



Original Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Causal Inference Can Reshape Post-Marketing Safety Assessment Without Replacing Clinical Judgment

Charlotte Dupuis^{1*}, Hugo Perrin¹, Elise Morel², Thomas Martin³

¹Department of Causal Inference and Post-Marketing Safety, Faculty of Pharmacy, University of Lille, Lille, France.

²Department of Drug Safety Assessment, Faculty of Pharmacy, University of Montpellier, Montpellier, France.

³Department of Clinical Judgment and Pharmacovigilance, Faculty of Pharmacy, University of Nantes, Nantes, France.

*Email: charlotte.dupuis@univ-lille.fr

ABSTRACT

Post-marketing drug safety assessment begins in an evidential environment that is exceptionally valuable for detecting unexpected harm but poorly suited to automatic causal interpretation. Spontaneous reports, electronic health records, claims databases, registries, and other real-world sources can reveal safety patterns that were not detectable before approval, yet each source embeds distinct reporting processes, treatment-selection mechanisms, temporal structures, measurement limitations, and clinical uncertainties. Causal inference can reshape this landscape by forcing safety questions to be specified more precisely. Target-trial thinking clarifies eligibility, treatment strategies, time zero, follow-up, outcomes, and estimands; active and clinically credible comparators improve treatment-group comparability; longitudinal methods address treatment switching and changing risk; negative controls can expose systematic error; and quantitative bias analysis makes assumptions about unmeasured bias explicit. None of these methods, however, converts observational safety evidence into causal certainty by technique alone. Post-marketing decisions also depend on event phenotype, seriousness, reversibility, alternative explanations, individual case chronology, biological plausibility, signal maturity, and regulatory context. This Current Opinion argues that causal inference should therefore function as a disciplined layer of safety assessment rather than as a replacement for clinical judgment. The most defensible architecture is one in which statistical signals generate causal questions, causal methods test increasingly explicit counterfactual contrasts, and clinical reasoning remains independently responsible for dimensions that population-level estimation cannot fully encode.

Keywords: Pharmacovigilance, Causal inference, Real-world evidence, Target trial emulation, Pharmacoepidemiology, Quantitative bias analysis

Received: 25 January 2026; **Revised:** 21 March 2026; **Accepted:** 23 March 2026

INTRODUCTION

Post-marketing safety assessment is often asked to perform two jobs at once. It must detect unexpected patterns early enough to protect patients, yet it must also distinguish genuine treatment-related harm from the enormous background variation produced by disease, comorbidity, treatment choice, reporting behavior, healthcare utilization, and chance. These jobs require different evidential standards. Disproportionality analysis can be highly useful for signal detection, but its interpretation depends on transparent specification of data handling, case selection, comparator construction, statistical methods, and reporting choices. READUS-PV was developed precisely because the credibility and reproducibility of such analyses depend on making these decisions visible

[1]. Transparency is therefore a prerequisite for inference, not evidence that an observed reporting association is itself causal.

The interpretive problem becomes more consequential when cautious signal language is converted into causal language without an accompanying increase in evidence strength. A 2025 meta-epidemiological study showed that causal statements in highly visible pharmacovigilance disproportionality publications could propagate into stronger causal wording in later citations [2]. This matters because the semantics used around a signal can change how clinicians, regulators, researchers, and patients perceive evidential maturity. An association may be clinically important long before its causal structure is resolved, but urgency does not erase the distinction between detection and explanation.

Real-world evidence offers the possibility of moving beyond detection toward explicitly comparative questions, provided that design, measurement, and decision context are aligned. Contemporary pharmacoepidemiology increasingly emphasizes that credible RWE depends not simply on data volume but on principled choices about the causal question, study design, data fitness, measurement, analysis, and interpretation [3]. The argument developed here is consequently not that causal inference should replace pharmacovigilance. Rather, it should reshape what happens after an association becomes important enough to interrogate. The key transformation is from asking whether a drug is “associated with” an event to asking which causal contrast is clinically meaningful, whether the available data can identify it, which biases threaten the comparison, how those biases can be diagnosed or quantified, and which aspects of the final safety judgment remain irreducibly clinical.

Association signals and causal questions

A disproportionality signal is an asymmetric pattern in reporting, not automatically an estimate of excess clinical risk. Formal work examining the relationship between disproportionality measures and causal rate ratios illustrates how demanding the necessary assumptions can be. Confounding, differential reporting, and differences in the probability that events become observable in spontaneous-report systems can break the correspondence between a reporting ratio and the causal contrast that decision makers actually want [4]. The appropriate response is not to dismiss disproportionality analysis. It is to recognize its principal inferential role as generating, prioritizing, and characterizing questions that may require stronger designs.

Even before confounding is considered, the underlying case stream can contain structural complications. Individual case safety reports may be transmitted repeatedly across organizations and jurisdictions, creating multiple case versions and recipient-specific information states. A multicompany investigation documented extensive replication across ICSR transmission networks and showed that the same underlying safety information may not be represented identically for every recipient [5]. This creates an important distinction between a pharmacological phenomenon and a reporting-system phenomenon. Repeated transmissions do not necessarily create a false signal, but provenance, deduplication, version control, and completeness can influence the evidential substrate from which a signal is derived. Causal assessment should therefore begin only after the upstream data-generating process has been considered.

The same discipline applies when structured healthcare data are introduced. A real-world database does not become a neutral causal laboratory merely because it provides denominators. Selection of drugs, outcomes, reference pairs, and analytic designs must be aligned with the question being asked. Work developing a taxonomy-based framework for real-world signal detection showed that the suitability of drug–outcome reference sets depends on the design used to evaluate them [6]. A control pair appropriate for benchmarking one method may be inappropriate for another. Likewise, an outcome chosen because it is easy to measure may be a poor surrogate for the clinically meaningful phenotype that prompted the original signal.

Rapid RWD analysis can nevertheless be highly valuable once the safety question is sufficiently specified. A structured framework for time-sensitive safety-signal assessment illustrates how real-world analyses can be organized around the question, available data, design choices, and operational constraints [7]. The advantage is not that speed generates stronger causality. It is that a signal can be moved efficiently from descriptive concern toward questions about occurrence, background frequency, comparative incidence, susceptible subgroups, temporality, and alternative explanations. In this sense, rapid RWD is most useful when it narrows uncertainty rather than when it produces a prematurely definitive effect estimate.

Regulatory experience reinforces the importance of matching the method to the actual decision problem. In a pilot of regulator-led RWE studies in the European medicines network, safety-related questions were prominent, but many analyses were descriptive rather than causal, and practical usefulness depended on data availability, representativeness, timeliness, and early clarification of the regulatory question [8]. That is not methodological

weakness. Some safety decisions require characterization rather than causal estimation. Others require an explicitly comparative counterfactual. The first task in causal safety assessment is therefore not to choose an estimator; it is to decide whether a causal estimand is both necessary and identifiable.

Target trial thinking

Target-trial thinking offers a disciplined way to perform that translation. The framework separates two operations that are sometimes conflated: specifying the randomized trial that would answer the causal question and then determining whether observational data can emulate its relevant components [9]. For post-marketing safety, those components include eligibility criteria, treatment strategies, assignment, time zero, follow-up, outcome definition, causal contrast, and analysis. Writing them explicitly often exposes ambiguities that were invisible in a generic exposed-versus-unexposed comparison.

This discipline is particularly useful because many observational biases arise before any statistical model is fitted. If eligibility is assessed after treatment assignment, if exposed patients accumulate required survival time before being classified as exposed, or if treatment strategies are defined using information unavailable at baseline, sophisticated adjustment can preserve rather than repair the design error. Target-trial emulation is therefore best understood as a strategy for avoiding self-inflicted bias through explicit design alignment [10]. It does not make treatment allocation random, and it cannot recover confounders or clinical details that the data source never measured.

Empirical comparisons with randomized trials provide an important but bounded validation lesson. In the RCT-DUPLICATE program, emulations that more closely reproduced the design and measurement characteristics of their corresponding trials showed stronger agreement than less-close emulations, while discordance remained for some comparisons [11, 12]. These results support the proposition that observational analyses can approximate selected randomized estimates when the target trial is sufficiently emulatable and design differences are minimized. They do not establish that database analyses and randomized trials are interchangeable, nor do they show that every adverse outcome can be studied with comparable fidelity.

A further distinction is therefore necessary between a valid causal question and data capable of answering it. Fit-for-purpose RWE frameworks emphasize that study-design fitness and data fitness must be considered separately [13]. A claims database may have excellent treatment capture but inadequate phenotype specificity. An EHR may contain rich clinical detail but incomplete care received outside the network. A registry may define a disease state well but provide an unsuitable comparator population. Causal inference is strengthened when “not identifiable with these data” is allowed to remain a legitimate conclusion instead of being treated as a methodological failure requiring a more complicated model.

Confounding control

Once the target contrast is explicit, comparator selection becomes one of the most consequential causal choices. The active-comparator new-user design is attractive because patients initiating treatments for similar indications may be more comparable than treated and untreated patients. Yet comparator selection is not mechanical. Alternative comparator strategies may be necessary when clinical practice, treatment sequencing, contraindications, rarity of exposure, or sample size makes the conventional active-comparator approach infeasible [14]. A comparator must therefore be judged by the clinical treatment decision it approximates, not merely by whether it produces acceptable covariate balance after modeling.

This point limits the increasingly common assumption that high-dimensional data will solve confounding if enough features are extracted. In an empirical study augmenting claims-based pharmacoepidemiology with features derived from electronic free-text notes, natural-language processing improved prediction of treatment choice, but the incremental improvement in prediction of outcome risk was small or absent in the reported examples [15]. The result is conceptually important. Better representation of why a patient received one treatment rather than another may improve adjustment, but predictive enrichment is not the same as demonstrating exchangeability. A machine-learning model can be exceptionally accurate while leaving the causal contrast biased.

Selection mechanisms require similar care because they are often described casually as “confounding” despite operating through different causal structures. Selection bias depends on how the analysed population differs from the referent population for the target causal effect, and restriction can alter the estimated contrast through mechanisms involving colliders, effect modification, or other selection structures [16]. In safety studies, such problems can arise through survival to treatment eligibility, availability of laboratory monitoring, specialist

referral, healthcare utilization, registry enrollment, persistence in treatment, or complete follow-up. These mechanisms should be represented explicitly because the remedy for unmeasured treatment confounding is not automatically the remedy for biased selection.

Residual confounding should also be confronted before results are known. Simulation-based bias analysis can be used during study design to examine how plausible characteristics of an unmeasured confounder, or imperfect proxies for it, could affect expected bias [17]. This does not identify the true hidden confounder or reveal the true causal effect. Its value is diagnostic. It can show whether a planned study is robust across plausible scenarios or whether a modest departure from exchangeability could materially change interpretation. In a post-marketing setting, this turns unmeasured confounding from a ritual limitation stated after analysis into a design property that can influence whether the study should proceed, which additional information should be sought, and how much causal language the resulting estimate can sustain.

Time-varying treatment and risk

Temporal structure is especially important in safety research because treatment status, disease severity, concomitant therapies, monitoring intensity, and susceptibility to harm can all change during follow-up. Immortal-time bias is one visible manifestation of this broader problem. Directed acyclic graph analyses show that apparent immortal-time problems can arise through several distinct structures involving exposure classification, eligibility, survival, and follow-up, and that simply deleting an “immortal” interval does not provide a universal solution [18]. The causal question must instead align treatment assignment, eligibility, and time zero so that patients are compared under interpretable treatment strategies.

The challenge becomes greater when treatment is not a single baseline decision. Patients discontinue, restart, switch, escalate, taper, or receive additional therapies in response to evolving health states. Those same states may predict the adverse outcome. Dynamic treatment-regime methods formalize the idea that treatment decisions can depend on accumulated patient history and therefore cannot always be represented faithfully by a static exposure label [19]. For pharmacovigilance, the implication is not that every safety study requires a dynamic regime. It is that the choice between baseline assignment, as-treated follow-up, switching strategies, cumulative exposure, dose trajectories, and other longitudinal representations should follow the biological and clinical structure of the hypothesized harm.

A coherent causal safety analysis must therefore preserve the logic of target-trial specification and comparator plausibility after baseline rather than applying them only at treatment initiation. The target-trial framework requires explicit treatment strategies and follow-up rules [9]. Comparator design likewise depends on clinically meaningful treatment alternatives [14]. This article proposes that longitudinal safety analyses make every change in treatment state auditable against those two questions: what treatment strategy is being contrasted, and whether the evolving comparator remains clinically interpretable. When those conditions fail, apparent treatment effects may instead reflect treatment adaptation to worsening disease, informative discontinuation, changing monitoring intensity, or selective survival.

The temporal architecture is easier to evaluate when treatment-state transitions, bias-generating mechanisms, and interpretive boundaries are displayed simultaneously rather than collapsed into a single exposure arrow.

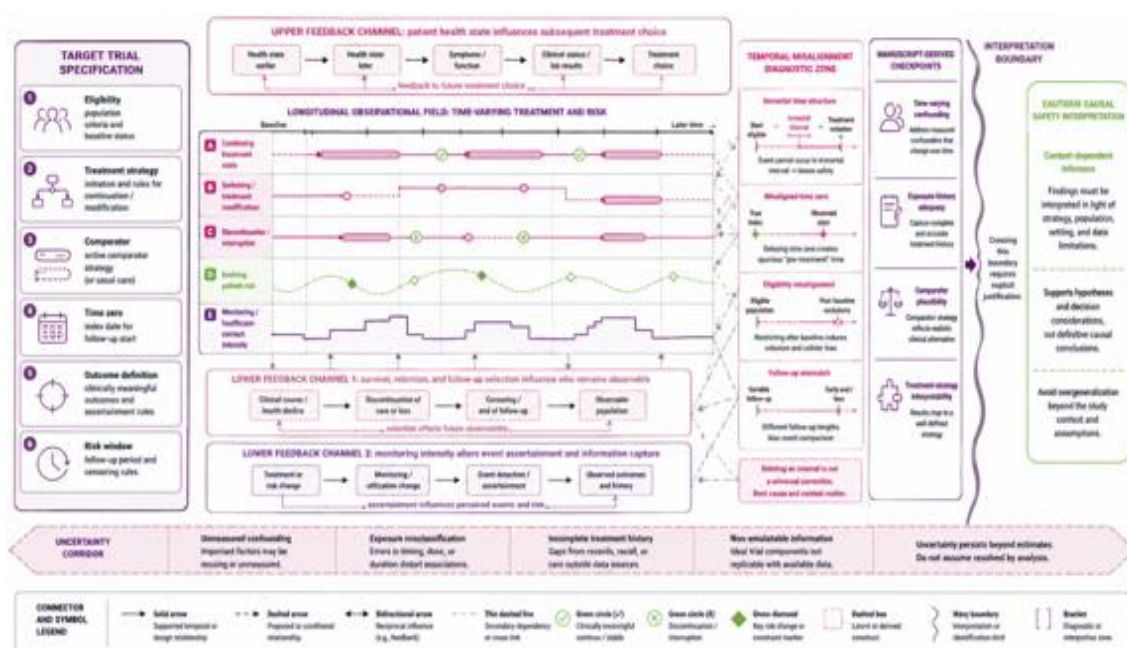


Figure 1. Temporal misalignment and treatment-state feedbacks that distort post-marketing causal interpretation

Negative and positive controls

Negative controls add a falsification layer that conventional covariate balance cannot provide. In pharmacoepidemiologic studies they have most commonly been used to detect residual systematic error, although some methods extend them toward calibration or bias correction. A systematic review of their use showed that diagnostic applications predominated, while methodological work demonstrates that stronger inferential uses require stronger assumptions about how the control shares the target study's confounding structure [20, 21]. A negative-control result is therefore informative only when the absence of the target causal relation is credible and the pathways generating bias are sufficiently comparable.

Control selection should consequently be treated as a design decision rather than as an afterthought. The same logic that makes drug–outcome reference sets design-specific [6] applies within a causal safety study. An outcome can fail as a negative control because the treatment plausibly affects it, because it is measured under a different healthcare-utilization process, because its risk window differs, or because its determinants do not reproduce the confounding structure of the target adverse event. This article proposes that controls be judged by their bias-sharing plausibility and causal exclusion logic rather than by convenience or statistical nullity alone.

Positive controls have a related but distinct role. A known or strongly expected association can help assess whether a database and analytic design are capable of detecting a signal under relevant measurement and follow-up conditions. That role is methodological calibration, not causal validation of the target drug–event pair. A design that recovers an expected positive control may still be biased for another outcome with different ascertainment, latency, confounding, or competing-risk structure. The most useful control strategy is therefore one that reveals which types of error the analysis can and cannot diagnose.

These distinctions are easier to use operationally when the control problem is organized around the specific bias pattern being challenged rather than around a generic list of advantages and disadvantages.

Table 1. Control-choice constraints for falsifying bias in post-marketing causal analyses

Manuscript-derived bias pattern	Context	Boundary	Interpretation	Validation and Decision
Scientific anchor	Governing condition Cross-source concordance	Preventability or reversibility	Escalation rule	Validation need Causal-safety endpoint protected

Residual treatment-selection confounding	Governing condition for an informative control Control must plausibly share treatment-selection determinants while lacking the target causal effect	Often partly reducible through improved comparator or covariate design	Unexpected association should trigger reassessment of exchangeability before stronger causal language	Validation need Clinical review of control validity and sensitivity to alternative confounding structures
	Cross-source concordance needed Concordance across claims, EHR, or other sources strengthens concern when the same unexpected control association persists			Causal-safety endpoint protected Protection against mistaking treatment-selection differences for drug-related harm
Outcome-ascertainment or surveillance bias	Governing condition for an informative control Control outcome should experience a comparable observation process without sharing the hypothesized drug effect	May be reduced by harmonizing measurement and observation windows	Divergent control behavior should prompt examination of monitoring intensity and healthcare contact	Validation need Verify that control and target outcomes are captured through comparable observation mechanisms
	Cross-source concordance needed Useful when ascertainment behavior is similar across sources or settings			Causal-safety endpoint protected Protection against interpreting intensified surveillance as higher biological event risk
Temporal misalignment or immortal-time structure	Governing condition for an informative control Control analysis must preserve the same eligibility, assignment and follow-up architecture as the target analysis	Frequently preventable through redesign when identified before estimation	Failure of the control under alternate time-zero definitions should trigger redesign rather than post hoc reassurance	Validation need Re-specify target-trial timing and repeat the diagnostic under plausible temporal definitions
	Cross-source concordance needed Cross-source agreement is secondary to internal temporal validity			Causal-safety endpoint protected Protection against survival or treatment-classification artifacts
Systematic database or analytic error	Governing condition for an informative control Control must traverse the same coding, extraction and analytic pipeline as the target pair	Potentially reversible when the error originates in implementation	Persistent unexpected control estimates should block automatic escalation of the target estimate	Validation need Independent code review, alternative specifications, and data-quality checks
	Cross-source concordance needed Replication of the same anomaly across independent data pipelines raises concern about shared design assumptions rather than one database alone			Causal-safety endpoint protected Protection against attributing pipeline artifacts to pharmacological effects

Proposed calibration use	Governing condition for an informative control	Does not by itself remove target-study bias	Successful calibration permits analysis to proceed but cannot waive target-specific bias assessment	Validation need
	Cross-source concordance needed			Causal-safety endpoint protected
	A positive or negative reference pair must be clinically and methodologically compatible with the target design Agreement across suitable reference pairs can support method performance but cannot establish target validity			Confirm that reference-pair detectability, measurement and follow-up resemble the target question Protection against treating benchmark recovery as proof that the target association is causal

Quantitative bias analysis

Traditional limitations sections often acknowledge residual confounding, exposure misclassification, outcome error, or selection bias without asking how strong those biases would need to be to change interpretation. Quantitative bias analysis changes the form of that discussion. A systematic review of applied QBA showed that epidemiologic studies have used deterministic and probabilistic methods to evaluate unmeasured confounding, misclassification, and selection mechanisms, although applications frequently concentrated on one bias at a time and varied substantially in reporting quality [22]. QBA therefore does not remove uncertainty. It converts selected assumptions about uncertainty into an analyzable structure.

The E-value is one compact example. It summarizes, on the risk-ratio scale, the minimum strength of association that an unmeasured confounder would need to have with both exposure and outcome, conditional on measured covariates, to explain away an observed association or move the confidence interval toward the null [23]. Used carefully, this can make otherwise vague statements about unmeasured confounding more interpretable. It is particularly attractive in post-marketing settings because investigators may need to communicate sensitivity without constructing a full probabilistic bias model for every analysis.

The attraction of a single sensitivity statistic also creates a risk of false closure. Methodological critiques of E-values emphasize that they do not capture every relevant dimension of confounding and do not address other forms of systematic error, such as selection bias or measurement error [24]. They can also be overinterpreted when divorced from substantive knowledge about plausible confounders. A large sensitivity value is not a probability of causation, a regulatory threshold, or proof that no realistic bias could explain the result. Conversely, a smaller value does not establish that the association is noncausal. The statistic is informative only within the assumptions it represents.

For safety assessment, QBA is most useful when applied prospectively and selectively. Design-stage simulations can examine plausible unmeasured-confounding scenarios before a database analysis is finalized [17]. After estimation, further analyses can explore how specific residual biases might change the magnitude or direction of the effect. This article proposes that the objective should not be to produce one global robustness score. Instead, each important bias should retain its identity, assumptions, and evidential consequence. A causal estimate becomes more interpretable when decision makers can see which uncertainties were quantified, which were only qualitatively assessed, and which remain outside the available data altogether.

Where causal methods remain fragile

Causal methods remain fragile when the data cannot represent the clinical process that generated treatment and outcome. Target-trial specification can remove avoidable ambiguity, but it cannot create an unmeasured disease-severity variable, recover an inaccurately coded phenotype, or reconstruct treatment received outside the observed system [9]. Richer data can improve proxy measurement without demonstrating that residual confounding has disappeared [15]. Selection and temporal bias also remain distinct problems: restriction can change the target causal contrast [16], while misaligned eligibility, treatment assignment, and follow-up can generate separate time-dependent distortions [18]. Sensitivity analyses inherit the same limitation. Quantitative bias analysis evaluates the biases that are parameterized [22], and no single metric summarizes all unmeasured error [24]. Even empirical agreement between selected trial emulations and randomized trials should therefore be treated as conditional validation rather than evidence that observational causal estimates transport automatically across drugs, outcomes, populations, or databases [11]. A study may be internally coherent yet still be poorly transferable when prescribing

patterns, baseline risk, event ascertainment, or treatment alternatives differ. Causal credibility must therefore be assessed in practice at the level of the specific question and data-generating setting.

Clinical judgment as an independent evidence layer

Clinical judgment contributes information that population-level causal estimation does not fully encode. Structured tools such as MONARCSi can improve consistency and documentation in individual-case causality assessment, but agreement with safety professionals is not equivalent to establishing causal truth [25]. Phenotype-specific problems make the boundary clearer. Drug-induced liver injury assessment depends on chronology, laboratory pattern, competing etiologies, concomitant exposure, dechallenge or rechallenge information, and expert adjudication [26]. These features may determine whether an apparently compatible population association is clinically persuasive in a particular case.

This article therefore proposes that clinical judgment be treated as an independent evidence layer rather than as an informal residual step. Independence does not mean isolation: case-level observations can motivate population analyses, and causal estimates can alter the plausibility assigned to individual cases. The non-substitution principle is that neither layer should automatically erase unresolved evidence in the other.

A combined causal–clinical decision framework

The proposed framework links four functions without collapsing them. First, signal characterization asks whether the observed pattern is sufficiently credible and clinically consequential to justify further inquiry [4]. Second, causal specification defines the target contrast, comparator, time structure, and data requirements [9]. Third, bias diagnosis should stress-test plausible unmeasured confounding [17], use controls to expose residual systematic error [20], and quantify specified uncertainty where feasible [22]. Fourth, clinical adjudication should preserve structured case-level assessment [25] and phenotype-specific evaluation of chronology and competing causes [26].

These functions feed a decision process rather than a numerical causality score. Regulatory RWE increasingly depends on principled study design and collaborative evidence generation across organizations [3, 27]. The framework therefore does not prescribe that a particular effect estimate, control result, E-value, or disproportionality measure should trigger action. It proposes instead that decision makers state which causal question was identified, which biases were addressed, which remained unresolved, what independent clinical evidence contributes, and which uncertainties are acceptable given the potential severity and reversibility of harm. Validation should likewise be layer-specific: study emulation, data fitness, control behavior, bias sensitivity, phenotype adjudication, and external transportability require different tests rather than one global performance metric.

CONCLUSION

Causal inference can make post-marketing safety assessment more explicit, falsifiable, and resistant to avoidable bias, but methodological sophistication does not remove evidential boundaries. Signals remain distinct from causal effects, robustness analyses remain conditional on their assumptions, and population estimates cannot absorb every clinically relevant feature of an adverse event. The most defensible future model is therefore integrative rather than substitutive: causal methods should discipline the path from signal to explanation while clinical judgment preserves the patient-level, phenotypic, and contextual evidence that statistical designs cannot legitimately replace.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Fusaroli M, Salvo F, Begaud B, AlShammari TM, Bate A, Battini V, et al. The reporting of a disproportionality analysis for drug safety signal detection using individual case safety reports in Pharmacovigilance (READUS-PV): Development and statement. *Drug Saf.* 2024;47(6):575-84. doi:10.1007/s40264-024-01421-9
2. Bernardeau C, Revol B, Salvo F, Fusaroli M, Raschi E, Cracowski JL, et al. Are causal statements reported in pharmacovigilance disproportionality analyses using individual case safety reports exaggerated in related citations? A meta-epidemiological study. *Drug Saf.* 2025;48(6):679-88. doi:10.1007/s40264-025-01524-x
3. Hernandez RK, Critchlow CW, Dreyer N, Lash TL, Reynolds RF, Sørensen HT, et al. Advancing principled pharmacoepidemiologic research to support regulatory and healthcare decision making: The era of real-world evidence. *Clin Pharmacol Ther.* 2025;117(4):927-37. doi:10.1002/cpt.3563
4. Scosyrev E, Behr S, Jain D, Ponnuru A, Michel C. Disproportionality analysis and causal inference in drug safety. *Pharmaceut Med.* 2025;39(2):97-107. doi:10.1007/s40290-024-00549-4
5. van Stekelenborg J, Kara V, Haack R, Vogel U, Garg A, Krupp M, et al. Individual case safety report replication: An analysis of case reporting transmission networks. *Drug Saf.* 2023;46(1):39-52. doi:10.1007/s40264-022-01251-7
6. Coste A, Wong AYS, Warren-Gash C, Matthewman J, Bate A, Douglas IJ. Implementation of a taxonomy-based framework for the selection of appropriate drugs and outcomes for real-world data signal detection studies. *Drug Saf.* 2024;47(2):183-92. doi:10.1007/s40264-023-01382-5
7. Wang L, Golchin N, von Klot S, Salinas CA, Manlik K, Patadia V, et al. Adopting a framework for rapid real-world data analyses in safety signal assessment. *Ther Innov Regul Sci.* 2024;58(6):1014-22. doi:10.1007/s43441-024-00694-7
8. Prilla S, Groeneveld S, Pacurariu A, Restrepo-Méndez MC, Verpillat P, Torre C, et al. Real-world evidence to support EU regulatory decision making—Results from a pilot of regulatory use cases. *Clin Pharmacol Ther.* 2024;116(5):1188-97. doi:10.1002/cpt.3355
9. Hernán MA, Dahabreh IJ, Dickerman BA, Swanson SA. The target trial framework for causal inference from observational data: Why and when is it helpful? *Ann Intern Med.* 2025;178(3):402-7. doi:10.7326/ANNALS-24-01871
10. Fu EL. Target trial emulation to improve causal inference from observational data: What, why, and how? *J Am Soc Nephrol.* 2023;34(8):1305-14. doi:10.1681/ASN.0000000000000152
11. 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. doi:10.1001/jama.2023.4221
12. 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. doi:10.1161/CIRCULATIONAHA.120.051718
13. Gatto NM, Vitioe SE, Rubinstein E, Reynolds RF, Campbell UB. A structured process to identify fit-for-purpose study design and data to generate valid and transparent real-world evidence for regulatory uses. *Clin Pharmacol Ther.* 2023;113(6):1235-9. doi:10.1002/cpt.2883
14. Wintzell V, Svanström H, Pasternak B. Selection of comparator group in observational drug safety studies: Alternatives to the active comparator new user design. *Epidemiology.* 2022;33(5):707-14. doi:10.1097/EDE.0000000000001521
15. Wyss R, Plasek JM, Zhou L, Bessette LG, Schneeweiss S, Rassen JA, et al. Scalable feature engineering from electronic free text notes to supplement confounding adjustment of claims-based pharmacoepidemiologic studies. *Clin Pharmacol Ther.* 2023;113(4):832-8. doi:10.1002/cpt.2826
16. Lu H, Cole SR, Howe CJ, Westreich D. Toward a clearer definition of selection bias when estimating causal effects. *Epidemiology.* 2022;33(5):699-706. doi:10.1097/EDE.0000000000001516
17. Desai RJ, Bradley MC, Lee H, Eworuke E, Weberpals J, Wyss R, et al. A simulation-based bias analysis to assess the impact of unmeasured confounding when designing nonrandomized database studies. *Am J Epidemiol.* 2024;193(11):1600-8. doi:10.1093/aje/kwae102
18. Yang G, Burgess S, Schooling CM. Illustrating the structures of bias from immortal time using directed acyclic graphs. *Int J Epidemiol.* 2025;54(1):dyae176. doi:10.1093/ije/dyae176

19. Williams NT, Hoffman KL, Díaz I, Rudolph KE. Learning optimal dynamic treatment regimes from longitudinal data. *Am J Epidemiol.* 2024;193(12):1768-75. doi:10.1093/aje/kwae122
20. Zafari Z, Park JE, Shah CH, dosReis S, Gorman EF, Hua W, et al. The state of use and utility of negative controls in pharmacoepidemiologic studies. *Am J Epidemiol.* 2024;193(3):426-53. doi:10.1093/aje/kwad201
21. Flanders WD, Waller LA, Zhang Q, Getahun D, Silverberg M, Goodman M. Negative control exposures: Causal effect identifiability and use in probabilistic-bias and Bayesian analyses with unmeasured confounders. *Epidemiology.* 2022;33(6):832-9. doi:10.1097/EDE.0000000000001528
22. Petersen JM, Ranker LR, Barnard-Mayers R, MacLehose RF, Fox MP. A systematic review of quantitative bias analysis applied to epidemiological research. *Int J Epidemiol.* 2021;50(5):1708-30. doi:10.1093/ije/dyab061
23. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: Introducing the E-value. *Ann Intern Med.* 2017;167(4):268-74. doi:10.7326/M16-2607
24. Ioannidis JPA, Tan YJ, Blum MR. Limitations and misinterpretations of E-values for sensitivity analyses of observational studies. *Ann Intern Med.* 2019;170(2):108-11. doi:10.7326/M18-2159
25. Comfort S, Dorrell D, Meireis S, Fine J. MODified NARanjo causality scale for ICSRs (MONARCSi): A decision support tool for safety scientists. *Drug Saf.* 2018;41(11):1073-85. doi:10.1007/s40264-018-0690-y
26. Hey-Hadavi J, Seekins D, Palmer M, Coffey D, Caminis J, Abdullaev S, et al. Overview of causality assessment for drug-induced liver injury (DILI) in clinical trials. *Drug Saf.* 2021;44(6):619-34. doi:10.1007/s40264-021-01051-5
27. Beck AE, Kampman M, Huynh C, Simon C, Plueschke K, Cohet C, et al. Collaborative real-world evidence among regulators: Lessons and perspectives. *Clin Pharmacol Ther.* 2025;117(2):368-73. doi:10.1002/cpt.3457