Available online **www.ijpras.com**

International Journal of Pharmaceutical Research & Allied Sciences, 2026, 15(1):33-43
https://doi.org/10.51847/4fl1lEm9bq

Review Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Do Comparative Efficacy Trials Still Resolve Meaningful Uncertainty in Biosimilar Development and Regulatory Decision-Making? A Critical Evidence Review

Aisha Bello[1]*, Zainab Sule[1], Ibrahim Musa[2], Grace Adeyemi[3], Ahmed Youssef[4]

[1]Department of Biosimilar Comparative Efficacy and Regulatory Decision, Faculty of Pharmacy, University of Ibadan, Ibadan, Nigeria.
[2]Department of Uncertainty Resolution in Trials, Faculty of Pharmacy, University of Ilorin, Ilorin, Nigeria.
[3]Department of Critical Evidence Review, Faculty of Pharmaceutical Sciences, Federal University of Agriculture Abeokuta, Abeokuta, Nigeria.
[4]Department of Biosimilar Development and Clinical Translation, Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan.

***Email:** aisha.bello@ui.edu.ng*

ABSTRACT

Comparative efficacy trials have historically served as visible clinical confirmation within biosimilar development, but their incremental scientific value has become increasingly contested as analytical characterization, functional testing, pharmacokinetic comparison, pharmacodynamic biomarkers, and immunogenicity assessment have become more sensitive. This critical evidence review evaluates whether comparative efficacy trials still resolve uncertainties that materially affect biosimilar development or regulatory decision-making. A transparent, question-led search and appraisal approach was used to prioritize literature addressing the totality of evidence, analytical similarity, reference-product variability, endpoint sensitivity, pharmacokinetic and immunogenicity uncertainty, and the consequences of retaining or waiving comparative efficacy testing. The central tension is that patient-level efficacy endpoints appear clinically direct but may be less capable than upstream methods of detecting or attributing small product-related differences. Disease heterogeneity, background treatment, saturated dose-response relationships, broad equivalence margins, endpoint measurement error, and reference-product variability can further reduce discrimination. Nevertheless, manufacturing-related immune risk, inadequately characterized functional differences, or uncertainty that cannot be resolved through analytical or clinical-pharmacology evidence may still justify targeted comparative clinical testing. The reviewed evidence supports neither automatic retention nor universal elimination of comparative efficacy trials. Their value depends on whether a clinically plausible and product-related uncertainty remains, whether the selected clinical model can detect its expected consequence, and whether the result could alter a development or regulatory conclusion. Comparative efficacy testing should therefore be treated as a conditional uncertainty-resolution tool rather than a routine final step. This judgement remains product-, mechanism-, assay-, population-, and regulatory-context dependent.

Key words: *Biosimilar, Biopharmaceuticals, Biosimilars, Comparability, Analytical similarity, Immunogenicity*
Received: 14 October 2025; **Revised:** 21 January 2026; **Accepted:** 27 January 2026

INTRODUCTION

Biosimilar development differs fundamentally from development of a new biological medicine. The objective is not to independently re-establish the reference product's benefit–risk profile but to demonstrate that the proposed

product is highly similar and has no clinically meaningful differences from an already characterized comparator. The European framework initially incorporated substantial comparative clinical testing because analytical methods, structure–function understanding, and regulatory experience were less mature. As these capabilities developed, the evidentiary emphasis shifted progressively toward analytical and functional characterization supported by targeted clinical pharmacology and immunogenicity assessment [1].

This evolution has intensified debate over the continuing role of comparative efficacy trials. Such studies are costly, lengthy, frequently conducted in patients receiving effective treatment, and often designed around equivalence rather than superiority. Their apparent clinical directness can obscure a more important question: whether the selected endpoint can sensitively detect a small product-related difference and attribute that difference to the biosimilar rather than to disease variability, background therapy, assay performance, or comparator fluctuation. Retrospective examination of development programs suggests that most comparative efficacy studies confirmed conclusions already supported by preceding evidence, whereas clinically consequential concerns were uncommon and often connected to product or process quality and immunogenicity [2].

A streamlined alternative does not imply that human evidence is unimportant. Rather, it proposes that each clinical component should answer a defined residual question. Comparative pharmacokinetic studies may identify exposure differences; pharmacodynamic measures may provide a proximal and sensitive comparison of biological activity; and targeted immunogenicity assessment may address specific immune risks. Under this model, a comparative efficacy trial is justified only when the preceding evidence leaves a clinically meaningful uncertainty that a sufficiently sensitive patient study can resolve [3]. Routine performance of a trial without such an uncertainty target risks generating confirmation without discrimination.

Regulatory decision analyses have further challenged the assumption that efficacy outcomes independently determine biosimilar acceptability. Publicly available application evidence indicates that major concerns tend to arise from quality, manufacturing, analytical, or pharmacokinetic findings rather than from an isolated failure to reproduce a reference product's efficacy estimate. Conversely, an apparently successful efficacy trial cannot compensate for unexplained analytical or functional differences. This review therefore critically examines when comparative efficacy trials change conclusions, when they add little incremental information, and how their value should be judged within an integrated totality of evidence [4].

Review question and decision context

The central review question is whether a comparative efficacy trial resolves meaningful uncertainty after analytical similarity, functional comparability, pharmacokinetic evidence, pharmacodynamic information where available, and immunogenicity risk have been considered. Meaningful uncertainty is defined here as uncertainty that is clinically plausible, remains unresolved by more direct evidence, is observable in the proposed clinical model, and could change a development or regulatory decision. This four-part definition is an original critical synthesis rather than a validated regulatory algorithm.

A comparative efficacy trial is a controlled clinical study intended to determine whether the biosimilar and reference product produce sufficiently similar efficacy outcomes within prespecified statistical margins. Its scientific value should not be inferred merely from the clinical importance of the endpoint. An endpoint can be clinically important yet insensitive for biosimilarity because it lies far downstream from the product attribute being compared. Disease heterogeneity, spontaneous variation, concomitant treatment, subjective assessment, delayed responses, and therapeutic dose saturation may dilute a small product-related difference that analytical or pharmacodynamic methods could detect more directly [3].

The decision context also requires separation of concepts that are often conflated. Analytical similarity concerns comparative structural, physicochemical, and functional attributes. Clinical comparability concerns whether any remaining difference has a relevant effect in humans. Indication extrapolation applies the totality of evidence and scientific justification across licensed uses rather than requiring independent efficacy trials in every population. Interchangeability concerns substitution or switching under a particular legal and regulatory framework. A comparative efficacy result may contribute to one element of this reasoning, but it does not independently establish all four.

Scientific uncertainty must also be distinguished from implementation uncertainty. Prescribers, patients, payers, and health systems may desire a familiar phase III trial because it appears reassuring. That preference may be understandable, but it does not demonstrate that the study is sensitive to the remaining product-related question. Regulatory analyses indicate that clinical findings require interpretation alongside the quality package and cannot

be treated as an independent evidentiary override [4]. A trial performed principally to communicate confidence should therefore not be misclassified as necessary uncertainty resolution.

Evidence-identification and critical-appraisal approach

This article uses a transparent, question-led critical-review design. The evidence-identification process focused on literature addressing comparative biosimilar efficacy trials, totality-of-evidence development, analytical and functional similarity, reference-product variability, critical quality attributes, pharmacokinetic and pharmacodynamic comparison, immunogenicity, switching, extrapolation, interchangeability, and tailored regulatory pathways. Searches were conducted across biomedical databases, publisher platforms, DOI-linked bibliographic records, and forward and backward citation pathways. The approach was designed to identify decision-relevant evidence and does not claim exhaustive systematic coverage.

Eligible sources were peer-reviewed journal articles with verifiable bibliographic records and direct relevance to at least one review question. Priority was given to primary analytical studies, randomized comparative trials, pharmacokinetic or pharmacodynamic investigations, immunogenicity analyses, regulatory dossier evaluations, and multicountry regulatory assessments. Reviews and perspectives were retained when they clarified field-level assumptions, historical rationale, methodological controversy, or stakeholder and policy boundaries. Sources were excluded when they provided only general biosimilar commentary, lacked an extractable connection to the decision question, or would require interpretation beyond the product, population, method, or regulatory context examined.

The appraisal compared each source according to the uncertainty addressed, proximity of the measured endpoint to the suspected product difference, analytical or clinical sensitivity, capacity for causal attribution, internal validity, external validity, and relevance to an actual decision. Particular attention was given to contradictions between analytical findings and clinical outcomes, differences caused by reference-product drift or non-product variability, and cases in which a trial was statistically inconclusive without demonstrating clinically meaningful dissimilarity. The synthesis is qualitative and interpretive; it does not pool treatment effects, assign numerical quality scores, or estimate the probability that a future efficacy trial will change a regulatory conclusion.

Historical rationale for comparative efficacy trials

The historical rationale for comparative efficacy trials emerged from the need to support an abbreviated but credible development pathway. Because a biosimilar relies on the established clinical knowledge of its reference product, the development program is comparative and inferential. Early regulators nevertheless required substantial clinical confirmation because small structural differences in complex proteins could not always be fully characterized, and the relationship between measured quality attributes and clinical outcomes was incompletely understood. Comparative trials therefore functioned as a precautionary final layer rather than as independent demonstrations of therapeutic efficacy [5].

Advances in risk-based characterization changed this evidentiary balance. Critical quality attributes can be selected according to their potential influence on biological activity, pharmacokinetics, safety, and immunogenicity, and analytical methods can be prioritized according to sensitivity and uncertainty. This tiered approach does not eliminate judgement: attribute criticality, acceptable variability, and clinical relevance remain context dependent. It does, however, allow the development program to investigate product differences directly rather than waiting for a downstream patient endpoint to reveal an unexplained consequence [6].

The limitations of the historical model are particularly visible in oncology. Comparative studies may use pathological response, progression-related measures, or other accepted outcomes in patients receiving combination regimens. Such endpoints can be clinically meaningful but may be influenced by tumour heterogeneity, diagnostic assessment, concomitant agents, treatment timing, and high-dose regimens operating near the plateau of the exposure–response curve. These features can reduce assay sensitivity and make it difficult to determine whether an observed difference is attributable to the biological product itself [7].

The increasing sensitivity of analytical and clinical-pharmacology methods therefore prompted calls to end routine phase III biosimilar trials. That position captured a genuine methodological concern: a large patient study should not be assumed to add value merely because it is clinically familiar. Yet the categorical removal of efficacy testing would also be difficult to defend when a clinically plausible uncertainty remains unresolved or when no sensitive pharmacodynamic alternative exists. The historical debate is best reframed not as “clinical trials versus no clinical trials,” but as a requirement that every comparative trial identify the specific uncertainty it is expected to resolve and demonstrate that its design can detect the corresponding difference [8].

Analytical and functional evidence preceding clinical trials

Modern similarity assessment begins with the premise that clinically relevant effects originate from product attributes and biological functions. Longitudinal analyses of trastuzumab reference lots identified shifts in Fc-related attributes and antibody-dependent cellular cytotoxicity, demonstrating that the comparator itself may change within its approved lifecycle [9]. A clinical trial conducted against a limited set of contemporary lots may therefore compare products without revealing how the reference-product target evolved.

Characterization of the trastuzumab biosimilar SB3 further showed the importance of integrating structural, binding, and functional evidence with reference-lot history. Differences observed across comparator lots could be distinguished from candidate-specific deviations only because the analytical program examined the relevant quality attributes directly [10]. Without this context, a downstream efficacy difference might be incorrectly attributed to the biosimilar or dismissed as clinical noise.

Orthogonal analytical methods can interrogate molecular structure, receptor binding, Fc-mediated activity, purity, aggregation, and other attributes through independent measurement principles. The HLX02 program illustrated how sensitive FcγRIIIa affinity and functional assays can localize differences that broad patient outcomes may dilute [11]. Such methods do not prove that every detected difference has a clinical consequence, but they clarify its mechanism and potential relevance.

Recent approval experience indicates that successful monoclonal-antibody programs commonly rely on convergent structural and functional methods rather than a single similarity test [12]. Across trastuzumab case studies, orthogonal analytical and functional methods exposed reference-product drift and structure–function differences that aggregate clinical efficacy endpoints were poorly positioned to localize [9-11]. Analytical evidence is therefore strongest for detection and attribution; clinical relevance must still be evaluated through mechanism, exposure, immunogenicity, and residual-risk reasoning.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for contested claims, hidden assumptions, and evidence standards in do comparative efficacy trials still resolve meaningful uncertainty in biosimilar.

Table 1. Contested claims, hidden assumptions, and evidence standards in Do Comparative Efficacy Trials Still Resolve Meaningful Uncertainty in Biosimilar.

Evidence domain or review construct	Review question	Eligible evidence type	Key extraction fields	Appraisal or relevance criterion	Synthesis role	Main limitation	Interpretation boundary
Reference-product variability	Is the comparator a stable analytical target?	Longitudinal lot characterization	Attribute drift, functional change, lot period	Ability to distinguish comparator evolution from candidate deviation	Tests a hidden assumption underlying clinical comparison	Product-specific evidence	Analytical drift does not alone establish patient harm [9].
Critical quality attributes	Which differences could plausibly affect clinical performance?	Risk-based analytical methodology	Attribute criticality, method sensitivity, mechanism	Biological relevance and orthogonality of methods	Connects analytical difference to residual uncertainty	Criticality remains partly judgement based	A measured difference is not automatically clinically meaningful [6].
Orthogonal analytical similarity	Can upstream methods detect differences more sensitively?	Structural, physicochemical, binding, and functional studies	Assay principle, sensitivity, variability, reference-lot range	Directness and ability to localize mechanism	Establishes the primary evidence base	Public packages may not disclose all methods	Analytical similarity cannot predict every clinical consequence [11].
Clinical efficacy confirmation	Does a patient endpoint add an independent discriminatory signal?	Comparative efficacy trial	Population, dose, endpoint, margin, background therapy	Assay sensitivity and decision-changing potential	Tests incremental value	Downstream variability may mask small differences	Clinical importance does not guarantee discriminatory sensitivity [7].
Immunogenicity risk	Could product or process	Aggregate, impurity, ADA,	Assay tolerance,	Product attribution	Identifies uncertainty	Rare events are difficult	Broad efficacy trials are not

	differences provoke a clinically relevant immune response?	neutralizing-antibody, and mechanistic studies	timing, titre, mechanism, exposure	and clinical plausibility	that may justify targeted human evidence	to study prospectively	automatically sensitive to immune risk [13].
Proposed meaningful-uncertainty test	Is the uncertainty plausible, unresolved, observable, and decision-changing?	Integrated review synthesis	Source of uncertainty, expected consequence, sensitive model, decision effect	All four conditions must be justified	Proposed rule for trial justification	Not externally validated	This construct is an original review synthesis, not a regulator-approved standard.

Sensitivity and statistical limitations of efficacy endpoints

A patient efficacy endpoint is often clinically recognizable but biologically distant from the quality attribute under comparison. Pharmacodynamic biomarkers may provide a more proximal measure when they reflect the product's mechanism and respond across a suitable exposure range [14]. Their potential advantage is not convenience alone; it is the ability to compare biological activity before disease variability and background treatment obscure a modest difference.

Regulatory data analyses support tailoring the clinical program according to what remains unresolved after quality and pharmacokinetic assessment [15]. If exposure is comparable and no functionally relevant difference remains, an additional efficacy study may have little independent capacity to change the conclusion. Conversely, unexplained exposure or functional findings should be investigated directly rather than averaged within a noisy patient endpoint.

Replacing efficacy outcomes with pharmacodynamic evidence requires more than identifying a measurable biomarker. The marker must be mechanistically relevant, technically reproducible, sufficiently responsive, and interpretable across the selected dose and sampling interval [16]. A biomarker measured on a response plateau may be no more sensitive than the efficacy endpoint it replaces.

A randomized PCSK9-inhibitor investigation demonstrated how careful dose selection and a responsive pharmacodynamic measure can magnify potential differences under controlled conditions [17]. Conventional efficacy trials may instead lose sensitivity through therapeutic-dose saturation, spontaneous disease variation, concomitant medication, endpoint misclassification, uncertain historical effects, or equivalence margins that tolerate differences larger than those detectable analytically. Statistical equivalence under such conditions establishes compatibility with the prespecified design, not complete biological identity.

Cases in which trials changed development conclusions

Clinically consequential product-related risk remains possible. Investigations of tungsten-induced erythropoietin aggregates connected a manufacturing-associated product feature with T-cell activation and pure red cell aplasia [13]. This case supports rigorous process control and targeted immune-risk investigation, but it does not establish that a routine efficacy-equivalence trial would be the most sensitive method for detecting such a problem.

The opposite attribution error is also important. Two pharmacokinetic studies of GP2017 reported differences in antidrug-antibody findings that were associated with non-product factors, including variation between study settings [18]. An immunogenicity imbalance should therefore trigger investigation of assay performance, sample handling, population characteristics, exposure, and study conduct before it is interpreted as evidence of biosimilar dissimilarity.

The SB3 comparative trial met its principal efficacy objective, but interpretation occurred within a program complicated by reference-product evolution and variability in pathological response assessment [19]. The trial supplied clinical confirmation, yet the analytical program was more capable of identifying the changing comparator characteristics that informed the biological explanation.

In the LILAC study, conclusions appeared differently when treatment effects were expressed through risk-ratio and risk-difference frameworks [20]. This does not establish clinically meaningful inferiority of the biosimilar. It illustrates that endpoint scale, estimand, margin construction, and random variation can generate an apparently concerning result without identifying a product-related mechanism. Trials change development conclusions most

credibly when the finding is reproducible, mechanistically plausible, attributable to the candidate, and consequential for the integrated evidence package.

Cases in which trials added little discriminatory information

The NOR-SWITCH trial provided controlled evidence that switching from reference infliximab to CT-P13 was not inferior to continued reference treatment under the study design [21]. Its pooled indications and limited capacity for rare or indication-specific outcomes nevertheless restricted the conclusions that could be drawn about every switching context or product-quality risk.

A randomized transition study involving SB2 similarly found no material efficacy, safety, or immunogenicity signal following transition from reference infliximab [22]. The study strengthened clinical reassurance, but absence of a detected difference did not independently explain why the products should perform similarly; that explanation remained grounded in the preceding comparability evidence.

The broader early switching literature was generally reassuring, although study designs, populations, follow-up durations, switch patterns, and immunogenicity methods varied substantially [23]. Such evidence is relevant to implementation and interchangeability debates but should not be treated as proof that every comparative clinical study is sensitive to rare, delayed, or mechanism-specific differences.

An updated synthesis likewise found no consistent loss of efficacy, safety, or immunogenicity after switching while preserving uncertainty concerning repeated switches and uncommon outcomes [24]. Controlled switching studies and the accumulated literature generally did not identify major losses of efficacy, safety, or immunogenicity after a single reference-to-biosimilar switch, although sensitivity to rare outcomes remained limited [21-23]. These studies supplied reassurance, but reassurance, discrimination, causal attribution, and rare-event detection are distinct evidentiary functions.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for evidence supporting and challenging major claims, including methodological weaknesses.

Table 2. Evidence supporting and challenging major claims, including methodological weaknesses.

Method, platform, or evidence category	Representative application	Reported strength	Validation context	Generalization concern	Contradictory evidence	Evidence gap	Review interpretation
Longitudinal analytical characterization	Reference trastuzumab lot drift	Detects temporal attribute and functional changes [9]	Product-specific lot series	Drift patterns may differ across products	A clinical consequence was not directly measured	Limited public longitudinal datasets	Strong for detecting comparator evolution, not for quantifying patient risk.
Orthogonal functional testing	Fc-receptor binding and ADCC assessment	Localizes mechanism-related differences [11]	Trastuzumab biosimilar program	Not every clinical pathway has a sensitive assay	Downstream effects may involve unmeasured biology	External validation across modalities	Usually more diagnostic than aggregate efficacy endpoints.
Comparative efficacy trial	SB3 and ABP 980 oncology studies	Provides patient-level outcome comparison [19]	Controlled phase III settings	Combination therapy and endpoint variability may dilute differences	Estimand-dependent interpretation occurred in LILAC [20]	Direct measures of incremental decision value	Useful only when the clinical model is sufficiently sensitive.
Switching trial	NOR-SWITCH and SB2 transition	Provides controlled reassurance [21]	Patients receiving established therapy	Limited rare-event and repeated-switch sensitivity	No consistent major signal in later synthesis [24]	Long-term and complex-switch evidence	Reassurance should not be equated with complete discrimination.
Immunogenicity investigation	EPO aggregate mechanism	Links a process feature to clinically consequential	Mechanistic and clinical immune context	Extreme case may not transfer to other biologics	ADA imbalances may be non-product-related [18]	Sensitive prospective methods for rare events	Targeted mechanistic assessment may outperform

		immunity [17]					broad efficacy testing.
Pharmacodynamic comparison	Mechanism-linked biomarker studies	Can provide proximal, responsive evidence [16]	Controlled clinical-pharmacology design	Suitable biomarkers are not universally available	Variability or dose selection may undermine sensitivity	Qualification across products and populations	A strong replacement option when biological and technical criteria are met.

Residual uncertainty in PK and immunogenicity evidence

Pharmacokinetic comparison can be highly sensitive because exposure integrates absorption, distribution, clearance, target-mediated disposition, and immune effects. Regulatory experience with monoclonal antibodies and fusion proteins nevertheless shows that immunogenicity must be interpreted across the complete evidence package and product lifecycle [25]. A conventional efficacy trial may be too small, too short, or too indirect to identify a rare immune-mediated event.

A case analysis of ustekinumab biosimilars concluded that analytical evidence and single-dose pharmacokinetic studies provided sufficient information to establish comparable immunogenicity in that product context [26]. The conclusion is informative but not universal. Its transferability depends on assay sensitivity, exposure, mechanism, observed ADA patterns, process knowledge, and the absence of an unresolved product-specific immune risk.

Pharmacodynamic evidence also has limitations. A randomized study of IL-5 antagonists showed that biological relevance alone does not guarantee a sensitive biosimilarity endpoint; variability, dose selection, response range, and sampling strategy remain decisive [27]. Failure of a biomarker strategy should lead to redesign or alternative evidence, not automatic assumption that a large efficacy trial will be more informative.

Interferon beta-1a provides a contrasting example in which responsive pharmacodynamic markers supported a controlled comparison [28]. Residual uncertainty therefore depends on the product and model. Drug tolerance of ADA assays, timing of sample collection, neutralizing activity, low event frequency, healthy-volunteer transportability, and delayed immune outcomes must be considered explicitly. PK and targeted immunogenicity evidence may be sufficient, but only when their sensitivity matches the uncertainty being addressed.

Decision framework for retaining, replacing, or waiving trials

Regulatory assessment packages already vary in the nature and scale of comparative clinical evidence submitted, indicating that no single clinical template is scientifically necessary for every biosimilar [29]. Variation alone does not prove that every reduced program is optimal, but it demonstrates that product characteristics and residual evidence gaps influence clinical-development expectations.

A developer-oriented framework proposes that trial requirements should follow residual uncertainty rather than a fixed sequence of development phases [30]. This principle is scientifically coherent, although developer perspectives may also reflect incentives to reduce development cost and time. The argument must therefore be assessed against independent analytical, regulatory, and clinical-pharmacology evidence.

Contemporary expectations for tailored development emphasize that waiver is not evidentiary relaxation. A reduced program requires convincing analytical similarity, functionally relevant methods, adequate pharmacokinetic evidence, appropriate immunogenicity assessment, and a transparent explanation of why remaining differences lack clinically meaningful consequences [31].

Scientific arguments for waiving routine efficacy testing are strongest when the proposed trial lacks assay sensitivity and cannot alter the integrated conclusion [32]. Waiver is weaker when an unexplained functional difference persists, a process-related immune risk is plausible, PK comparability is uncertain, or no alternative method can observe the expected clinical consequence.

Recent assessment, developer, and regulator-facing syntheses converge on a tailored residual-uncertainty approach rather than an automatic comparative efficacy requirement [29-31]. The proposed review framework therefore asks four sequential questions: Does meaningful uncertainty remain? Is it plausibly product-related? Can a sensitive clinical model detect it? Could the result change a decision? Failure at any gate favors targeted analytical, manufacturing, PK, PD, immunogenicity, or lifecycle evidence rather than a broad efficacy trial.

Figure 1 presents the original conceptual synthesis for decision framework for the value of comparative biosimilar efficacy trials, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

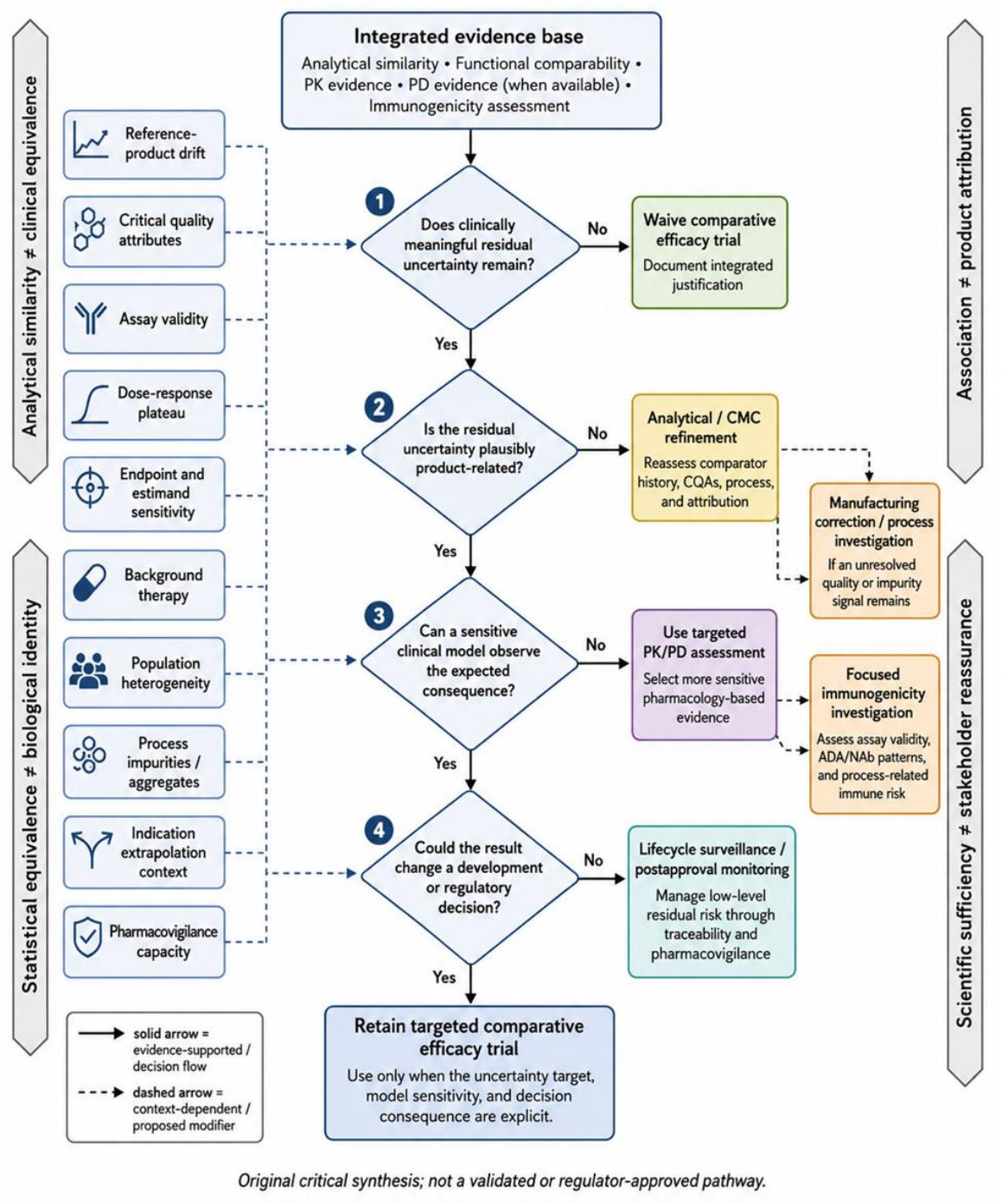

Figure 1. Decision Framework for the Value of Comparative Biosimilar Efficacy Trials

Counterarguments, boundary conditions, and policy implications

A prominent counterargument is that comparative efficacy trials support prescriber and patient confidence even when their scientific discrimination is limited. Prescriber discussions confirm that clinical familiarity, switching concerns, and uncertainty about regulatory reasoning can influence adoption [33]. These concerns deserve transparent communication, but stakeholder reassurance is not equivalent to unresolved product-related uncertainty.

A second boundary concerns unequal regulatory capacity. Jurisdictions differ in analytical expertise, access to appropriate reference products, manufacturing oversight, pharmacovigilance, traceability, and reliance mechanisms [34]. A streamlined pathway that is credible in a mature regulatory system may be difficult to implement safely where these supporting functions are incomplete.

International guidance has moved toward greater flexibility while preserving the core requirement for robust comparability and lifecycle oversight [35]. Flexibility should therefore mean product-specific evidence selection, not reduced scrutiny. Waiving an insensitive efficacy trial may require stronger analytical review, clearer residual-risk justification, and more reliable postapproval monitoring.

Policy should separate three functions that are frequently combined: establishing biosimilarity, managing residual risk, and communicating confidence. Insensitive trials should not be used as communication substitutes when better explanation of analytical science, regulatory standards, traceability, and pharmacovigilance is available. Conversely, clinical evidence should be retained when a specific uncertainty cannot be addressed reliably through other methods.

Review limitations

This review used a transparent critical search but did not claim exhaustive systematic coverage. The selection prioritized decision-relevant literature, and additional product-specific studies may provide different examples without altering the central conceptual distinction between confirmation and uncertainty resolution.

The evidence base is heterogeneous. Analytical studies, randomized efficacy trials, switching investigations, pharmacodynamic studies, regulatory analyses, reviews, and stakeholder perspectives answer different questions. Their findings cannot be combined as though they measured one common outcome.

CONCLUSION

Comparative efficacy trials still can resolve meaningful uncertainty, but only under restricted conditions. Their value does not arise automatically from measuring a patient outcome; it depends on whether an unresolved and clinically plausible product-related difference remains, whether the selected model can detect its consequence, and whether the result could change a decision. Modern analytical, functional, pharmacokinetic, and qualified pharmacodynamic methods often provide more sensitive and causally interpretable evidence than broad efficacy endpoints. A successful efficacy-equivalence result cannot rescue inadequate analytical similarity, unexplained manufacturing or functional differences, or failed pharmacokinetic comparison. Equally, universal waiver would be premature where process-related immune risk, uncertain mechanism, inadequate biomarkers, or an observable clinical consequence remains unresolved. The appropriate regulatory question is therefore not whether comparative efficacy trials are intrinsically necessary, but whether a particular trial has a defined uncertainty target and credible discriminatory capacity. Retaining, replacing, or waiving the trial should follow that product-specific judgement.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Wolff-Holz E, Tiitso K, Vleminckx C, Weise M. Evolution of the EU biosimilar framework: Past and future. BioDrugs. 2019;33(6):621-34. doi:10.1007/s40259-019-00377-y
2. Schiestl M, Ranganna G, Watson K, Jung B, Roth K, Capsius B, et al. The path towards a tailored clinical biosimilar development. BioDrugs. 2020;34(3):297-306. doi:10.1007/s40259-020-00422-1
3. Bielsky MC, Cook A, Wallington A, Exley A, Kauser S, Hay JL, et al. Streamlined approval of biosimilars: Moving on from the confirmatory efficacy trial. Drug Discov Today. 2020;25(11):1910-8. doi:10.1016/j.drudis.2020.09.006
4. Kirsch-Stefan N, Guillen E, Ekman N, Barry S, Knippel V, Killalea S, et al. Do the outcomes of clinical efficacy trials matter in regulatory decision-making for biosimilars? BioDrugs. 2023;37(6):855-71. doi:10.1007/s40259-023-00631-4
5. Webster CJ, Wong AC, Woollett GR. An efficient development paradigm for biosimilars. BioDrugs. 2019;33(6):603-11. doi:10.1007/s40259-019-00371-4

6. Vandekerckhove K, Seidl A, Gutka H, Kumar M, Gratzl G, Keire D, et al. Rational selection, criticality assessment, and tiering of quality attributes and test methods for analytical similarity evaluation of biosimilars. AAPS J. 2018;20(4):68. doi:10.1208/s12248-018-0230-9
7. Stebbing J, Mainwaring PN, Curigliano G, Pegram M, Latymer M, Bair AH, et al. Understanding the role of comparative clinical studies in the development of oncology biosimilars. J Clin Oncol. 2020;38(10):1070-80. doi:10.1200/JCO.19.02953
8. Frapaise FX. The end of phase 3 clinical trials in biosimilars development? BioDrugs. 2018;32(4):319-24. doi:10.1007/s40259-018-0287-0
9. Kim S, Song J, Park S, Ham S, Paek K, Kang M, et al. Drifts in ADCC-related quality attributes of Herceptin®: Impact on development of a trastuzumab biosimilar. MAbs. 2017;9(4):704-14. doi:10.1080/19420862.2017.1305530
10. Lee JH, Paek K, Moon JH, Ham S, Song J, Kim S. Biological characterization of SB3, a trastuzumab biosimilar, and the influence of changes in reference product characteristics on the similarity assessment. BioDrugs. 2019;33(4):411-22. doi:10.1007/s40259-019-00362-5
11. Xie L, Zhang E, Xu Y, Gao W, Wang L, Xie MH, et al. Demonstrating analytical similarity of trastuzumab biosimilar HLX02 to Herceptin® with a panel of sensitive and orthogonal methods including a novel FcγRIIIa affinity chromatography technology. BioDrugs. 2020;34(3):363-79. doi:10.1007/s40259-020-00407-0
12. Aithal R, Majedi OM, McCarthy D, Atouf F, Peckham N. Review of quality attributes and analytical methods used for comparative analytical assessment of monoclonal antibodies as part of successful biosimilar approvals in the United States and European Union. BioDrugs. 2026;40(4):713-32. doi:10.1007/s40259-026-00786-w
13. Rubic-Schneider T, Kuwana M, Christen B, Aßenmacher M, Hainzl O, Zimmermann F, et al. T-cell assays confirm immunogenicity of tungsten-induced erythropoietin aggregates associated with pure red cell aplasia. Blood Adv. 2017;1(6):367-79. doi:10.1182/bloodadvances.2016001842
14. Li J, Florian J, Campbell E, Schrieber SJ, Bai JPF, Weaver JL, et al. Advancing biosimilar development using pharmacodynamic biomarkers in clinical pharmacology studies. Clin Pharmacol Ther. 2020;107(1):40-2. doi:10.1002/cpt.1653
15. Guillen E, Ekman N, Barry S, Weise M, Wolff-Holz E. A data driven approach to support tailored clinical programs for biosimilar monoclonal antibodies. Clin Pharmacol Ther. 2023;113(1):108-23. doi:10.1002/cpt.2785
16. Strauss DG, Wang YM, Florian J, Zineh I. Pharmacodynamic biomarkers evidentiary considerations for biosimilar development and approval. Clin Pharmacol Ther. 2023;113(1):55-61. doi:10.1002/cpt.2761
17. Sheikhy M, Schrieber SJ, Sun Q, Gershuny V, Matta MK, Bai JPF, et al. Considerations for use of pharmacodynamic biomarkers to support biosimilar development—(I) A randomized trial with PCSK9 inhibitors. Clin Pharmacol Ther. 2023;113(1):71-9. doi:10.1002/cpt.2769
18. von Richter O, Lemke L, Haliduola H, Balfour A, Zehnpfennig B, Skerjanec A, et al. Differences in immunogenicity associated with non-product related variability: Insights from two pharmacokinetic studies using GP2017, an adalimumab biosimilar. Expert Opin Biol Ther. 2019;19(10):1057-64. doi:10.1080/14712598.2019.1603959
19. Pivot X, Bondarenko I, Nowecki Z, Dvorkin M, Trishkina E, Ahn JH, et al. Phase III, randomized, double-blind study comparing the efficacy, safety, and immunogenicity of SB3 (trastuzumab biosimilar) and reference trastuzumab in patients treated with neoadjuvant therapy for human epidermal growth factor receptor 2-positive early breast cancer. J Clin Oncol. 2018;36(10):968-74. doi:10.1200/JCO.2017.74.0126
20. von Minckwitz G, Colleoni M, Kolberg HC, Morales S, Santi P, Tomasevic Z, et al. Efficacy and safety of ABP 980 compared with reference trastuzumab in women with HER2-positive early breast cancer (LILAC study): A randomised, double-blind, phase 3 trial. Lancet Oncol. 2018;19(7):987-98. doi:10.1016/S1470-2045(18)30241-9
21. Jørgensen KK, Olsen IC, Goll GL, Lorentzen M, Bolstad N, Haavardsholm EA, et al. Switching from originator infliximab to biosimilar CT-P13 compared with maintained treatment with originator infliximab (NOR-SWITCH): A 52-week, randomised, double-blind, non-inferiority trial. Lancet. 2017;389(10086):2304-16. doi:10.1016/S0140-6736(17)30068-5
22. Smolen JS, Choe JY, Prodanovic N, Niebrzydowski J, Staykov I, Dokoupilova E, et al. Safety, immunogenicity and efficacy after switching from reference infliximab to biosimilar SB2 compared with continuing reference infliximab and SB2 in patients with rheumatoid arthritis: Results of a randomised,

double-blind, phase III transition study. Ann Rheum Dis. 2018;77(2):234-40. doi:10.1136/annrheumdis-2017-211741

23. Cohen HP, Blauvelt A, Rifkin RM, Danese S, Gokhale SB, Woollett GR. Switching reference medicines to biosimilars: A systematic literature review of clinical outcomes. Drugs. 2018;78(4):463-78. doi:10.1007/s40265-018-0881-y
24. Barbier L, Ebbers HC, Declerck P, Simoens S, Vulto AG, Huys I. The efficacy, safety, and immunogenicity of switching between reference biopharmaceuticals and biosimilars: A systematic review. Clin Pharmacol Ther. 2020;108(4):734-55. doi:10.1002/cpt.1836
25. Kurki P, Barry S, Bourges I, Tsantili P, Wolff-Holz E. Safety, immunogenicity and interchangeability of biosimilar monoclonal antibodies and fusion proteins: A regulatory perspective. Drugs. 2021;81(16):1881-96. doi:10.1007/s40265-021-01601-2
26. Schiestl M, Roy N, Trieb M, Park JP, Guillen E, Woollett G, et al. Analytical data and single-dose PK are sufficient to conclude comparable immunogenicity for biosimilars: An ustekinumab case study. BioDrugs. 2025;39(5):769-76. doi:10.1007/s40259-025-00733-1
27. Gershuny V, Sun Q, Schrieber SJ, Matta MK, Weaver JL, Ji P, et al. Considerations for use of pharmacodynamic biomarkers to support biosimilar development—(II) A randomized trial with IL-5 antagonists. Clin Pharmacol Ther. 2023;113(1):80-9. doi:10.1002/cpt.2760
28. Florian J, Gershuny V, Sun Q, Schrieber SJ, Matta MK, Hazel A, et al. Considerations for use of pharmacodynamic biomarkers to support biosimilar development—(III) A randomized trial with interferon beta-1a products. Clin Pharmacol Ther. 2023;113(2):339-48. doi:10.1002/cpt.2784
29. Kuribayashi R, Hariu A, Nakano A, Kishioka Y. Survey of data package and sample size of comparative clinical studies for biosimilar developments from PMDA assessments. Pharmaceut Med. 2024;38(3):225-39. doi:10.1007/s40290-024-00525-y
30. Cohen HP, Turner M, McCabe D, Woollett GR. Future evolution of biosimilar development by application of current science and available evidence: The developer's perspective. BioDrugs. 2023;37(5):583-93. doi:10.1007/s40259-023-00619-0
31. Guillen E, Barry S, Jost N, Ekman N, Knippel V, Kuhlmann-Gottke J, et al. The tailored biosimilar approach: Expectations and requirements. Drugs. 2025;85(5):601-8. doi:10.1007/s40265-025-02168-y
32. Niazi SK. Scientific rationale for waiving clinical efficacy testing of biosimilars. Drug Des Devel Ther. 2022;16:2803-15. doi:10.2147/DDDT.S378813
33. Shubow S, Sun Q, Nguyen Phan AL, Hammell DC, Kane M, Lyman GH, et al. Prescriber perspectives on biosimilar adoption and potential role of clinical pharmacology: A workshop summary. Clin Pharmacol Ther. 2023;113(1):37-49. doi:10.1002/cpt.2765
34. Kang HN, Thorpe R, Knezevic I, Casas Levano M, Chilufya MB, Chirachanakul P, et al. Regulatory challenges with biosimilars: An update from 20 countries. Ann N Y Acad Sci. 2021;1491(1):42-59. doi:10.1111/nyas.14522
35. Kang HN, Wadhwa M, Knezevic I, Ondari C, Simao M. WHO guidelines on biosimilars: Toward improved access to safe and effective products. Ann N Y Acad Sci. 2023;1521(1):96-103. doi:10.1111/nyas.14965