab] OR discrepanc*[tiab] OR "effect estimate*"[tiab] OR replicat*[tiab])	Publication date: 2020/01/01 to 2026/05/31
Embase (Elsevier)	('meta epidemiology'/exp OR 'meta-epidemiolog*':ti,ab OR 'ratio of ratios':ti,ab OR 'ratio of odds ratios':ti,ab OR 'target trial emulation':ti,ab OR 'real world evidence'/exp OR 'real-world evidence':ti,ab) AND ('randomized controlled trial (topic)'/exp OR 'randomi?ed controlled trial*':ti,ab OR 'randomi?ed trial*':ti,ab) AND ('observational study'/exp OR observational:ti,ab OR 'non-randomi?ed':ti,ab OR cohort:ti,ab) AND (agreement:ti,ab OR concordance:ti,ab OR discrepanc*:ti,ab OR 'effect estimate*':ti,ab)	[2020-2026]/py; publication date until May 2026
Scopus	TITLE-ABS-KEY(("meta-epidemiolog*" OR "ratio of ratios" OR "ratio of odds ratios" OR "target trial emulation" OR "real-world evidence") AND ("randomized controlled trial*" OR "randomised controlled trial*" OR "randomized trial*") AND (observational OR "non-randomized" OR nonrandomized OR cohort OR "study design*") AND (agreement OR concordance OR discrepanc* OR "effect estimate*"))	PUBYEAR > 2019 AND PUBYEAR < 2027; registrations through May 2026
Web of Science (Core Collection)	TS=(("meta-epidemiolog*" OR "ratio of ratios" OR "ratio of odds ratios" OR "target trial emulation" OR "real-world evidence") AND ("randomi?ed controlled trial*" OR "randomi?ed trial*") AND (observational OR "non-randomi?ed" OR cohort OR "study design*") AND (agreement OR concordance OR discrepanc* OR "effect estimate*"))	PY=2020-2026; registrations until May 2026

Cochrane Library	("meta-epidemiolog*" OR "ratio of ratios" OR "trial emulation" OR "real-world evidence"):ti,ab,kw AND (observational OR "non-randomized" OR cohort):ti,ab,kw AND (randomized OR randomised):ti,ab,kw	Published between January 2020 and May 2026; Cochrane Database of Systematic Reviews and Methodology Review Group
Complementary search (all bases)	("platform trial*" OR "adaptive trial*" OR "stepped wedge" OR "test-negative" OR "mendelian randomi?ation" OR "wastewater surveillance" OR "wastewater-based epidemiology" OR seroprevalence OR "genomic surveillance" OR "implementation science" OR "hybrid effectiveness-implementation" OR "community-based participatory research" OR "participatory action research" OR ("machine learning" AND "prediction model*")) AND ("systematic review" OR "meta-analysis" OR evaluation OR framework OR guidance)	Same period; field syntax adapted to each base

Note: these are the strings and interfaces declared in the manuscript; no exact date log, number of results per base, interface filters or original exports, necessary to repeat exactly the set of 1,287 records, were attached. [Mesh]: MeSH; /exp: Emtree; [tiab], :ti,ab and TITLE-ABS-KEY: search fields; TS: topic; * and ?: wildcards according to platform. The syntax must be checked in the corresponding interface. **Source.** Own elaboration.

Table S2.

Documentary base of the cells with score 3 in the regional application matrix (Figure 4)

Methodological outline	**Regions with a score of 3**	**Corpus Evidence**	**Sustenance**
Explanatory ECA	North America, Europe	Origin of data from meta-epidemiological studies (EPI) (5)(16)	Indirect
Pragmatic Testing / Platform	Europe	RECOVERY (UK) (4)	Direct
Conglomerates / stepped wedge	Sub-Saharan Africa, Asia	Methodological review without regional disaggregation (3)	Indirect
Cohort / case-control	North America, Europe, Asia	Sources of observational comparisons, without regional disaggregation (5)(15)	Indirect

Target Assay Emulation / RWE	North America; Europe	United States Complaint Bases (19)(21); Europe without its own study in the corpus	Direct (North America); indirect (Europe)
Test-negative design	North America; Europe	Methodological analysis from the United States (7); Europe without its own study in the corpus	Direct (North America); indirect (Europe)
Mendelian randomization	North America, Europe	Methodological review without regional disaggregation (8)	Indirect
Mixed methods/implementation	Sub-Saharan Africa	Systematic review of CFIR in LMICs (26)	Direct
IAP / CBPR (participatory)	North America, Latin America; Oceania	Meta-review of the United States (14); Guatemala Guideline and Critical Epidemiology (27)(28); Oceania without its own study in the corpus	Direct (North America, Latin America); indirect (Oceania)
Wastewater epidemiology	North America; Europe	Analysis from the United States (11); Europe without its own study in the corpus	Direct (North America); indirect (Europe)
Seroepidemiology	North America, Europe, Asia	WHO meta-analysis with estimates by region (9)	Direct
Genomic epidemiology	North America, Europe	Sequencing Disparities by Income Level (10)	Direct
Predictive Models (ML/AI)	North America, Europe, Asia	Systematic review without regional disaggregation (24)	Indirect

Note: Direct = study of the region or with results disaggregated by region; indirect = general origin, global initiatives or antecedents without particular regional measurement. Only 12 of the 28 cells with a score of 3 are directly supported; the other 16 do not demonstrate effective regional deployment. The scores correspond to the judgment of a single evaluator. Prepared by the authors.

Table S3.

Matrix of potential overlapping comparisons and inclusion decisions

Sources Compared	**Documentary relationship**	**Decision in the synthesis**	**Scope of Control**
Franklin (21) / Wang (19) y Heyard (20)	RCT-DUPLICATE publications with integrated pairs	Retain Heyard (20); omit the others from the aggregation	Known overlap; exclusion rule applied

Schwingshackl (17) y Stadelmaier (16)	Related Nutrition Comparisons	Retener Stadelmaier (16)	Overlap treated according to the declared selection
Bröckelmann (18) y Toews (5)	Comparisons from Cochrane Evidence	Retener Toews (5)	Overlap treated according to the declared selection
Toews (5) y Kimachi (15)	Possible reuse of comparisons in general review and nephrology	Both in published analysis; exclude both simultaneously in sensitivity	Residual dependence not ruled out; no trial list available for study-by-study matrix

https://doi.org/10.71068/hs000106