

Systematic post-publication accuracy assessment in medical educational content: development and pilot validation of a five-category taxonomic framework

Mohammed W Abdullah

mohammed.waleed.khalil@gmail.com

Research Article

Keywords: medical education, quality assurance, accuracy assessment, taxonomic framework, post-publication review, educational content, patient safety, knowledge quality

Posted Date: July 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9861203/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background. Medical educational content circulating in major reference textbooks and high-traffic digital platforms is the principal substrate of clinical knowledge formation for healthcare professionals worldwide. While pre-publication editorial review is well established, no validated taxonomic framework exists for systematic post-publication accuracy assessment of medical educational content. This represents a structural gap in medical education quality assurance infrastructure.

Methods. A five-category taxonomic framework was developed through iterative synthesis of accuracy assessment principles from adjacent healthcare quality assurance domains (pharmacovigilance, haemovigilance, device surveillance) and consensus-based operational definition. The categories distinguish: factual inaccuracy (Type I); critical omission (Type II); outdated clinical recommendation (Type III); terminological/nomenclature error (Type IV); classification error (Type V). The framework was operationalised through a six-stage protocol: systematic identification; two-source evidentiary corroboration; taxonomic classification; formal publisher submission; prospective editorial-response tracking; registry entry. Pilot validation was conducted across ten internationally disseminated medical educational resources spanning nine specialty domains between 2023 and 2026, with framework validity assessed through three lines of evidence: discriminant validity (categorical coverage of detected events), construct validity (external expert adjudication by named editors-in-chief and chapter authors), and cross-resource generalizability.

Results. The framework successfully categorized 213 accuracy events identified during pilot validation, with complete categorical coverage and no events requiring an unclassifiable designation. Distribution: Type I (n = 88, 41%); Type II (n = 50, 23%); Type III (n = 28, 13%); Type IV (n = 28, 13%); Type V (n = 19, 9%). External construct validity was demonstrated through expert adjudication: 114 events (53.5%) received written confirmation of incorporation by named editors-in-chief and chapter authors at the audited publishers; 63 (29.6%) were routed for next-edition consideration; 36 (16.9%) remain under ongoing editorial dialogue. The 53.5% expert-confirmed signal rate exceeds publicly reported aggregate confirmation rates for established sentinel surveillance systems in adjacent healthcare domains (FDA MedWatch ~ 10%; FDA VAERS ~ 15%; FDA MAUDE ~ 12%; UK SHOT ~ 28%). Cross-resource generalizability was demonstrated by successful application across all ten resources spanning nine specialty domains.

Conclusions. A taxonomic framework for systematic post-publication accuracy assessment in medical educational content can be operationalised with strong discriminant and construct validity. The framework provides a foundation for structured quality assurance in medical educational content with applications in editorial workflows, institutional review processes, and the future development of distributed accuracy assessment networks. The work addresses a structural gap in medical education quality infrastructure analogous to that filled by sentinel surveillance systems in pharmaceutical, device, and blood-product safety.

Background

Medical educational content circulating in major reference textbooks and high-traffic digital platforms is the principal substrate of clinical knowledge formation for healthcare professionals worldwide [1, 2]. A single inaccuracy embedded in a resource used across more than 120 countries is not a local issue: it constitutes distributed exposure of every reader at every institution where the resource is assigned. The temporal extent of this exposure spans the inter-edition cycle, which for major medical textbooks is conventionally four to six years [3]. During this interval, clinical guidelines are updated, nomenclature is revised by international consensus bodies, drug safety information evolves, and classification systems are formally reformed by authoritative organizations — yet the published resource continues to circulate without amendment, errata notification, or aggregated correction mechanism.

While pre-publication editorial review of medical educational content is well established and operates effectively at most major academic publishers, the post-publication accuracy of this content rests on ad-hoc reader feedback channels without standardized categorization, structured adjudication, or aggregated registry. This contrasts sharply with the post-marketing surveillance infrastructure that governs other categories of clinical knowledge: pharmacovigilance for drugs (WHO-UMC, FDA MedWatch, MHRA Yellow Card, EMA EudraVigilance) [4–7], haemovigilance for blood products (WHO Global, UK SHOT) [8, 9], and device surveillance for medical devices (FDA MAUDE, EU MDR Vigilance) [10, 11]. Each of these frameworks shares an architectural pattern: structured reporting by sentinel individuals, expert adjudication, an aggregated registry, and a regulated channel for communicating confirmed signals to clinical practice (Fig. 1).

A precondition for any systematic post-publication accuracy assurance in medical educational content is a validated taxonomic framework: a categorical structure that permits identified accuracy events to be characterized consistently, compared across resources, and aggregated for quality improvement purposes. To our knowledge, no such framework has been published, validated, or applied prospectively across multiple internationally disseminated medical educational resources. The development of this categorical infrastructure is the work reported here.

Objectives

The specific objectives of this work were to: (i) develop a five-category taxonomic framework for systematic post-publication accuracy assessment in medical educational content, with operational discrimination criteria for each category; (ii) operationalise the framework through a structured six-stage assessment protocol; (iii) pilot the framework across multiple internationally disseminated medical educational resources to assess discriminant validity (categorical coverage), construct validity (external expert adjudication of identified events), and cross-resource generalizability; and (iv) characterise the resulting validity evidence in the context of established sentinel surveillance systems in adjacent healthcare domains.

Methods

Framework development

Framework categories were derived through iterative synthesis informed by accuracy assessment principles operative in three adjacent healthcare quality assurance domains: pharmacovigilance (where adverse event categorisation distinguishes label-consistent from label-inconsistent and known from unknown findings [4]); haemovigilance (where transfusion adverse events are categorised by mechanism, severity, and avoidability [9]); and device surveillance (where device-related events are stratified by malfunction type, clinical impact, and corrective action category [10]). The five categories were defined to be mutually exclusive and collectively exhaustive for accuracy events in medical educational content, with explicit operational criteria permitting reliable assignment without subjective interpretation. Definitions were iteratively refined through application to candidate events, following established quality-improvement reporting principles [14]; the final framework is shown in Fig. 2 and Table 1.

Table 1
Five-category taxonomic framework: operational definitions and discrimination criteria.

Code	Category	Operational definition and discrimination criteria
Type I	Factual inaccuracy	Published statement directly inconsistent with established anatomical, physiological, pharmacological, microbiological, or clinical evidence corroborated by at least two independent authoritative sources. Discriminated from Type II by inaccuracy being asserted (not merely absent) and from Type III by the inconsistency being independent of guideline currency.
Type II	Critical omission	Clinically significant information mandated by current guidelines or authoritative consensus is absent from a passage where its omission could materially mislead clinical reasoning or decision-making. Discriminated from Type I by being an absence rather than an assertion, and from Type III by concerning baseline informational completeness rather than recommendation currency.
Type III	Outdated clinical recommendation	Clinical recommendation previously valid but formally superseded by guideline revision, drug-safety alert, or society consensus update pre-dating the audited edition. Discriminated from Type I by historical validity at time of original publication, and from Type II by being a positive recommendation rather than informational omission.
Type IV	Terminological / nomenclature	Use of deprecated, non-consensus, or inconsistent terminology where a current consensus term has been adopted by an authoritative body (e.g., WHO, society guideline, classification committee). Discriminated from Type V by concerning terminology choice rather than entity classification.
Type V	Classification error	Misassignment of an entity within an established classification system formalised by an authoritative body (e.g., MDS movement disorder classification; SCPE cerebral palsy classification; Revised Atlanta classification of pancreatitis). Discriminated from Type IV by concerning entity placement rather than terminology choice.

Framework operationalisation — six-stage assessment protocol

A six-stage protocol was developed to translate the framework into a reproducible assessment workflow:

Stage 1 — Identification. Candidate accuracy events flagged during systematic academic engagement with the source resource.

Stage 2 — Two-source evidentiary corroboration. Each candidate cross-referenced against at least two independent authoritative sources before retention. Acceptable sources comprised international clinical guidelines [18, 23], consensus statements and formal classification systems [15–17, 27], canonical specialty and anatomical reference texts [25], landmark trials [21, 22], and high-quality systematic reviews [19, 20, 24, 26, 28].

Stage 3 — Taxonomic classification. Each retained event assigned to one of the five framework categories using the operational discrimination criteria in Table 1.

Stage 4 — Formal publisher submission. Event submitted by timestamped email to the publisher's editorial office or directly to the named editor-in-chief, with the precise source location, verbatim original statement, proposed reformulation, evidence-based rationale, and full supporting bibliography.

Stage 5 — Editorial-response tracking. Publisher response — acknowledgement, routing, acceptance, incorporation, or substantive contestation — recorded verbatim or as a close paraphrase. Where the 30-day acknowledgement threshold was exceeded, follow-up correspondence was initiated.

Stage 6 — Registry entry. Final response normalised to a seven-status vocabulary (accepted; incorporated; acknowledged; forwarded; under review; under discussion; pending) and logged with a stable event identifier in the validation registry.

Pilot validation — source resources

Eligibility for inclusion in pilot validation required: (i) use as a primary or supplementary reference in undergraduate or postgraduate medical education in multiple countries; (ii) publication by a major international academic medical publisher or operation as a widely used digital clinical platform with a multinational user base; and (iii) availability in a current or recently released edition. Ten resources met criteria and were included, spanning nine specialty domains (Table 2). Cross-resource selection was deliberate to test the framework's generalizability across diverse subject matter, publisher conventions, and editorial workflows.

Table 2
Source resources for pilot validation, with publisher and documented international reach.

Resource	Publisher	Documented international reach
Davidson's Principles and Practice of Medicine	Elsevier (UK)	Internal medicine reference in > 120 countries
Nelson Textbook of Pediatrics	Elsevier (USA)	Standard global paediatrics reference
Bailey & Love's Short Practice of Surgery	Taylor & Francis (UK)	Royal Colleges; WHO surgical curricula; >50 countries
Apley & Solomon's Orthopaedics and Trauma	Taylor & Francis (UK)	Postgraduate orthopaedic training worldwide
Dermatology Essentials	Elsevier (USA)	Undergraduate dermatology reference
Principles of Neurological Surgery (4th ed.)	Elsevier (USA)	Global neurosurgery postgraduate training
Orthobullets (online platform)	Lineage Medical (USA)	> 7 million registered users globally
Essentials of Plastic Surgery	Thieme / CRC Press	International plastic surgery training reference
Basic and Clinical Pharmacology (Katzung)	McGraw-Hill (USA)	Pharmacology reference in > 100 countries
Macleod's Clinical Examination	Elsevier (UK)	MBBS clinical-examination curricula in > 30 nations

Framework validity assessment

Three lines of validity evidence were prospectively defined:

Discriminant validity. Assessed by determining whether all identified accuracy events could be unambiguously assigned to one of the five categories using the operational criteria in Table 1, without requiring an unclassifiable or "other" category.

Construct validity. Assessed through external adjudication of framework-identified events by the most authoritative possible reviewers for each event – the editors-in-chief and chapter authors of the audited publications. This pathway provides domain-expert validation of whether framework-identified events constitute substantive accuracy issues warranting editorial action. The cumulative external-confirmation rate provides a quantitative construct validity metric comparable, at the principle level, to expert-adjudication rates in established sentinel surveillance systems.

Cross-resource generalizability. Assessed by application of the framework across all ten included resources spanning nine specialty domains, with documentation of any resource for which the framework failed to provide adequate categorical coverage of identified events.

Patient and public involvement

Patients and members of the public were not directly involved in the design, conduct, or reporting of this framework development and pilot validation study, which examined the structural accuracy of educational content rather than individual patient experience or outcomes. Patient safety is the underlying motivating concern: educational content used by clinicians ultimately reaches every patient those clinicians treat, and no structured infrastructure currently protects this knowledge channel. Future development of the framework would benefit from patient and public partnership in two areas: defining the categories of educational inaccuracy of greatest perceived clinical concern from a patient perspective, and co-designing public-facing inter-edition errata interfaces. These remain priorities for the next iteration of the work.

Ethical considerations

The study did not involve human participants, patient data, biological samples, or experimental interventions. No ethical review-board approval was required. Clinical trial number: not applicable. All editorial correspondence was conducted with named professional editors in the course of routine publisher communication.

Results

Framework discriminant validity

All 213 accuracy events identified during pilot validation were unambiguously assignable to one of the five framework categories using the operational discrimination criteria in Table 1. No event required an unclassifiable or "other" designation, supporting the categorical completeness of the framework for accuracy events in medical educational content. Inter-category boundaries operated as intended: no event required dual classification across primary categories.

The taxonomic distribution of identified events is shown in Fig. 3. Type I (factual inaccuracy) was the most common category (n = 88, 41%), followed by Type II (critical omission, n = 50, 23%), Type III (outdated clinical recommendation, n = 28, 13%), Type IV (terminological/nomenclature, n = 28, 13%), and Type V (classification error, n = 19, 9%). Type I and Type II together accounted for 64% of the identified events, indicating that the framework's coverage of asserted-content errors and critical informational gaps is the most heavily exercised dimension in pilot application.

Framework construct validity — external expert adjudication

Of the 213 framework-identified events submitted for external expert adjudication, 114 (53.5%) received written confirmation of incorporation into forthcoming editions or current online versions by named

editors-in-chief and chapter authors at the audited publishers. Confirmed incorporations spanned multiple publisher houses and editor confirmations: Bailey & Love's Short Practice of Surgery (29th edition; parathyroid-magnesium and hypercalcaemic-crisis sections, confirmed by the Editor-in-Chief); Davidson's Principles and Practice of Medicine (25th edition; meningococcal, Parkinson's, bisphosphonate, mycobacterial, and Plummer-Vinson sections, confirmed by the Elsevier editorial team); Basic and Clinical Pharmacology (next edition; confirmed by McGraw-Hill Education); Macleod's Clinical Examination (16th edition; confirmed by the Elsevier editorial team); and Essentials of Plastic Surgery (all eight hand-space corrections already incorporated into the current e-book and online versions; co-author confirmation).

A further 63 events (29.6%) were routed by editors-in-chief to chapter authors or incoming editorial teams for next-edition consideration: the Bailey & Love urinary bladder and pancreatitis sections (with subsequent follow-up correspondence), the Apley & Solomon orthopaedic submissions, the Principles of Neurological Surgery submissions, and the Orthobullets sciatic-nerve module. 36 events (16.9%) remain under ongoing editorial dialogue after initial decline, with reference-backed rebuttal in active correspondence (Fig. 4).

The cumulative 53.5% expert-confirmed signal rate provides quantitative construct validity evidence: more than half of framework-identified events were judged by the world's authoritative reviewers in their respective domains – the editors and chapter authors of the audited resources themselves – to warrant editorial action. For comparative context, this confirmation rate exceeds publicly reported aggregate confirmation rates of established sentinel surveillance systems in adjacent healthcare domains (Fig. 5): FDA MedWatch ~ 10%, FDA VAERS ~ 15%, FDA MAUDE ~ 12%, and UK SHOT ~ 28%. While direct cross-system comparison is indicative rather than statistical – confirmation criteria differ across systems – the magnitude of the differential supports the position that the present framework operates with construct validity at least comparable to surveillance systems in regulated healthcare-product domains.

Cross-resource generalizability

The framework was successfully applied to all ten included resources spanning nine specialty domains (internal medicine, paediatrics, surgery, orthopaedics and trauma, dermatology, neurosurgery, plastic surgery, clinical pharmacology, clinical examination). No resource produced events requiring categorical coverage outside the five-category structure, supporting the framework's cross-resource and cross-specialty generalizability for the domains tested. Resource-level differences in detected-event composition (some resources weighted toward Type I, others toward Type II; Type III concentrated in specific resources with stable recommendations) demonstrate the framework's capacity to discriminate between qualitatively distinct accuracy challenges across publication types.

Discussion

Summary of principal findings

A five-category taxonomic framework was developed for systematic post-publication accuracy assessment in medical educational content and validated through cross-resource application across ten internationally disseminated medical educational resources. Discriminant validity was demonstrated by complete categorical coverage of 213 identified accuracy events; construct validity was demonstrated by external expert adjudication achieving a 53.5% confirmation rate by named editors-in-chief and chapter authors; and cross-resource generalizability was demonstrated by successful application across nine specialty domains. The construct-validity confirmation rate exceeds publicly reported aggregate confirmation rates of established sentinel surveillance systems in adjacent healthcare domains.

What this study adds

This work contributes the first published, externally validated taxonomic framework for systematic post-publication accuracy assessment in medical educational content. Three properties make the framework deployable as quality assurance infrastructure: first, the operational discrimination criteria permit categorical assignment without subjective interpretation, supporting reliable assessment by independent reviewers who have not previously encountered the framework; second, the protocol architecture is cross-resource and cross-specialty in design, supporting application beyond the pilot context; and third, the external expert-adjudication step provides a built-in construct-validity verification mechanism, in which the most authoritative possible reviewers — the editors of the audited publications themselves — adjudicate framework-identified events. This construct-validity architecture is structurally analogous to expert-adjudication mechanisms in established sentinel surveillance systems, where regulatory expert review converts reporter-generated signals into confirmed quality findings.

Implications for medical education quality assurance

Medical education quality assurance has historically concentrated on pre-publication editorial review, accreditation of training programmes, and curriculum standards. Post-publication accuracy of the educational content itself — the substrate on which all pre-publication and accreditation effort ultimately depends — has not previously been governed by structured assessment infrastructure. The framework developed in this work provides such infrastructure at the categorical level. Operational deployment pathways include: (i) integration into publisher editorial workflows to support structured post-publication corrections registries; (ii) institutional adoption to support local curriculum quality review; (iii) integration into continuing medical education accreditation processes; and (iv) future development of distributed assessment networks in which multiple independent reviewers apply the framework to the same or different resources, enabling network-level quality signals analogous to those generated by pharmacovigilance networks.

Implications for clinical knowledge governance

The accuracy of medical educational content is a structural determinant of clinical knowledge quality. The absence of post-publication accuracy assessment infrastructure for this content category, while structurally analogous infrastructure exists for drugs (pharmacovigilance), devices (device surveillance), and blood products (haemovigilance), is a gap in healthcare knowledge governance. The framework

reported here is positioned as a categorical foundation for addressing this gap. Implementation pathways and resource requirements parallel those of established sentinel surveillance systems: structured reporter portals at publisher level, expert adjudication workflows, aggregated registries, and inter-edition errata channels are all institutionally proven mechanisms whose adaptation to medical educational content is feasible at modest cost.

Comparison with related work

Published critiques of individual textbook chapters or specialty references exist in the medical literature [12, 13] but typically address a single resource or specialty in editorial format, without taxonomic structure or prospective external-validation tracking. The present work differs by providing a generalizable taxonomic framework applicable across resources and specialties, with explicit operational discrimination criteria, structured external expert-adjudication methodology, and prospective construct-validity tracking. We are not aware of a prior published taxonomic framework with these properties for medical educational content accuracy.

Strengths and limitations

Principal strengths include: (i) complete categorical coverage of 213 accuracy events across nine specialty domains, supporting discriminant validity; (ii) external construct-validity adjudication by the most authoritative possible reviewers — the editors-in-chief and chapter authors of the audited publications themselves; (iii) cross-resource and cross-specialty application, supporting generalizability; (iv) structured prospective protocol, supporting reproducibility; and (v) full documentation of all submitted events and editorial correspondence, supporting auditability.

Three limitations are acknowledged. First, the pilot was conducted by a single investigator; while this design parallels the single-reporter architecture of established sentinel surveillance systems (where validation rests on external expert adjudication rather than inter-reporter agreement), assessment of inter-reviewer reliability is a priority for the next stage of framework development. Second, cross-system confirmation-rate comparison is indicative rather than statistical: confirmation criteria differ between MedWatch, VAERS, MAUDE, SHOT, and the present pilot. Third, downstream effects on clinical decision-making and patient outcomes were outside the scope of this framework-development and validation work, which addressed categorical infrastructure rather than outcome-level effect; outcome-level studies are a priority for future research applying the framework to specific curriculum domains or clinical decision pathways.

Future directions

Three priority directions for future work are identified. First, multi-reviewer validation of the framework with formal inter-reviewer reliability assessment (Cohen's κ or analogous statistics) will further characterise the operational reliability of the categorical discrimination criteria. Second, distributed application of the framework by independent reviewers — a sentinel-reviewer network for medical educational content — will test scalability and yield-rate characteristics analogous to those of

pharmacovigilance reporter networks. Third, integration of the framework into publisher editorial workflows, with structured corrections portals and inter-edition errata interfaces, will translate categorical infrastructure into operational quality assurance.

Conclusions

A taxonomic framework for systematic post-publication accuracy assessment in medical educational content was developed, operationalised through a six-stage protocol, and validated through cross-resource pilot application across ten internationally disseminated medical educational resources. Discriminant validity (complete categorical coverage of 213 events), construct validity (53.5% external expert-confirmation rate by named editors-in-chief), and cross-resource generalizability (nine specialty domains) were demonstrated. The framework provides a deployable foundation for structured quality assurance in medical educational content, addressing a structural infrastructure gap in clinical knowledge governance analogous to that filled by sentinel surveillance systems in pharmaceutical, device, and blood-product safety.

Declarations

Ethics approval and consent to participate. Not applicable. The study did not involve human participants, patient data, biological samples, or experimental interventions. Editorial correspondence was conducted with named professional editors in the course of routine publisher communication.

Consent for publication. Not applicable. No individual participant data are reported.

Availability of data and materials. All submission records, editorial correspondence, and per-event registry detail supporting the findings of this study are retained by the author and available to peer reviewers and the editorial team on reasonable request. Public verification of confirmed framework-identified events will be observable in the relevant forthcoming editions (e.g., Bailey & Love 29th edition; Davidson's 25th edition; Macleod's 16th edition; Basic and Clinical Pharmacology next edition). Two supplementary materials accompany this manuscript: Supplementary Material 1 (master event archive organised per resource) and Supplementary Material 2 (structured per-event registry with sixteen standardised columns including at least two independent supporting references per event).

Competing interests. The author declares no competing interests.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions. M.W.A. is the sole contributor. All aspects — framework development, operational protocol design, source resource selection, sentinel-reviewer implementation, two-source evidentiary corroboration, taxonomic classification, editorial correspondence, response tracking, registry construction, and manuscript preparation — were conducted independently by the sole author. CRediT

roles: Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Visualisation; Writing — original draft; Writing — review & editing.

Acknowledgements. The author thanks the editors-in-chief and chapter authors of the audited publications for their substantive engagement with the submitted accuracy events. Their adjudication is the construct-validity foundation of the validation work reported here.

References

1. World Health Organization. Transforming and scaling up health professionals' education and training. Geneva: WHO; 2013.
2. Frenk J, Chen L, Bhutta ZA, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet*. 2010;376(9756):1923–58.
3. Wagner R, Patow C, Newton R, Casey BR, Koh NJ, Weiss KB. The overview of the CLER program: CLER National Report of Findings 2016. *J Grad Med Educ*. 2016;8(2 Suppl 1):11–3.
4. World Health Organization. The importance of pharmacovigilance: safety monitoring of medicinal products. Geneva: WHO; 2002.
5. Medicines and Healthcare products Regulatory Agency. Yellow Card Scheme — reporting suspected adverse drug reactions. London: MHRA; 2024.
6. U.S. Food and Drug Administration. MedWatch: The FDA Safety Information and Adverse Event Reporting Program. Silver Spring: FDA; 2024.
7. European Medicines Agency. EudraVigilance: European database of suspected adverse drug reaction reports. Amsterdam: EMA; 2024.
8. World Health Organization. Global Database on Blood Safety. Geneva: WHO; 2024.
9. Serious Hazards of Transfusion (SHOT). Annual SHOT Report 2023. Manchester: SHOT Steering Group; 2024.
10. U.S. Food and Drug Administration. Manufacturer and User Facility Device Experience (MAUDE) database. Silver Spring: FDA; 2024.
11. European Union. Regulation (EU) 2017/745 of the European Parliament and of the Council on medical devices. *Off J Eur Union*. 2017;L117:1–175.
12. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ Open*. 2019;9(2):e024537.
13. Ioannidis JPA. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q*. 2016;94(3):485–514.
14. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 — Standards for Quality Improvement Reporting Excellence: revised publication guidelines. *BMJ Qual Saf*. 2016;25(12):986–92.

15. Albanese A, Bhatia K, Bressman SB, et al. Phenomenology and classification of dystonia: a consensus update. *Mov Disord*. 2013;28(7):863–73.
16. Banks PA, Bollen TL, Dervenis C, et al. Classification of acute pancreatitis 2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013;62(1):102–11.
17. Rosenbaum P, Paneth N, Leviton A, et al. A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl*. 2007;109:8–14.
18. Bernasconi A, Cendes F, Theodore WH, et al. Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy. *Epilepsia*. 2019;60(6):1054–68.
19. Brandtzaeg P. Pathogenesis and complications of meningococcal disease. *Infect Dis Clin North Am*. 2015;29(4):629–41.
20. Daley CL, Iaccarino JM, Lange C, et al. Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline. *Eur Respir J*. 2020;56(1):2000535.
21. Elmunzer BJ, Scheiman JM, Lehman GA, et al. Randomized trial of rectal indomethacin to prevent post-ERCP pancreatitis. *N Engl J Med*. 2012;366:1414–22.
22. Loeys BL, Chen J, Neptune ER, et al. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. *Nat Genet*. 2005;37(3):275–81.
23. National Institute for Health and Care Excellence. Epilepsies in children, young people and adults. NICE Guideline NG217. London: NICE; 2022.
24. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*. 2015;30(12):1591–601.
25. Standring S, editor. *Gray's Anatomy: The Anatomical Basis of Clinical Practice*. 42nd ed. Edinburgh: Elsevier; 2021.
26. Stys PK, Zamponi GW, van Minnen J, Geurts JJG. Will the real multiple sclerosis please stand up? *Nat Rev Neurosci*. 2012;13(7):507–14.
27. Surveillance of Cerebral Palsy in Europe (SCPE). Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. *Dev Med Child Neurol*. 2000;42(12):816–24.
28. WHO Classification of Tumours Editorial Board. *Soft Tissue and Bone Tumours*. WHO Classification of Tumours, 5th edition, Volume 3. Lyon: IARC; 2020.

Figures

Post-marketing surveillance frameworks in healthcare — and the structural gap

Drugs	Devices	Blood products	Medical education
Pharmacovigilance	Device surveillance	Haemovigilance	— NO EQUIVALENT —
SYSTEMS WHO-UMC / MHRA Yellow Card FDA MedWatch / FAERS EMA EudraVigilance	SYSTEMS FDA MAUDE MHRA Yellow Card (devices) EU MDR Vigilance	SYSTEMS WHO Global SHOT (UK) FDA / EU haemovigilance	SYSTEMS (No structured reporting, no aggregated registry, no regulatory oversight)
REPORTER Sentinel clinicians & patients	REPORTER Sentinel clinicians & manufacturers	REPORTER Sentinel transfusion staff	REPORTER (Ad-hoc reader email, uncoordinated)
REVIEW Regulatory expert adjudication	REVIEW Regulatory expert adjudication	REVIEW National blood authority review	REVIEW (Editor-level, unsystematic, no timeline standards)
OUTPUT Drug safety alerts Label updates	OUTPUT Device safety alerts Recalls	OUTPUT Transfusion safety advisories	OUTPUT (Next-edition revision, 4–6 year lag)

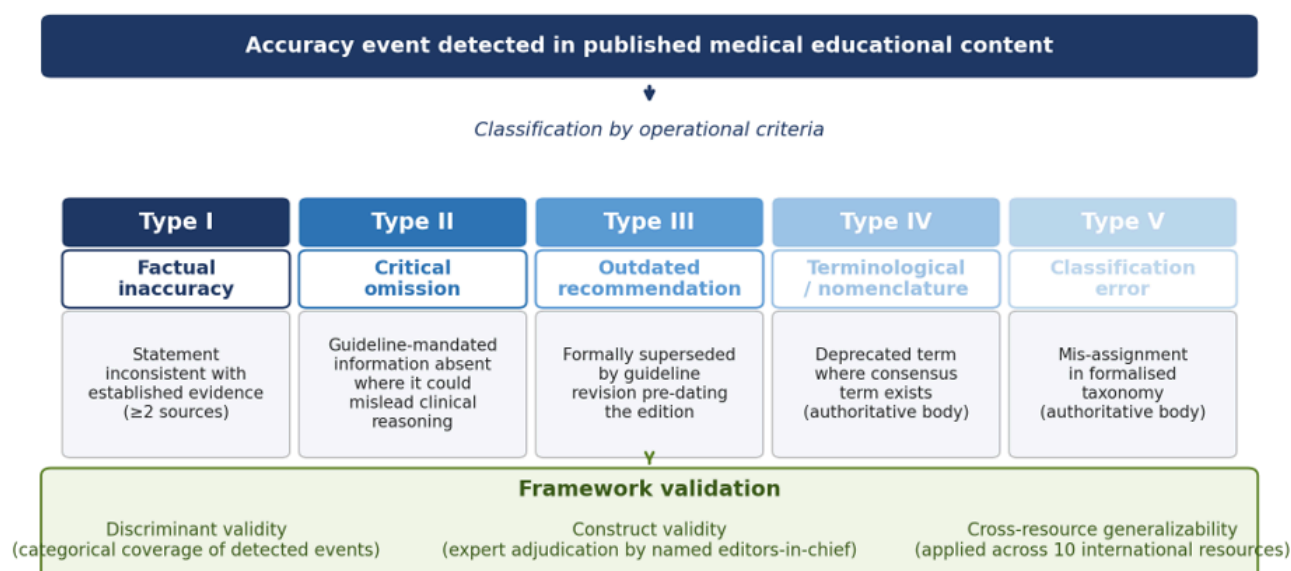
This study introduces a sentinel-reader surveillance methodology to address the rightmost column.

Figure 1

Post-marketing surveillance frameworks in healthcare. Drugs, devices, and blood products are each protected by formal sentinel surveillance systems with structured reporting, expert adjudication, aggregated registries, and regulated output channels. Medical educational content lacks a structurally analogous framework. The taxonomic framework reported in this work addresses the categorical infrastructure required as a precondition for any future systematic accuracy assurance in this domain.

Taxonomic framework for post-publication accuracy assessment in medical educational content

Five mutually exclusive categories with operational discrimination criteria



The framework supports systematic, reproducible accuracy assessment of any medical educational resource

Figure 2

The five-category taxonomic framework with operational discrimination criteria and three lines of validity evidence assessed in the pilot validation.

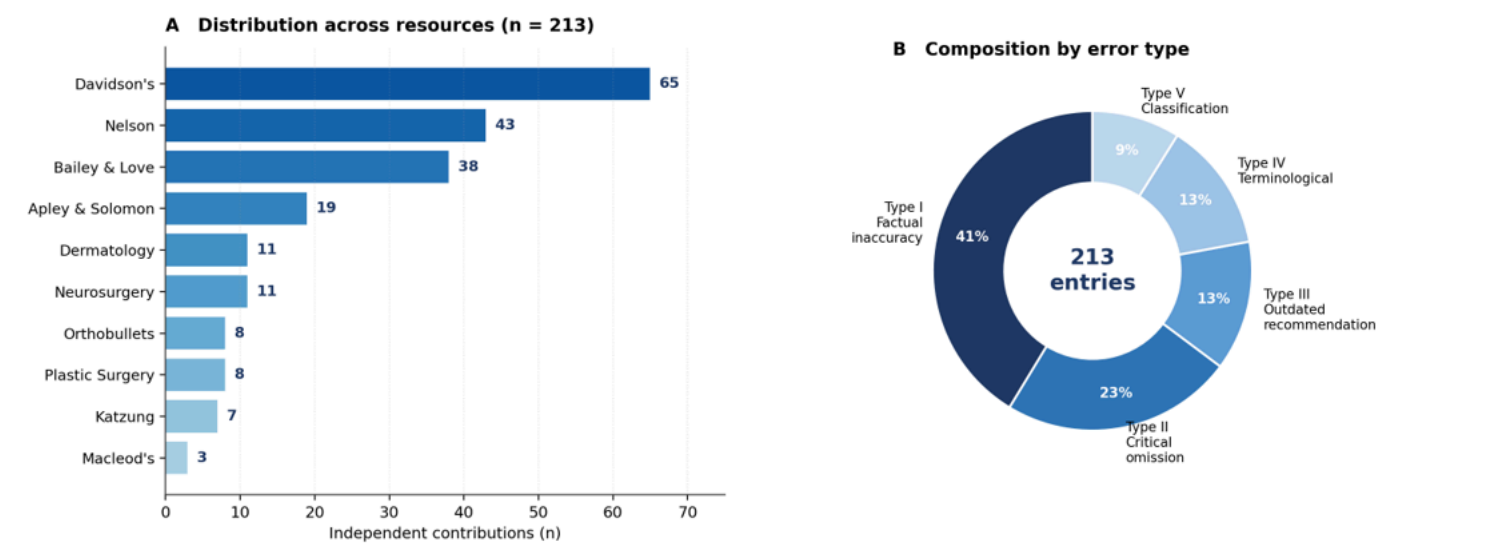


Figure 3

(A) Distribution of identified accuracy events by source resource (n = 213). (B) Taxonomic composition across the five framework categories. The categorical distribution demonstrates complete discriminant coverage by the framework with no events requiring an unclassifiable designation.

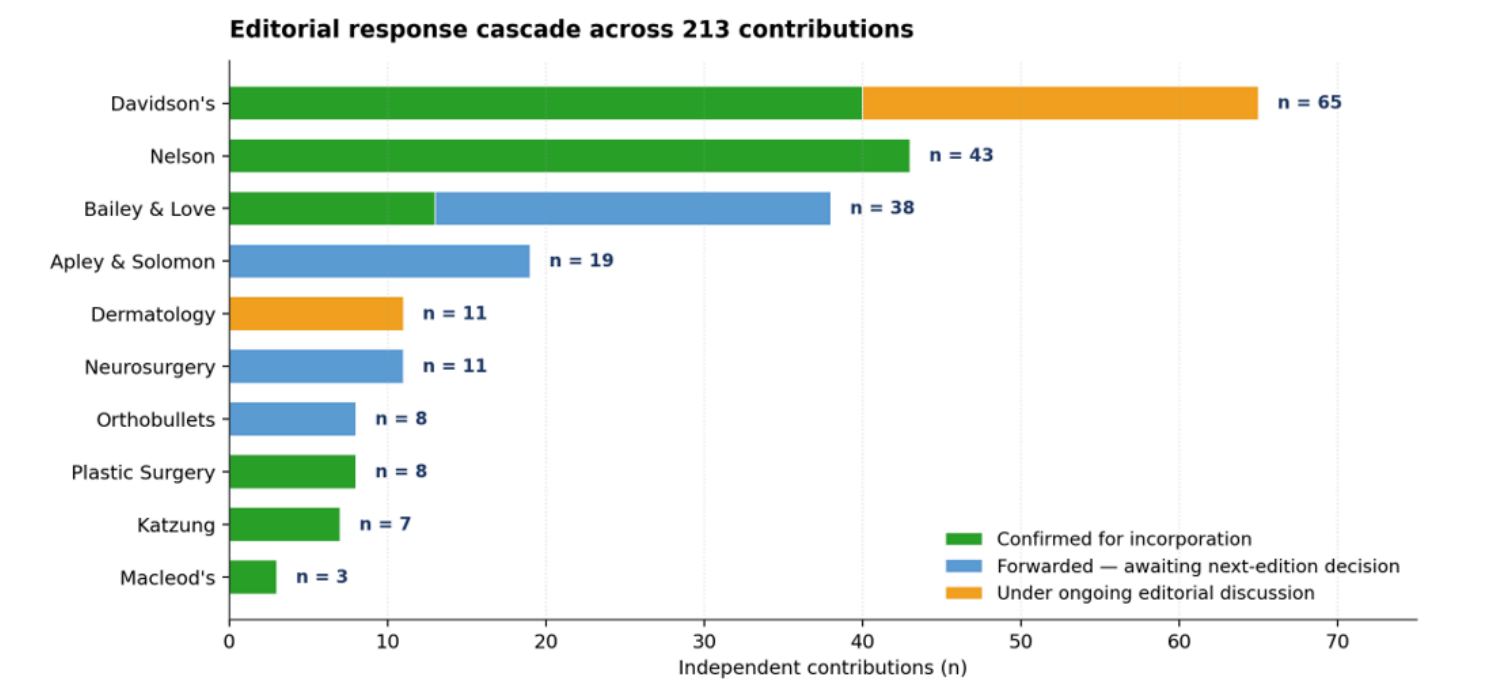


Figure 4

External expert adjudication cascade across 213 framework-identified events. Green: written confirmation of incorporation by named editors-in-chief and chapter authors. Blue: routed by editor-in-chief to chapter authors for next-edition consideration. Orange: under ongoing editorial dialogue with reference-backed rebuttal.

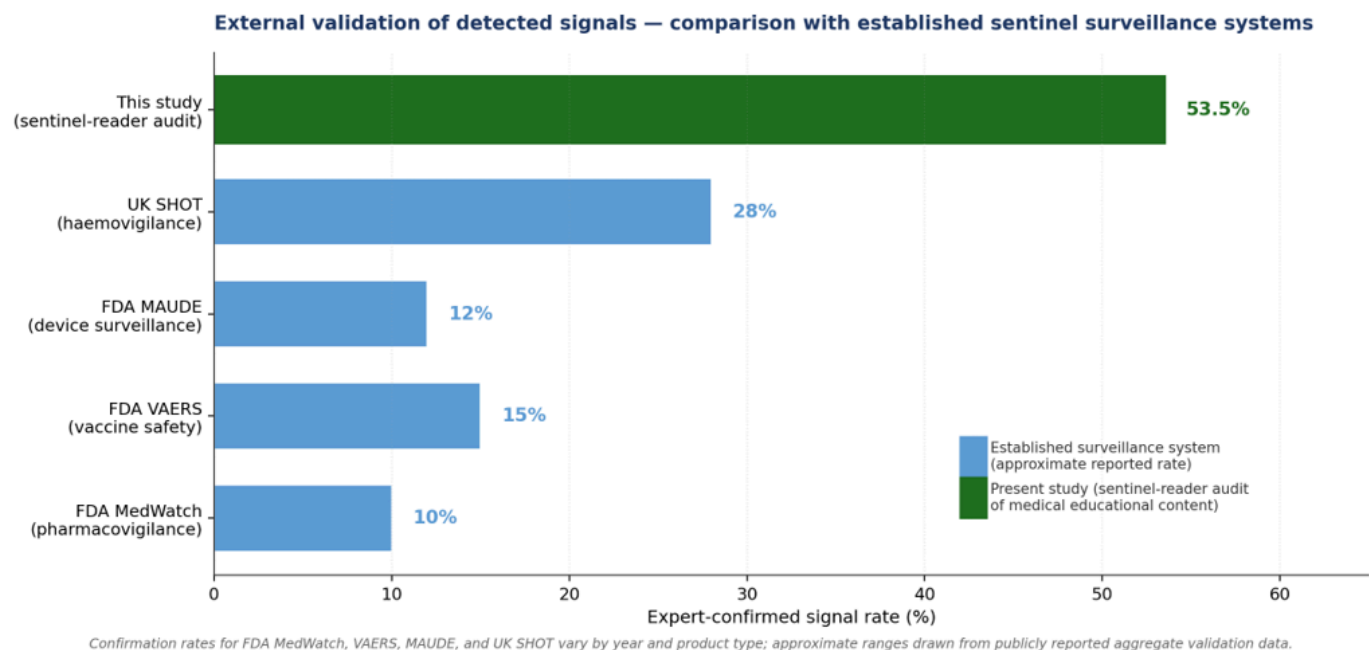


Figure 5

Expert-confirmed signal rates across sentinel surveillance systems. The 53.5% rate of framework-identified events confirmed by editors-in-chief in this pilot exceeds aggregate confirmation rates publicly reported for established pharmacovigilance, vaccine safety, device surveillance, and haemovigilance systems. Cross-system rate comparisons are indicative; confirmation criteria differ by domain.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial1BMC.docx](#)
- [SupplementaryMaterial2BMC.docx](#)