Available online **www.ijpras.com**

International Journal of Pharmaceutical Research & Allied Sciences, 2025, 14(4):47-56
https://doi.org/10.51847/2vx9iuAvw1

Original Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Safety after the Average Patient: Detecting Heterogeneous Drug Harm across Age, Sex, Comorbidity, Genetics, and Concomitant Treatment

Charlotte Dupuis[1]*, Hugo Perrin[1], Elise Morel[2], Thomas Martin[3]

[1]Department of Heterogeneous Drug Harm Detection, Faculty of Pharmacy, University of Lille, Lille, France.
[2]Department of Subgroup Safety Analysis and Comorbidity, Faculty of Pharmacy, University of Montpellier, Montpellier, France.
[3]Department of Genetics and Concomitant Treatment, Faculty of Pharmacy, University of Nantes, Nantes, France.

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

ABSTRACT

Population-average safety estimates can obscure clinically important differences in drug harm because patients differ in baseline susceptibility, realized exposure, biological response, concomitant treatment, and the probability that an adverse outcome will be observed or reported. This Original Subgroup-Safety Framework Article proposes a multidimensional architecture for detecting and evaluating heterogeneous drug harm across age, sex, pregnancy, physiological state, comorbidity, frailty, organ impairment, genetics, ancestry, metabolic phenotype, and polypharmacy. The framework separates five interacting explanatory layers: baseline susceptibility, exposure modification, biological response, concomitant-treatment context, and observation or data-generation processes. It then connects subgroup-signal detection to cross-source triangulation, causal assessment, external validation, bounded communication, and decision-specific evaluation. Polypharmacy is represented as a time-varying interaction structure rather than a medication count, allowing drug–drug, drug–gene, and drug–disease relationships to be considered jointly. To limit false subgroup conclusions, the framework requires explicit subgroup definitions, interaction assessment, multiplicity control, partial pooling where appropriate, evaluation of sparse intersections, and independent replication. Validation is treated as a staged requirement encompassing construct validity, data validity, analytical validity, external transportability, causal coherence, decision utility, and continuing performance under clinical and dataset drift. The original contribution is an integrated conceptual synthesis that distinguishes observed subgroup differences from drug-attributable heterogeneous harm and separates signal detection from causal, clinical, and regulatory conclusions. The framework remains unvalidated and does not prescribe treatment, establish subgroup-specific safety, or define regulatory action. Its value must be tested through multisource retrospective studies, prospective evaluation, fairness assessment, and monitoring of subgroup-specific calibration over time.

Key words: *Pharmacovigilance, Drug safety, Safety signal, Causal inference, Adverse event, Real-world data*
Received: 26 June 2025; **Revised:** 29 October 2025; **Accepted:** 02 November 2025

INTRODUCTION

Drug-safety evidence is commonly summarized through population-level event rates, relative risks, disproportionality measures, or average treatment effects. These summaries are necessary, but they may conceal variation in absolute harm and treatment response across patients who differ in prognosis, physiology, treatment context, and susceptibility. Methodological work on heterogeneous treatment effects therefore emphasizes that

average estimates cannot, by themselves, determine whether clinically meaningful variation exists across individuals or subgroups [1].

Heterogeneity is also not a single statistical phenomenon. Differences in baseline outcome risk may alter absolute harm even when a relative treatment effect is constant, whereas genuine effect modification implies that the causal effect of treatment changes across patient contexts. Predictive risk modelling and treatment-effect modelling consequently answer different questions and require different assumptions, estimands, and validation strategies [2]. Conflating them can transform an expected difference in prognosis into an unsupported claim of differential drug toxicity.

Pharmacovigilance refers to the detection, assessment, understanding, and prevention of adverse effects or other medicine-related problems, whereas drug safety encompasses the broader evidence and decisions concerning medicine-associated harm. A reported adverse event is an observation following treatment and need not be caused by the drug; a safety signal is information suggesting a possible new causal association or a new aspect of a known association. Contemporary surveillance increasingly combines spontaneous reports, longitudinal healthcare data, active surveillance, and computational methods because no single source captures all relevant harms or explanations [3].

This article develops an original, non-empirical Subgroup-Safety Framework for examining heterogeneous drug harm across age, sex, pregnancy, comorbidity, frailty, organ impairment, genetics, metabolic phenotype, and concomitant treatment. It proposes a common architecture for distinguishing susceptibility, exposure, response, interaction, and observation processes; organizing evidence escalation; and defining decision boundaries. The framework is a conceptual synthesis rather than a validated prediction model, causal estimator, clinical guideline, or regulatory standard.

Why population-average safety estimates conceal risk

A population-average estimate can conceal risk through variation in baseline susceptibility. Patients with greater underlying probability of an outcome may experience larger absolute harm even when the drug's relative effect does not vary. Conversely, an observed difference in outcome frequency between groups may reflect prognosis rather than a causal interaction with treatment. Real-world heterogeneity assessment must therefore distinguish prognostic variation from treatment-effect modification before describing a subgroup as differentially harmed [4].

Real-world safety comparisons introduce additional difficulties because treatment is rarely assigned independently of patient characteristics. Disease severity, prior treatment, organ function, healthcare access, monitoring intensity, and clinician preference can influence both treatment selection and adverse-outcome detection. Measurement error, missing exposure information, treatment switching, limited overlap between comparison groups, and unrecorded confounding may further create or obscure subgroup differences [5]. Statistical adjustment can reduce selected biases but cannot guarantee causal identification.

Exploratory models can search many patient characteristics, interactions, outcomes, time windows, and model specifications. Such flexibility may identify apparent heterogeneity that is unstable outside the development sample. Predictive approaches to heterogeneous treatment effects therefore require evaluation of model specification, calibration, discrimination, subgroup definition, and external corroboration before a subgroup claim is accepted [6]. A high-risk prediction or statistically significant interaction does not independently establish drug-attributable harm.

This article proposes that apparent subgroup harm may arise from five sources: baseline susceptibility, modified drug exposure, altered biological response, interaction with concomitant treatment, and variation in observation or data generation. These sources may coexist. For example, an older patient with kidney impairment may have higher baseline vulnerability, increased exposure because of reduced clearance, more interacting medicines, and more intensive laboratory monitoring. The framework therefore treats an observed subgroup difference as a hypothesis requiring explanation rather than a causal conclusion [4].

Dimensions of treatment-effect heterogeneity

Clinically relevant heterogeneity may arise through sex-related physiology, inherited variation in drug disposition and response, and interaction-dependent polypharmacy effects [7-9]. These dimensions are neither interchangeable nor necessarily independent. Their influence may operate through exposure, pharmacodynamic sensitivity, disease processes, treatment selection, or healthcare observation. The proposed framework consequently avoids treating demographic or molecular characteristics as self-sufficient explanations of harm.

Susceptibility heterogeneity refers to variation in the probability or seriousness of harm arising from pre-treatment patient characteristics and evolving physiological states. It may include age-related physiological change, sex-related biology, pregnancy, frailty, comorbidity, prior disease, and organ impairment. Exposure heterogeneity refers to differences in the concentration and duration of drug exposure created by dose, adherence, absorption, distribution, metabolism, elimination, treatment timing, and physiological change. Sex and gender must remain conceptually distinct because biological mechanisms and socially patterned treatment contexts may generate different pathways [7].

Response heterogeneity describes variation in biological consequences at a comparable exposure, including pharmacodynamic, immune, genomic, and metabolic-phenotype differences. Contextual interaction heterogeneity arises when harm depends on a combination of treatment and patient factors, such as drug–drug, drug–gene, drug–disease, or treatment-sequence interactions. Computational representations of polypharmacy demonstrate the importance of modelling combination-dependent adverse effects, but predicted relations remain hypotheses until replicated and clinically evaluated [9].

Observation heterogeneity describes variation in whether an underlying event becomes recorded, coded, reported, or linked to a medicine. Spontaneous reporting may be shaped by awareness, notoriety, seriousness, access, and reporting behaviour, while longitudinal data may be shaped by diagnostic opportunity, coding, monitoring, and missingness. These processes can produce apparent subgroup differences without corresponding biological effect modification. Surveillance architecture must therefore evaluate how evidence was generated before interpreting a signal as heterogeneous drug harm [3].

Proposed subgroup-safety framework

The proposed Safety After the Average Patient Multidimensional Subgroup-Safety Framework integrates complementary signal sources, fit-for-purpose longitudinal real-world evidence, and structured representations of drug–event relationships rather than relying on a single surveillance stream [10-12]. Its purpose is to organize the questions, evidence, validation stages, and decision boundaries needed to investigate subgroup harm. The architecture does not assume that a reported event, statistical association, or computationally predicted relation is causal.

At its centre is observed subgroup drug harm, defined as a detected difference in the occurrence, severity, timing, or predicted probability of an adverse outcome across a specified patient-and-treatment context. Five interacting layers surround this observation: baseline susceptibility, exposure modification, biological response, concomitant-treatment context, and observation or data-generation processes. **Figure 1** presents the original conceptual synthesis for multidimensional architecture of heterogeneous drug harm, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

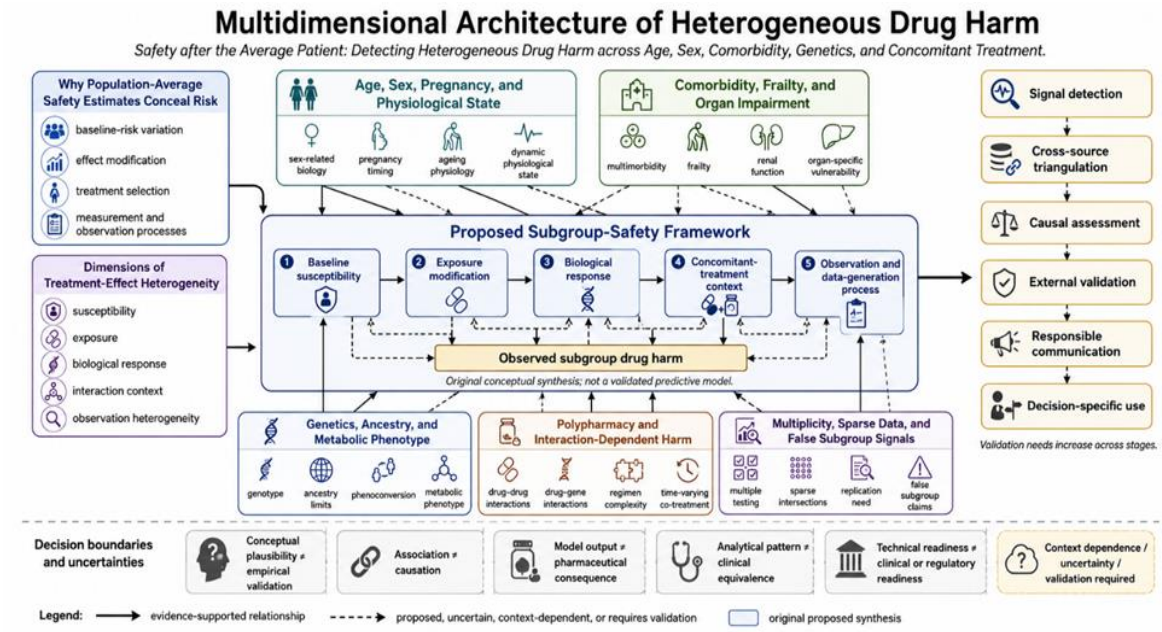

Figure 1. Multidimensional Architecture of Heterogeneous Drug Harm.

The figure is an original conceptual synthesis developed for "Safety after the Average Patient: Detecting Heterogeneous Drug Harm across Age, Sex, Comorbidity, Genetics, and Concomitant Treatment." Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, clinical recommendation, regulatory determination, or deployment-ready system.

The framework proceeds through six operations. First, context specification defines the medicine, comparator, adverse outcome, time zero, follow-up, subgroup dimensions, and intended decision. Second, multidimensional detection evaluates joint patient–drug–context configurations rather than isolated demographic splits. Third, explanation separation asks whether an observed difference is most consistent with baseline susceptibility, exposure modification, biological response, treatment interaction, observation bias, or a combination. Complementary use of spontaneous reports and longitudinal databases is particularly important because their strengths and biases differ [10].

Fourth, cross-source triangulation compares direction, timing, outcome definitions, dose patterns, treatment context, and plausible mechanisms across spontaneous reports, real-world data, active surveillance, clinical evidence, and structured relational evidence. Fifth, validation escalation requires evidence proportional to the proposed use, progressing from construct and data validity to analytical, external, causal, decision, and continuing validity. Fit-for-purpose data, transparent analytical questions, reproducibility, and explicit inference boundaries are necessary throughout this process [11]. Knowledge graphs may organize relationships among medicines, phenotypes, genes, diseases, and adverse outcomes, but their outputs require independent clinical and epidemiological validation [12].

The framework generates seven testable propositions. Multidimensional representations should replicate more reliably than one-variable subgroup splits; some apparent signals should weaken after exposure and observation processes are modelled; and cross-source concordance should increase causal credibility. Genotype-only models should be less stable when disease, inflammation, organ function, or concomitant treatment alters metabolic phenotype. Prespecification, shrinkage, interaction assessment, and replication should reduce false subgroup claims. Validity should remain decision specific, because successful detection does not establish causal attribution, treatment selection, or clinical utility. Finally, subgroup-specific calibration should change when populations, clinical practice, or data capture change. These propositions require empirical testing and external corroboration [6].

Age, sex, pregnancy, and physiological state

Sex-associated pharmacokinetic differences may contribute to variation in adverse drug reactions through differences in body composition, enzyme activity, transport, clearance, and exposure at comparable doses [13]. Such patterns do not establish that sex alone determines individual risk. Sex-related biology should be distinguished from gender-related differences in prescribing, adherence, healthcare access, symptom recognition, and reporting.

Pregnancy represents a dynamic physiological state rather than a fixed demographic category. Safety assessment must consider gestational timing, dose, comparator choice, outcome specificity, maternal disease, co-treatment, and surveillance opportunity. Comparative evidence on lithium illustrates how pregnancy-safety questions require carefully timed exposure definitions and confounding-aware designs rather than undifferentiated classification of exposure during pregnancy [14].

Chronological age is similarly an incomplete proxy for vulnerability. Medication-related harm in older adults depends on frailty, organ function, treatment goals, regimen complexity, prior adverse effects, and the balance between therapeutic benefit and cumulative burden. Decision-support evidence in hospitalized older adults demonstrates the importance of patient- and regimen-level medication review without establishing a universal deprescribing rule [15].

The proposed framework therefore represents age, pregnancy, hormonal state, organ function, inflammation, and other physiological conditions as time-indexed modifiers. These variables may alter baseline susceptibility, drug exposure, biological response, or observation intensity. They should guide hypothesis formation and validation requirements, but they should not be used as deterministic labels of vulnerability or as substitutes for patient-specific evidence.

Comorbidity, frailty, and organ impairment

Multimorbidity, frailty-linked medication complexity, and impaired kidney function create overlapping susceptibility and exposure pathways that may increase the probability or seriousness of drug harm [16-18]. They remain distinct constructs: multimorbidity describes coexisting diseases, frailty describes reduced physiological reserve, and organ impairment may directly alter drug disposition, tolerance, or recovery from an adverse event.

Comorbidity may affect safety through several routes. It can modify treatment indication, competing risks, symptom interpretation, adherence, monitoring, and clinician selection of therapy. Because multimorbidity and polypharmacy frequently coexist and are defined inconsistently, a simple disease count cannot determine whether observed harm reflects biological susceptibility, treatment complexity, or differential healthcare contact [16].

Frailty should be treated as a potentially dynamic vulnerability state rather than as a synonym for older age. Its relationship with polypharmacy may be bidirectional: medication burden may contribute to functional decline, while frailty may increase prescribing complexity and sensitivity to adverse effects. Variation among frailty measures further limits direct comparison across studies and settings [17].

Organ impairment introduces more specific pharmacological pathways. Declining kidney function may increase exposure to renally eliminated medicines, alter metabolite accumulation, and increase serious adverse-reaction risk [18]. Nevertheless, kidney evidence cannot automatically be generalized to hepatic, cardiac, or other impairment. Each proposed relationship requires drug-specific pharmacology, longitudinal organ-function measurement, and evaluation of competing explanations.

Genetics, ancestry, and metabolic phenotype

Pharmacogenetic information may support safety improvement when a functional variant, drug pathway, toxicity mechanism, and treatment strategy are sufficiently established. Prospective evidence for DPYD-guided fluoropyrimidine dosing provides an example of context-specific genotype-informed risk reduction [19]. It does not establish that all genetic associations are clinically actionable or transferable across medicines and populations.

Genotype is not equivalent to expressed metabolic phenotype. Inflammation, disease, diet, environmental exposures, organ dysfunction, and concomitant medicines can alter enzyme activity, producing phenoconversion between genotype-predicted and observed metabolism [20]. A subgroup-safety model based only on inherited variation may therefore misclassify exposure and risk when the patient's physiological and treatment context changes.

Ancestry requires particular caution. Unequal representation in genetic research constrains discovery, calibration, and transportability across populations [21]. Broad ancestry labels may combine genetic, environmental, social, and healthcare factors and should not be treated as direct biological mechanisms. Using ancestry as a substitute for measured functional variation can reinforce unsupported assumptions and obscure within-group diversity.

The framework consequently proposes a dynamic genotype–phenotype–exposure representation. Genetic variation is linked to expressed metabolic function, current disease state, organ function, dose, and interacting treatment before a safety hypothesis is advanced. This architecture is conceptually plausible but requires cross-ancestry validation, functional interpretation, repeated phenotype assessment, and evidence that its use improves the intended decision.

Polypharmacy and interaction-dependent harm

Polypharmacy is often represented by a medication-count threshold, but count alone does not describe pharmacological compatibility, treatment necessity, dose, timing, or patient susceptibility. Its clinical consequences depend on the composition and purpose of the regimen, the vulnerability of the patient, and the presence of pharmacokinetic or pharmacodynamic interactions [22].

Drug–drug interactions may contribute to adverse-drug-reaction-related hospital admissions, particularly among older patients with complex treatment regimens [23]. However, the presence of a listed interaction does not prove that it caused an individual event. Indication, disease severity, adherence, dose changes, treatment duration, and monitoring intensity may influence both interaction exposure and outcome detection.

The proposed framework therefore represents concomitant treatment as a time-varying graph rather than a static list. Nodes may represent medicines, diseases, genes, organ functions, and outcomes, while edges represent hypothesized interactions. Graph-based modelling can expose combination-dependent patterns that are difficult to identify through isolated drug–event pairs, but computationally predicted relations remain hypotheses until externally validated [9].

Interaction-dependent harm should be evaluated through explicit temporal ordering, biological plausibility, dose and dechallenge information where available, alternative-treatment comparisons, and replication across data sources. The framework does not define a universally unsafe medication count or treat every theoretical interaction as clinically consequential.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for integrated construct, components, evidence requirements, failure modes, and decision boundaries for safety after the average patient.

Table 1. Integrated construct, components, evidence requirements, failure modes, and decision boundaries for Safety after the Average Patient

Component or scientific dimension	Problem addressed	Inputs or determinants	Proposed mechanism or relationship	Expected contribution	Evidence required	Failure or uncertainty risk	Boundary statement
Baseline susceptibility	Average estimates may obscure differences in absolute vulnerability	Age, sex-related biology, pregnancy, frailty, comorbidity, prior disease	**Proposed:** Patient state modifies baseline probability or seriousness of harm	Separates prognosis from drug-effect modification	Validated clinical phenotypes and longitudinal risk measurement [16]	Residual confounding and use of demographic proxies	Higher baseline risk does not establish a differential causal drug effect
Exposure modification	Prescribed dose may not equal realized exposure	Dose, adherence, absorption, clearance, organ function, physiological state	**Proposed:** Patient and treatment factors alter concentration and exposure duration	Identifies pharmacokinetic explanations for subgroup differences	Drug-specific exposure, timing, dose, and organ-function evidence [18]	Misclassified adherence, dose, or time-varying physiology	Exposure plausibility does not establish that the observed event was drug caused
Biological response	Similar exposure may produce different consequences	Pharmacodynamics, immune function, genotype, metabolic phenotype	**Proposed:** Molecular or physiological variation modifies response at a given exposure	Organizes mechanistic subgroup hypotheses	Functional and externally replicated evidence [19]	Weak functional interpretation and limited population representation	Genetic association alone is not an individual safety determination
Concomitant-treatment context	Harm may depend on treatment combinations	Medication sequence, dose, duration, indication, drug–drug and drug–gene interactions	**Proposed:** Regimen structure modifies exposure or response	Moves analysis beyond medication count	Temporally resolved treatment and interaction evidence [23]	Confounding by indication and false interaction alerts	Co-prescription or database co-occurrence does not prove interaction-dependent harm
Observation and data generation	Recorded differences may reflect surveillance processes	Reporting propensity, coding, monitoring, access, diagnostic opportunity, missingness	**Proposed:** Data-generation processes modify apparent subgroup-harm estimates	Distinguishes biological variation from observation bias	Cross-source comparison and phenotype validation [10]	Differential reporting and incomplete denominators	A reporting imbalance is a signal, not a causal conclusion
Cross-source triangulation	One data source may provide an incomplete or biased view	Spontaneous reports, longitudinal data, active surveillance, clinical and mechanistic evidence	**Proposed:** Concordance across complementary sources increases credibility	Supports evidence escalation and explanation testing	Fit-for-purpose data and reproducible analyses [11]	Correlated biases and inconsistent definitions	Concordance strengthens but does not prove causality

Credibility, validation, and drift	Exploratory subgroup findings may be unstable	Prespecification, interaction tests, replication, calibration, setting and time	**Proposed:** Validation should be proportional to intended decision use	Separates detection from causal, clinical, and regulatory claims	Independent validation, transparent reporting, and continuing monitoring [24]	Multiplicity, sparse intersections, transport failure, and drift	Performance in one dataset does not establish clinical utility or continuing validity

Multiplicity, sparse data, and false subgroup signals

Subgroup analysis creates a large search space involving patient characteristics, treatments, outcomes, time windows, interactions, and analytical choices. Credibility therefore depends on prior specification, biological rationale, formal interaction assessment, consistency, direction, and attention to multiplicity [25]. A nominally significant subgroup estimate is not sufficient evidence of effect modification.

Meta-research has shown that subgroup claims may receive limited statistical support or corroboration outside the analysis in which they were identified [26]. This risk is especially important when several correlated modifiers are examined or when investigators selectively emphasize the most extreme estimate. Exploratory findings should be reported as hypotheses unless they reproduce under independent conditions.

The proposed framework uses partial pooling, shrinkage, sensitivity analysis, negative controls, alternative outcome definitions, and external replication as credibility safeguards. These methods may reduce instability but cannot eliminate false discovery. Confirmatory status should require explicit estimands, predefined directions where justified, sufficient overlap, transparent handling of missing data, and consistency across clinically meaningful analyses [25].

Sparse intersections are unavoidable when age, sex, comorbidity, genetics, organ function, and concomitant treatment are considered jointly. Instead of assigning unstable estimates to every intersection, the framework permits findings to remain unresolved. Uncertainty should be displayed directly, and absence of evidence in a sparsely represented subgroup should not be communicated as evidence of safety or absence of heterogeneity [26].

Validation and responsible communication

Validation must begin by examining whether participants, predictors, outcomes, and analyses match the intended subgroup-safety question. Risk-of-bias and applicability assessment can identify weaknesses in selection, measurement, modelling, and transportability [24]. Such assessment does not certify that a model is causally valid or useful in clinical practice.

Responsible reporting should describe data provenance, eligibility, variable definitions, missingness, modelling choices, internal and external validation, performance, subgroup representation, and intended use. TRIPOD+AI extends these transparency requirements to regression and machine-learning prediction models [27]. Compliance with reporting guidance improves interpretability but does not itself establish safety benefit.

The framework proposes seven validation levels: conceptual validity, data validity, analytical validity, external validity, causal and mechanistic coherence, decision validity, and continuing validity. The sequence is not a regulatory maturity scale. Progress at one level does not guarantee progress at the next, and validation requirements should reflect whether the intended use is detection, explanation, prediction, communication, or treatment support [24].

Communication should distinguish a detected difference, a replicated association, a causally supported relationship, and a decision-evaluated tool. Reports should state the affected context, uncertainty, alternative explanations, data limitations, and unsupported uses. Transparent reporting must prevent a model output from being translated directly into categorical labels such as “safe” or “unsafe” for an individual or population [27].

Limitations and research agenda

The framework may organize subgroup-safety investigation, but its components and relationships remain unvalidated as an integrated system. Even a well-performing model can lose calibration when patient populations, prescribing, coding, laboratory practices, or outcome incidence change over time [28]. Aggregate performance may also conceal deterioration within a smaller subgroup.

Dataset shift can arise from changes in clinical practice, population composition, measurement, labels, or the interaction between clinicians and analytical tools [29]. These changes may alter apparent subgroup performance

without any change in underlying pharmacology. Continuing validity therefore requires surveillance of data distributions, subgroup-specific calibration, missingness, and clinical use.

Additional limitations include sparse intersections, incomplete adherence and dose histories, residual confounding, inconsistent event definitions, selective monitoring, missing genetic information, and unequal data quality across populations. Underrepresentation in genetic studies further constrains transportability and may amplify uncertainty for populations already poorly served by existing evidence [21]. The framework cannot recover causal truth from absent or systematically biased data.

Research should first evaluate whether the proposed taxonomy improves explanation and replication of known subgroup-safety problems. Subsequent studies should compare one-variable and multidimensional detection, test cross-source concordance, evaluate dynamic metabolic phenotype and regimen representations, and assess subgroup-specific calibration. Prospective studies should then determine whether framework-informed communication or decision support improves outcomes without increasing inequity, alert burden, or inappropriate treatment restriction [29].

CONCLUSION

Safety assessment after the average patient requires more than dividing a population by one demographic or molecular characteristic. The proposed framework treats heterogeneous drug harm as the possible result of interacting susceptibility, exposure, biological response, concomitant treatment, and observation processes. It separates detection from explanation, causal assessment, external validation, communication, and decision use. Its constructs, propositions, and validation levels provide a research architecture rather than a proven system. Empirical testing, cross-source replication, fairness assessment, and continuing monitoring are required before any clinical, regulatory, or operational application can be justified.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Kent DM, Paulus JK, van Klaveren D, D'Agostino R, Goodman S, Hayward R, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement. Ann Intern Med. 2020;172(1):35-45. doi:10.7326/M18-3667
2. Kent DM, van Klaveren D, Paulus JK, D'Agostino R, Goodman S, Hayward R, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement: Explanation and elaboration. Ann Intern Med. 2020;172(1). doi:10.7326/M18-3668
3. Bate A, Stegmann JU. Safety of medicines and vaccines—building next generation capability. Trends Pharmacol Sci. 2021;42(12):1051-63. doi:10.1016/j.tips.2021.09.007
4. Segal JB, Varadhan R, Groenwold RHH, Henderson NC, Li X, Nomura K, et al. Assessing heterogeneity of treatment effect in real-world data. Ann Intern Med. 2023;176(4):536-44. doi:10.7326/M22-1510
5. Thelen H, Hennessy S. Characterizing treatment effect heterogeneity using real-world data. Clin Pharmacol Ther. 2025;117(5):1209-16. doi:10.1002/cpt.3627
6. Selby JV, Maas CCHM, Fireman BH, Kent DM. Predictive modeling of heterogeneous treatment effects in RCTs: A scoping review. JAMA Netw Open. 2025;8(7). doi:10.1001/jamanetworkopen.2025.22390
7. Mauvais-Jarvis F, Bairey Merz N, Barnes PJ, Brinton RD, Carrero JJ, DeMeo DL, et al. Sex and gender: Modifiers of health, disease, and medicine. Lancet. 2020;396(10250):565-82. doi:10.1016/S0140-6736(20)31561-0
8. Roden DM, McLeod HL, Relling MV, Williams MS, Mensah GA, Peterson JF, et al. Pharmacogenomics. Lancet. 2019;394(10197):521-32. doi:10.1016/S0140-6736(19)31276-0

9. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13). doi:10.1093/bioinformatics/bty294
10. Crisafulli S, Bate A, Brown JS, Candore G, Chandler RE, Hammad TA, et al. Interplay of spontaneous reporting and longitudinal healthcare databases for signal management: Position statement from the Real-World Evidence Working Group of the International Society of Pharmacovigilance. Drug Saf. 2025;48(9):959-76. doi:10.1007/s40264-025-01548-3
11. Norén GN, Donegan K, Muñoz MA, Alshammari TM, Pratt N, Candore G, et al. Recommendations to enable broader use of real-world evidence to inform decision-making throughout pharmacovigilance signal management. Pharmacoepidemiol Drug Saf. 2025;34(10). doi:10.1002/pds.70231
12. Bean DM, Wu H, Iqbal E, Dzahini O, Ibrahim ZM, Broadbent M, et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. Sci Rep. 2017;7(1):16416. doi:10.1038/s41598-017-16674-x
13. Zucker I, Prendergast BJ. Sex differences in pharmacokinetics predict adverse drug reactions in women. Biol Sex Differ. 2020;11(1):32. doi:10.1186/s13293-020-00308-5
14. Patorno E, Huybrechts KF, Bateman BT, Cohen JM, Desai RJ, Mogun H, et al. Lithium use in pregnancy and the risk of cardiac malformations. N Engl J Med. 2017;376(23):2245-54. doi:10.1056/NEJMoa1612222
15. McDonald EG, Wu PE, Rashidi B, Goodwin Wilson M, Bortolussi-Courval É, Atique A, et al. The MedSafer study—electronic decision support for deprescribing in hospitalized older adults: A cluster randomized clinical trial. JAMA Intern Med. 2022;182(3):265-73. doi:10.1001/jamainternmed.2021.7429
16. Nicholson K, Liu W, Fitzpatrick D, Hardacre KA, Roberts S, Salerno J, et al. Prevalence of multimorbidity and polypharmacy among adults and older adults: A systematic review. Lancet Healthy Longev. 2024;5(4). doi:10.1016/S2666-7568(24)00007-2
17. Gutiérrez-Valencia M, Izquierdo M, Cesari M, Casas-Herrero A, Inzitari M, Martínez-Velilla N. The relationship between frailty and polypharmacy in older people: A systematic review. Br J Clin Pharmacol. 2018;84(7):1432-44. doi:10.1111/bcp.13590
18. Laville SM, Gras-Champel V, Hamroun A, Moragny J, Lambert O, Metzger M, et al. Kidney function decline and serious adverse drug reactions in patients with CKD. Am J Kidney Dis. 2024;83(5):601-14.e1. doi:10.1053/j.ajkd.2023.09.012
19. Henricks LM, Lunenburg CATC, de Man FM, Meulendijks D, Frederix GWJ, Kienhuis E, et al. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: A prospective safety analysis. Lancet Oncol. 2018;19(11):1459-67. doi:10.1016/S1470-2045(18)30686-7
20. Klomp SD, Manson ML, Guchelaar HJ, Swen JJ. Phenoconversion of cytochrome P450 metabolism: A systematic review. J Clin Med. 2020;9(9):2890. doi:10.3390/jcm9092890
21. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. Cell. 2019;177(1):26-31. doi:10.1016/j.cell.2019.02.048
22. Wastesson JW, Morin L, Tan ECK, Johnell K. An update on the clinical consequences of polypharmacy in older adults: A narrative review. Expert Opin Drug Saf. 2018;17(12):1185-96. doi:10.1080/14740338.2018.1546841
23. Hughes JE, Moriarty F, Bennett KE, Cahir C. Drug-drug interactions and the risk of adverse drug reaction-related hospital admissions in the older population. Br J Clin Pharmacol. 2024;90(4):959-75. doi:10.1111/bcp.15970
24. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-8. doi:10.7326/M18-1376
25. Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, et al. Development of the instrument to assess the credibility of effect modification analyses (ICEMAN) in randomized controlled trials and meta-analyses. CMAJ. 2020;192(32). doi:10.1503/cmaj.200077
26. Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JPA. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. JAMA Intern Med. 2017;177(4):554-60. doi:10.1001/jamainternmed.2016.9125
27. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385. doi:10.1136/bmj-2023-078378

28. Davis SE, Greevy RA Jr, Lasko TA, Walsh CG, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc. 2017;24(6):1052-61. doi:10.1093/jamia/ocx030
29. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med. 2021;385(3):283-6. doi:10.1056/NEJMc2104626