

Protocol

AI for Causality Assessment in Pharmacovigilance: Protocol for a Scoping Review

Miki Ohta¹, MD, MPH, PhD; Miki Ota², PharmD; Mikihiro Ohta³, JD

¹Clinical Research Promotion Center, The University of Tokyo Hospital, Tokyo, Japan

²Office of Informatic and Management for Safety, Pharmaceuticals and Medical Devices Agency, Tokyo, Japan

³Fuji Partnership Law Firm, Tokyo, Japan

Corresponding Author:

Miki Ohta, MD, MPH, PhD

Clinical Research Promotion Center

The University of Tokyo Hospital

7-3-1 Hongo, Bunkyo-ku

Tokyo, 113-8655

Japan

Phone: 81 0338155411

Email: mjohta@g.ecc.u-tokyo.ac.jp

Abstract

Background: Pharmacovigilance aims to protect patient safety by identifying and managing adverse events associated with pharmaceuticals. Determining the causality of these adverse events is central at both the individual case and population levels; however, it is increasingly challenging as the volume and complexity of safety data grow. Although AI and related technologies have been proposed to support causality assessment, limited research has examined how these methods are used, their information and quality requirements, or how associated risks are addressed.

Objective: This scoping review aims to determine the available evidence on AI-based methods for causality assessment in pharmacovigilance. The primary objective is to characterize how these methods are applied or proposed with a focus on their functional roles, reported data inputs and information needs, and associated risks. Secondary objectives include comparing applications at the individual case and population levels; describing the types of AI-based techniques and automation tools used in causality assessment workflows; and summarizing reported data quality considerations and governance mechanisms, including risk management approaches.

Methods: Sources describing or proposing AI-based approaches, including data-driven models (machine learning, natural language processing, knowledge graphs, and causal inference) and knowledge- or rule-based systems implementing causal assessment logic, will be eligible. Searches will be conducted in PubMed, Web of Science Core Collection, ProQuest, EBSCOhost, and Ichushi Web and will be restricted to English- and Japanese-language sources. Two reviewers will independently screen records and full-text articles, with disagreements resolved by a third reviewer. Data will be charted on use cases, information inputs, data quality dimensions, model characteristics, governance mechanisms, and identified risks. Synthesis will follow a reflexive thematic analysis approach and be reported in accordance with PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines informed by applicable PRISMA-S (Preferred Reporting Items for Systematic reviews and Meta-Analyses literature search extension) elements.

Results: This protocol was registered in the Open Science Framework platform on December 23, 2025. The registration was subsequently updated on May 19, 2026, to reflect an extension to the data collection period. A preliminary database search was conducted in December 2025, retrieving a total of 760 records, of which the preliminary title and abstract screening identified 196 (25.8%) articles for full-text review. Database searches are scheduled for July 2026. Data charting is scheduled for August 2026, and synthesis is scheduled for September 2026. Findings are expected to be submitted for publication by the end of December 2026.

Conclusions: This review is expected to provide a structured map of AI-based applications for causality assessment in pharmacovigilance, clarify reported information inputs and data quality dimensions, and synthesize risk management and governance approaches. The findings are expected to inform methodological development, practical implementation, and the governance of AI-supported causality assessment.

Trial Registration: Open Science Framework 10.17605/OSF.IO/QVF5C; <https://osf.io/qvf5c/overview>

International Registered Report Identifier (IRRID): DERR1-10.2196/101691

(*JMIR Res Protoc* 2026;15:e101691) doi: [10.2196/101691](https://doi.org/10.2196/101691)

KEYWORDS

artificial intelligence; AI; pharmacovigilance; causality assessment; machine learning; scoping review

Introduction

Pharmacovigilance aims to protect patient safety by detecting, assessing, understanding, and preventing adverse events associated with medicinal products [1]. A central activity in pharmacovigilance is the assessment of causal relationships between a drug and an adverse event at both the individual-case and population levels [2]. At the individual-case level, causality assessment is conducted in clinical settings, by regulatory authorities, and by the pharmacovigilance departments of pharmaceutical companies, most often through clinical judgment alone or in combination with standardized tools (eg, the World Health Organization Uppsala Monitoring Centre scale [3] and the Naranjo algorithm [4]). At the population level, statistical and epidemiological methods such as disproportionality analyses and observational studies support signal detection and benefit-risk assessment [5].

The rapid expansion in both the volume and complexity of safety data, driven by electronic reporting, large numbers of spontaneous reports, and diverse real-world data sources, has made timely and consistent causality assessment increasingly difficult to achieve through manual review alone. Accordingly, interest has grown in the use of advanced IT, particularly AI, sometimes in combination with automation tools such as robotic process automation to support and potentially transform causality assessment in pharmacovigilance.

Recent developments in AI, including machine learning, natural language processing, knowledge graphs, and causal inference methods, have been explored for various pharmacovigilance tasks such as automated coding [6], duplicate detection [7], case triage, signal detection and prioritization [8], and literature screening [9]. Concurrently, robotic process automation has been introduced to automate repetitive tasks such as case intake, data transfer, and report generation [10]. Although many of these applications primarily aim to improve efficiency and consistency, AI-based models are increasingly being developed to contribute directly to causality assessment—for example, by predicting causality categories for individual cases [11] or identifying patterns indicative of causal associations in large databases [12].

The application of AI-based methods to causality assessment raises several methodological, operational, and governance issues [11,13,14]. These systems must integrate structured and unstructured data at the individual case level and heterogeneous real-world data at the population level, often supplemented by external knowledge databases [11,14]. In addition, pharmacovigilance databases are affected by missing data, underreporting, misclassification, and reporting bias, all of which may influence AI performance [15-17]. Consequently,

data quality metrics and performance indicators beyond standard measures of accuracy must be considered to ensure that outputs align with clinically and regulatorily meaningful outcomes. Explainability and human-in-the-loop designs also remain essential as ultimate decisions regarding causality and safety actions continue to rest with experts [18-20]. Furthermore, AI-based systems and associated automation introduce risks such as algorithmic bias, model drift, overfitting, automation bias, and privacy and security concerns, which may manifest differently at the individual case and population levels and require tailored governance approaches [13,18,21-23].

In response to these emerging applications and challenges, international and national organizations, including the Council for International Organizations of Medical Sciences, the European Medicines Agency, and the US Food and Drug Administration, have begun to articulate principles for the use of AI across the medicinal product life cycle [20,23,24]. These initiatives emphasize data quality, model validation, explainability, human oversight, and governance throughout the AI life cycle.

Despite these advancements, several knowledge gaps remain. First, existing guidance documents are largely high level and cross-cutting and rarely specify, in a concrete or systematic manner, the types of information required, the data quality thresholds that should be met, or the risk management approaches needed in this context. Second, although the literature on AI in pharmacovigilance is expanding, most reviews provide broad overviews of applications and do not focus specifically on causality assessment, nor do they systematically distinguish between individual case- and population-level applications. Third, there is limited synthesis regarding which AI techniques and related automation tools are used at specific stages of causality assessment, the information and data quality assumptions on which they rely, and how associated risks are identified and addressed. Finally, the interplay and potential complementarity between AI-supported processes at the individual case and population levels remain insufficiently described.

Given this combination of emerging applications and persistent knowledge gaps, a scoping review is an appropriate approach for mapping the existing evidence, clarifying key concepts, and identifying areas requiring further research. Rather than quantifying the effect of a specific intervention, this review will provide an overview of how AI-based methods and, where applicable, AI-enabled automation are currently used, proposed, or discussed for causality assessment in pharmacovigilance. It will also summarize reported information inputs, data quality considerations, and associated risks. A preliminary search of MEDLINE (via PubMed), the Cochrane Database of Systematic

Reviews, and *JBIM Evidence Synthesis* identified no current or ongoing systematic or scoping reviews on this topic.

The main objectives of this study are described in [Textbox 1](#).

Textbox 1. Objectives of this scoping review.

Primary objective • Characterize how AI-based methods are applied or proposed for causality assessment between medicinal products and adverse events in pharmacovigilance, with a focus on their functional roles, reported data inputs, information needs, and associated risks Secondary objectives • Compare how these AI-based approaches differ between individual case-level and population-level applications, including potential complementarities between the 2 levels • Describe the types of AI-based techniques and supporting automation tools used in causality assessment workflows • Summarize reported data quality considerations and governance mechanisms (eg, human oversight, explainability, monitoring, and model maintenance) that support the safe use of AI in this context

Methods

Overview

A scoping review methodology was selected to map concepts, approaches, and evidence within a heterogeneous and rapidly evolving field rather than to estimate effect sizes or conduct a meta-analysis. This protocol follows guidance from the *JBIManual for Evidence Synthesis* [25] and will be reported, as applicable, in accordance with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [26] and PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for reporting literature searches in systematic reviews) [27]. The PRISMA-ScR and PRISMA-S checklists are provided in [Multimedia Appendices 1](#) and [2](#), respectively. The protocol was prospectively registered in the Open Science Framework [28].

Operational Definitions

Causality Assessment

In this review, causality assessment is defined as the plausibility evaluation of a causal relationship between a drug and an observed adverse event through the structured, qualitative synthesis of all available evidence streams [2]. This definition encompasses both individual case-level assessment, in which evidence from a specific report or patient is evaluated, and population-level assessment, in which evidence from aggregated safety data, real-world data, or epidemiological analyses is used to evaluate potential causal associations or safety signals.

Population-level causality assessment refers specifically to the evaluation of causal plausibility at the population level. This includes the explicit causal interpretation of detected safety signals, such as signal evaluation or validation, confounding assessment, causal modeling, or structured benefit-risk reasoning. In contrast, routine signal detection based solely on frequency-based disproportionality measures without causal interpretation is not considered population-level causality assessment and is excluded.

AI-Based Methods

AI-based methods are operationally defined as computational approaches that use data-driven learning, natural language processing, knowledge representation, causal modeling, or explicit knowledge- or rule-based reasoning to support or perform causality assessment. These include but are not limited to the following: (1) data-driven approaches such as machine learning models, deep learning models, natural language processing models, knowledge graph-based approaches, and causal inference models; (2) knowledge- or rule-based systems, including expert systems, that explicitly implement causal assessment logic, structured causality criteria, or rules for evaluating drug-event relationships; and (3) hybrid systems that combine statistical, machine learning, natural language processing, and knowledge- or rule-based components.

AI-Based Causality Assessment

AI applications relevant to causality assessment can be classified into 4 categories, as shown in [Textbox 2](#). The applications covered in this review fall into categories 1 to 3, which are distinguished from upstream support functions and pure business automation (category 4).

Textbox 2. AI application categories.**Category 1: AI used for extracting or structuring information relevant to causality assessment**

This category is limited to tools that extract, structure, or synthesize evidence elements explicitly relevant to causal review, such as temporal relationships between exposure and event, de- or rechallenge information, dose-response information, alternative explanations, concomitant medications, comorbidities, relevant laboratory or clinical findings, and prior evidence. General information extraction, coding, or data processing will not be considered sufficient for inclusion unless the extracted information is explicitly used or intended to be used to support causal judgment or structured causality assessment.

Category 2: AI used to assign or predict individual case-level causality or relatedness categories

This category includes systems that support or automate the application of structured causality criteria or assign relatedness categories for observed adverse events at the individual case level. AI used exclusively for adverse event forecasting (ie, estimating the likelihood of adverse events occurring in patients who have not yet experienced them) is outside the scope of this review.

Category 3: AI used for population-level causal inference or signal evaluation

This category includes methods that explicitly address causal interpretation, confounding, counterfactual reasoning, causal modeling, signal evaluation, or totality-of-evidence assessment in relation to a potential drug-event association. The use of frequency-based disproportionality measures alone or the term “signal detection” alone will not be sufficient for inclusion unless the method is explicitly linked to causal interpretation, confounding assessment, signal evaluation, causal inference, or totality-of-evidence assessment.

Category 4: AI or automation used solely for operational efficiency

This category includes tools used for case intake, routing, duplicate detection, Medical Dictionary for Regulatory Activities coding, data transfer, report generation, workflow management, or administrative prioritization when they do not incorporate any causal reasoning component or explicit linkage to causal evidence synthesis.

Causality assessment algorithms such as the World Health Organization Uppsala Monitoring Centre system or the Naranjo algorithm will not be considered AI-based methods when they are used only as manually applied criteria or as simple automated score calculators. They will be considered within scope only when AI- or knowledge-based methods automate the interpretation or application of the criteria, for example, by extracting relevant information from case narratives or other evidence sources and applying the criteria to support a causality judgment.

Simple statistical models such as logistic regression will be included only when explicitly implemented as a causal model or as part of a rule-based system for causality assessment. Models used solely for outcome prediction, prioritization, or classification without explicit causal framing will be excluded. Similarly, conventional descriptive statistics, disproportionality analyses used only as frequency-based screening without causal interpretation, and noncausal workflow automation will not be considered AI-based causality assessment by themselves.

For methods that fall near the eligibility boundary, inclusion will be determined by whether they are explicitly used to support a causal judgment regarding the relationship between a medicinal product and an adverse event.

Eligibility Criteria

The review will address the population, concept, and context framework recommended by the JBI for scoping reviews [25].

The central concept of interest is the use of AI-based methods to support or perform causality assessment for relationships

between medicinal products for human use, including vaccines and biologics, and adverse events at either the individual case or population level. Medical devices, as well as device components of drug-device combination products, are excluded; studies concerning the medicinal component of such products remain eligible. At the population level, causality assessment is distinguished from routine signal detection: it encompasses approaches that address confounding or bias, apply causal inference methods, or provide explicit causal interpretation of a signal.

Sources that describe or evaluate systems, methods, or workflows in which AI-based methods directly contribute to causality assessment or its interpretation in pharmacovigilance; propose frameworks, requirements, or governance models specifically for AI-based causality assessment in pharmacovigilance; or both will be included.

The context of interest is pharmacovigilance and medicinal product safety for human use, encompassing marketing authorization holders and pharmaceutical companies, regulatory agencies and public health authorities, health care settings (eg, hospitals and clinics), and academic or collaborative research networks conducting pharmacovigilance or pharmacoepidemiological studies. No restrictions will be placed on country, health system, or regulatory framework.

Nonempirical publications, including opinion pieces, editorials, and narrative commentaries that do not include substantive descriptions of AI-based methods or their application to causality assessment will be excluded. The complete eligibility criteria are summarized in [Table 1](#).

Table 1. Eligibility criteria (population, concept, and context framework).

Category	Inclusion criteria	Exclusion criteria
Population and data sources	Human pharmacovigilance or drug safety–related data, including individual case safety reports and spontaneous reports; electronic health records and hospital information systems; disease or drug registries; pharmacoepidemiological datasets; clinical trial safety datasets; published literature for safety assessment; and social media, patient forums, and direct patient reports used for pharmacovigilance purposes	Studies focusing exclusively on veterinary pharmacovigilance and sources not involving medicinal product safety, adverse events, pharmacovigilance, pharmacoepidemiology, or drug safety–related evidence
Concept	AI-based methods directly contributing to causality assessment, including (1) data-driven approaches (machine learning, deep learning, natural language processing, knowledge graphs, and causal inference models) and (2) knowledge- or rule-based systems explicitly implementing causal assessment logic. Both individual case- and population-level causality assessment are eligible. Eligible applications include AI used to extract or structure causality-relevant evidence elements, assign or predict case-level causality or relatedness categories, or support population-level causal inference or signal evaluation with explicit causal interpretation. Frameworks, requirements, or governance models for AI-based causality assessment in pharmacovigilance are also eligible.	AI applied only to pharmacovigilance tasks unrelated to causality assessment with no explicit causal link, including general information extraction, MedDRA ^a coding, duplicate detection, case routing, or workflow automation without extraction or structuring of causality-relevant evidence elements; automation-only tools without an AI component; simple automated scoring of causality algorithms without AI-enabled evidence interpretation or application; statistical or prediction models without explicit causal framing; frequency-based disproportionality analyses or “signal detection” with no explicit causal interpretation, confounding assessment, signal evaluation, causal inference, or totality-of-evidence assessment; and AI used solely for purposes unrelated to drug safety or pharmacovigilance
Context	Human pharmacovigilance and medicinal product safety settings, including marketing authorization holders; pharmaceutical companies; regulatory agencies; public health authorities; health care settings; and academic or collaborative research networks conducting pharmacovigilance, pharmacoepidemiological, or drug safety research, with no restrictions based on country, health system, or regulatory framework	Veterinary pharmacovigilance; AI applications used solely outside pharmacovigilance or medicinal product safety, such as marketing, adherence prediction, general clinical decision support unrelated to safety, or nonsafety administrative processes; and medical devices (including device-only vigilance) and combination products with respect to their device component
Types of sources	Quantitative, qualitative, and mixed methods empirical studies; methodological or technical papers; evaluation, feasibility, and pilot implementation studies; case studies; regulatory or industry reports with sufficient methodological details; and theses and conference proceedings with sufficient methodological details	Opinion pieces, editorials, or narrative commentaries without substantive descriptions of AI-based methods, requirements, or governance
Language and time frame	English or Japanese; no restriction based on publication year	All languages other than English or Japanese

^aMedDRA: Medical Dictionary for Regulatory Activities.

Information Sources

Searches will be conducted in the following bibliographic databases: (1) PubMed (MEDLINE), (2) Web of Science Core Collection (Science Citation Index Expanded, Social Sciences Citation Index, Arts and Humanities Citation Index, Emerging Sources Citation Index, Book Citation Index, Conference Proceedings Citation Index, Current Chemical Reactions, and Index Chemicus), (3) ProQuest, (4) EBSCOhost platform (Academic Search Complete, Bibliography of Asian Studies, Business Source Complete, CINAHL Plus with Full Text, Communication Abstracts, EconLit, Education Resources Information Center, GreenFILE, Historical Abstracts, Index to Legal Periodicals & Books, Information Science & Technology Abstracts, Linguistics Abstracts Online, MEDLINE, MLA Directory of Periodicals, MLA International Bibliography, Peace Research Abstracts, Philosopher's Index, PsycArticles, PsycInfo, Race Relations Abstracts, Regional Business News, SocINDEX with Full Text, Teacher Reference Center, and American Antiquarian Society Historical Periodicals Collection: Series 1-5), and (5) Ichushi Web (Japanese medical literature database).

To identify additional sources not captured through database searches, the following supplementary strategies will be used. First, the reference lists of all included articles and relevant reviews or guidelines will be screened. Second, forward citation tracking may be conducted using Web of Science or Google Scholar where appropriate. Third, targeted searches of the websites of the following regulatory and organizational bodies will be conducted: the US Food and Drug Administration, the European Medicines Agency, the Council for International Organizations of Medical Sciences, the World Health Organization, the Uppsala Monitoring Centre, and the Pharmaceuticals and Medical Devices Agency. For each website, searches will use terms related to AI, pharmacovigilance, and causality assessment, focusing on guidance documents, reflection papers, and technical reports. The date of each search will be recorded.

The database searches will be limited to English and Japanese. Finally, no study design filter will be applied.

Search Strategy

A 3-step search strategy as recommended by the JBI [25] will be used. The search strategy was developed a priori in

consultation with an information specialist and will be reported in accordance with applicable elements of the PRISMA-S [27] adapted for the scoping review context in 3 steps.

First, the development of the final strategy (initial limited search): an initial limited search of PubMed and Web of Science was conducted using a small set of key terms related to AI, pharmacovigilance, and causality assessment. Titles, abstracts, and indexing terms (eg, MeSH terms and keywords) of relevant articles were examined to identify additional terms.

Second, the comprehensive database search: a full search strategy was developed for PubMed and adapted for each

database, combining controlled vocabulary and free-text terms covering 3 main concept blocks (Table 2). All 3 blocks were combined using the Boolean operator AND; terms within each block were combined using OR.

Third, citation tracking and supplementary searches: the reference lists of all the included studies and key reviews or guidelines will be screened for additional relevant sources. Where appropriate, forward citation tracking may be conducted using Web of Science or Google Scholar. Targeted searches of relevant websites and repositories (as prespecified in the Information Sources section) will also be performed.

Table 2. Search concept blocks and representative terms (all 3 blocks will be combined using “AND”)^a.

Block	Concept domain	MeSH terms or controlled vocabulary terms (PubMed)	Free-text keywords
1	Pharmacovigilance	“Pharmacovigilance” (MeSH), “Drug-Related Side Effects and Adverse Reactions” (MeSH), “Adverse Drug Reaction Reporting Systems” (MeSH), and “Product Surveillance, Postmarketing” (MeSH)	“Pharmacovigilance,” “drug safety,” “adverse drug reaction*,” “adverse drug event*,” “adverse reaction,*” “drug-related adverse event*,” “ADR,” “adverse event*,” “safety signal*,” “safety report*,” “spontaneous report*,” and “ICSR*”
2	Causality and causal assessment	“Causality” (MeSH), “Bayes Theorem” (MeSH), and “Models, Statistical” (MeSH)	“Causality assessment,” “causal inference,” “causal model*,” “causal relationship*,” “signal assessment,” “Naranjo,” and “WHO-UMC”
3	AI	“Artificial Intelligence” (MeSH), “Machine Learning” (MeSH), “Deep Learning” (MeSH), “Natural Language Processing” (MeSH), and “Expert Systems” (MeSH)	“Artificial intelligence,” “machine learning,” “deep learning,” “natural language processing,” “NLP,” “knowledge graph*,” “expert system*,” “prediction model*,” “generative AI,” “agentic AI,” “large language model*,” “LLM*,” “ChatGPT,” “Claude,” “Gemini,” “GPT-5*,” “GPT-4*,” and “GPT-3*”

^aThe PubMed database search strategy uses MeSH terms combined with free-text keywords in the title and abstract field. For the other information sources, the strategy was translated into database-appropriate field codes and, where available, controlled vocabulary terms while maintaining the same conceptual structure. Complete search strings for all 5 databases are provided in [Multimedia Appendix 3](#).

Handling of Japanese- and English-Language Records

As all members of the review team are fluent in both Japanese and English, records in both languages will be screened and charted in their original language. For Japanese-language records retrieved from Ichushi Web and other sources, title and abstract and full-text screening will be performed directly in Japanese by at least 2 reviewers. Data charting for Japanese sources will likewise be conducted based on the original Japanese texts, with key data items (eg, study design, data sources, AI methods, outcomes, and main conclusions) entered into the shared data charting form in English to facilitate integrated synthesis across languages. Where important concepts or terminology do not have a straightforward English equivalent, the original Japanese wording will be recorded in the data charting form alongside an English summary. Any uncertainties or disagreements regarding language interpretation will be resolved through discussion within the review team using the same multistep consensus process as that for other reviewer disagreements.

Study and Source Selection

Overview

All search results will be exported to a reference manager (eg, EndNote; Clarivate Analytics), deduplicated, and then imported into a screening tool (eg, Rayyan; Rayyan Systems Inc) for

study selection. When multiple publications describe the same AI system or implementation, they will be grouped as a single system unit during data charting. The most recent and most methodologically comprehensive publication will serve as the primary reference, with others consulted as supplementary sources. Descriptive statistics will report both the number of distinct systems identified and the total number of publications included.

Study selection will proceed in 2 stages as follows. First, for title and abstract screening, 2 reviewers will independently screen titles and abstracts against the eligibility criteria. Records clearly not meeting the inclusion criteria will be excluded. Records retained by either reviewer will proceed to the full-text screening stage. Second, for full-text screening, the full texts of potentially relevant records will be retrieved. Two reviewers will independently assess each full text against the inclusion and exclusion criteria. Reasons for exclusion at this stage will be recorded (eg, not a pharmacovigilance context, AI not used for causality assessment, or insufficient methodological details).

Calibration and Agreement

Before full screening, the reviewers will pilot the eligibility criteria on a sample of approximately 20 to 30 records to ensure shared understanding and refine operational definitions if needed.



Disagreement Resolution

At both the title and abstract and full-text stages, disagreements between the 2 reviewers will first be flagged within the screening tool. A prespecified third reviewer will then independently assess the record. If consensus is not reached after the third reviewer’s assessment, disagreements will be resolved through discussion within the review team and, where necessary, by involving additional members. Study selection will be documented using a PRISMA-ScR flow diagram [26].

Data Charting (Data Extraction)

A structured data charting form will be developed in NVivo (Lumivero) or Microsoft Excel informed by the review objectives and the *JBIManual for Evidence Synthesis* [25]. This form will be piloted on a small number of included sources prior to full implementation and refined as necessary. The planned data charting items are summarized in Table 3.

Table 3. Planned data charting items.

Category	Data items
Bibliographic and general information	Authors, year of publication, and country or region; journal or source type; and funding sources and author-declared conflicts of interest
Study characteristics	Study design (eg, methodological paper, retrospective cohort study, cross-sectional study, qualitative study, mixed methods study, case study, or pilot implementation) and setting and organization type
Population and data sources	Types of safety data used (eg, individual case safety reports, electronic health records, registries, claims data, clinical trial data, published literature, and social media), key patient population or therapeutic area characteristics, and sample size or data volume
Causality assessment level and role	Levels targeted: individual case level, population level, or both and position of the AI-based method within the workflow (eg, information extraction and structuring, application of causality criteria, signal detection and causal interpretation, or governance and reporting)
AI techniques and implementation details	AI methods used (eg, supervised, unsupervised, or reinforcement learning; deep learning; natural language processing; knowledge graphs; rule-based systems; causal models; and hybrid systems), any automation technologies and their role, inputs and features, training and validation data, implementation maturity level (conceptual proposal, prototype or proof of concept, retrospective or prospective validation, pilot implementation, routine deployment, or regulatory or organizational framework), and source type
Information requirements for causality assessment	Types and granularity of the information required at the individual case level (eg, exposure and event timing, de- or rechallenge information, concomitant medications, comorbidities, and laboratory results), information requirements at the population level (eg, exposure populations, comparator groups, follow-up time, confounders, and health care use patterns), and any explicitly stated minimal datasets or mandatory data elements
Data quality and performance metrics	Data quality dimensions (completeness, accuracy, timeliness, consistency, and representativeness) and preprocessing steps; model performance metrics (eg, sensitivity, specificity, PPV ^a , NPV ^b , AUROC ^c , F ₁ -score, and calibration); validation strategies (internal, cross-validation, temporal, external, and comparison with expert assessment); and fairness, generalizability, or robustness evaluation
Governance and oversight	Human-in-the-loop design (thresholds for manual review, override mechanisms, and escalation pathways); approaches to explainability and transparency; and monitoring, maintenance, and updating (drift detection, retraining schedules, and performance auditing)
Risks and their management	Types of risks discussed (eg, algorithmic bias, model drift, overfitting, misclassification, automation bias, lack of explainability, and accountability and liability) and strategies for risk mitigation
Interaction between individual case- and population-level applications	How outputs from individual-level tools may inform population-level signal detection or causal modeling and vice versa, mechanisms for cross-validation, and shared infrastructure or data pipelines
Main findings and authors’ conclusions	Reported benefits, limitations, and recommendations and identified knowledge gaps and future research directions

^aPPV: positive predictive value.
^bNPV: negative predictive value.
^cAUROC: area under the receiver operating characteristic curve.

Two reviewers will independently chart data from each included source. One reviewer may perform the initial charting, and the second reviewer will verify the entries. Disagreements will be handled using the same 3-reviewer and consensus process used for study selection. The data charting form will remain flexible, and new fields may be added if important themes emerge during the review. Modifications will be reported in the final review.

Critical Appraisal of Individual Sources of Evidence

In accordance with JBI guidance for scoping reviews, a formal risk-of-bias assessment is not required when the objective is to map existing evidence and concepts [25]. However, given this review’s focus on data quality, validation, and governance, methodological characteristics relevant to interpreting AI-based methods will be documented in a structured manner during data charting. For each included AI-based or prediction model study,

the following will be recorded: (1) whether training data provenance is reported, (2) whether the reference standard or labeling process is described, (3) whether subgroup performance is evaluated, (4) whether data leakage controls are described, and (5) whether transportability or generalizability is addressed. These items are informed by the TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis for AI studies) reporting guidelines [29] and the PROBAST+AI (Prediction Model Risk-of-Bias Assessment Tool for AI studies) appraisal tool [30], used here as organizing frameworks rather than formal scoring instruments. These methodological characteristics will be reported descriptively in the completed review.

Data Synthesis and Presentation

Thematic synthesis will follow a reflexive thematic analysis approach [31] using both inductive and deductive strategies. An initial deductive coding framework will be derived from the review's population, concept, and context framework elements; primary and secondary objectives; and planned data items. Additional themes will be identified inductively. Findings will be presented using a combination of tables, figures (eg, workflow diagrams), and narrative synthesis. Descriptive summaries will be stratified by source type and implementation maturity level to avoid conflating conceptual proposals, prototype systems, validation studies, pilot implementations, and routine deployments. Analysis will be conducted using NVivo, structured Microsoft Excel matrices, or both.

Each analytic product corresponds to a specific research objective: descriptive summaries and thematic findings on requirements and risks address primary objective 1, cross-level comparisons address secondary objective 1, AI technique characterization addresses secondary objective 2, and governance synthesis addresses secondary objective 3.

Patient and Public Involvement

No patients or members of the public are involved in the design, conduct, reporting, or dissemination of this scoping review as it relies solely on previously published and publicly available data and documents.

Results

This protocol was registered in the Open Science Framework on December 23, 2025 (osf.io QVF5C) [28]. The registration was subsequently updated on May 19, 2026, to reflect an extension to the data collection period. A preliminary database search was conducted in December 2025, retrieving a total of 760 records, of which the preliminary title and abstract screening

identified 196 (25.8%) articles for full-text review. The formal search based on this protocol will be conducted in July 2026. Data charting is scheduled for August 2026, and synthesis is scheduled for September 2026. The completed review is expected to be submitted for publication by the end of December 2026. Deviations from this protocol will be documented and reported in the final review. This research was funded by the Ministry of Health, Labor and Welfare from April 2025 to March 2026 and by the Japan Society for the Promotion of Science from April 2026 to the present.

Discussion

Expected Findings

This review is expected to offer several contributions. First, it will provide a structured map of AI-based applications for causality assessment at the individual case and population levels, identifying areas of concentration and evidence gaps. Second, mapping reported information inputs, data quality dimensions, and validation practices is expected to support pharmacovigilance departments and marketing authorization holders when designing or procuring AI-based tools. Third, synthesizing reported risks and mitigation strategies may assist regulators and guideline developers in refining governance frameworks and help AI developers align system designs with pharmacovigilance needs.

Despite these contributions, several limitations are expected. For example, restricting inclusion to English- and Japanese-language sources introduces a potential source of selection bias; relevant work—particularly implementation descriptions and regulatory guidance documents from non-English- or Japanese-speaking regions—may be underrepresented. The broad scope of AI-based methods and the heterogeneous nature of implementations may constrain the ability to compare performance quantitatively. Additionally, the review will be limited by what is documented in the published literature; practical implementation details and internal organizational practices may not be captured.

Conclusions

This scoping review protocol sets out a systematic and transparent approach to mapping how AI-based methods are used, proposed, and discussed for causality assessment in pharmacovigilance across the individual case and population levels. By consolidating evidence that is currently dispersed across methodological, implementation, and regulatory sources, the planned review is intended to clarify the current landscape and establish a foundation for future empirical and methodological research on AI-supported causality assessment.

Acknowledgments

The authors would like to thank Professor Daisuke Koide (the University of Tokyo) and Professor Shinichiroh Yokota (Chiba University) for their helpful advice on developing the search strategy and for their constructive comments on the search strings for this scoping review protocol. They would also like to thank Editage for English-language editing. The views expressed in this paper are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the Pharmaceuticals and Medical Devices Agency. The authors declare the use of generative AI (GenAI) in the research and writing process. According to the Generative AI Delegation Taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: evaluation of the research novelty and identification of knowledge gaps, translation, and

recommendations. The GenAI tool used was GPT-5.1 Pro (OpenAI) and DeepL Pro (DeepL SE). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Data Availability

Data sharing is not applicable to this article as no data sets were generated or analyzed during this study.

Funding

This work is supported by the Ministry of Health, Labor, and Welfare of Japan (grant 23KC1005) and the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (grant 26K09796). The funders had no role in the study design, data collection, data analysis, decision to publish, or manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR checklist.

[[PDF File \(Adobe PDF File\), 517 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for reporting literature searches in systematic reviews) checklist.

[[PDF File \(Adobe PDF File\), 221 KB-Multimedia Appendix 2](#)]

Multimedia Appendix 3

Search strategies.

[[DOCX File , 24 KB-Multimedia Appendix 3](#)]

References

1. What is pharmacovigilance? World Health Organization. URL: <https://www.who.int/teams/regulation-prequalification/regulation-and-safety/pharmacovigilance> [accessed 2026-05-10]
2. Hammad TA, Afsar S, McAvoy LB, Le Louet H. Aspects to consider in causality assessment of safety signals: broadening the thought process. *Front Drug Saf Regul*. May 25, 2023;3:1193413. [[FREE Full text](#)] [doi: [10.3389/fdsfr.2023.1193413](https://doi.org/10.3389/fdsfr.2023.1193413)] [Medline: [40980101](https://pubmed.ncbi.nlm.nih.gov/40980101/)]
3. The use of the WHO-UMC system for standardised case causality assessment. World Health Organization. Jun 05, 2013. URL: <https://www.who.int/publications/m/item/WHO-causality-assessment> [accessed 2026-05-10]
4. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther*. Aug 1981;30(2):239-245. [doi: [10.1038/clpt.1981.154](https://doi.org/10.1038/clpt.1981.154)] [Medline: [7249508](https://pubmed.ncbi.nlm.nih.gov/7249508/)]
5. Practical Aspects of Signal Detection in Pharmacovigilance: Report of CIOMS Working Group VIII. Geneva, Switzerland. Council for International Organizations of Medical Sciences; 2010.
6. Martin GL, Jouganous J, Savidan R, Bellec A, Goehrs C, Benkebil M, et al. Validation of artificial intelligence to support the automatic coding of patient adverse drug reaction reports, using nationwide pharmacovigilance data. *Drug Saf*. May 2022;45(5):535-548. [[FREE Full text](#)] [doi: [10.1007/s40264-022-01153-8](https://doi.org/10.1007/s40264-022-01153-8)] [Medline: [35579816](https://pubmed.ncbi.nlm.nih.gov/35579816/)]
7. Barrett JW, Erlanson N, China JF, Norén GN. A scalable predictive modelling approach to identifying duplicate adverse event reports for drugs and vaccines. *arXiv*. Preprint posted online on March 31, 2025. [doi: [10.48550/arXiv.2504.03729](https://doi.org/10.48550/arXiv.2504.03729)]
8. Lieber T, Gosselt HR, Kools PC, Kruijssen OC, Van Lierop SN, Härmark L, et al. Natural language processing for automated triage and prioritization of individual case safety reports for case-by-case assessment. *Front Drug Saf Regul*. Feb 7, 2023;3:1120135. [[FREE Full text](#)] [doi: [10.3389/fdsfr.2023.1120135](https://doi.org/10.3389/fdsfr.2023.1120135)] [Medline: [40980103](https://pubmed.ncbi.nlm.nih.gov/40980103/)]
9. Tao Y, Li X, Yisha Z, Yang S, Zhan S, Sun F. LitAutoScreener: development and validation of an automated literature screening tool in evidence-based medicine driven by large language models. *Health Data Sci*. Sep 2, 2025;5:0322. [[FREE Full text](#)] [doi: [10.34133/hds.0322](https://doi.org/10.34133/hds.0322)] [Medline: [40904687](https://pubmed.ncbi.nlm.nih.gov/40904687/)]
10. Ghosh R, Kempf D, Pufko A, Barrios Martinez LF, Davis CM, Sethi S. Automation opportunities in pharmacovigilance: an industry survey. *Pharmaceut Med*. Feb 2020;34(1):7-18. [doi: [10.1007/s40290-019-00320-0](https://doi.org/10.1007/s40290-019-00320-0)] [Medline: [32036574](https://pubmed.ncbi.nlm.nih.gov/32036574/)]
11. Cherkas Y, Ide J, van Stekelenborg J. Leveraging machine learning to facilitate individual case causality assessment of adverse drug reactions. *Drug Saf*. May 2022;45(5):571-582. [doi: [10.1007/s40264-022-01163-6](https://doi.org/10.1007/s40264-022-01163-6)] [Medline: [35579819](https://pubmed.ncbi.nlm.nih.gov/35579819/)]
12. Zhang F, Sun B, Diao X, Zhao W, Shu T. Prediction of adverse drug reactions based on knowledge graph embedding. *BMC Med Inform Decis Mak*. Feb 04, 2021;21(1):38. [[FREE Full text](#)] [doi: [10.1186/s12911-021-01402-3](https://doi.org/10.1186/s12911-021-01402-3)] [Medline: [33541342](https://pubmed.ncbi.nlm.nih.gov/33541342/)]

13. Zhao Y, Yu Y, Wang H, Li Y, Deng Y, Jiang G, et al. Machine learning in causal inference: application in pharmacovigilance. *Drug Saf.* May 2022;45(5):459-476. [FREE Full text] [doi: [10.1007/s40264-022-01155-6](https://doi.org/10.1007/s40264-022-01155-6)] [Medline: [35579811](https://pubmed.ncbi.nlm.nih.gov/35579811/)]
14. Kreimeyer K, Dang O, Spiker J, Muñoz MA, Rosner G, Ball R, et al. Feature engineering and machine learning for causality assessment in pharmacovigilance: lessons learned from application to the FDA Adverse Event Reporting System. *Comput Biol Med.* Aug 2021;135:104517. [doi: [10.1016/j.compbimed.2021.104517](https://doi.org/10.1016/j.compbimed.2021.104517)] [Medline: [34130003](https://pubmed.ncbi.nlm.nih.gov/34130003/)]
15. Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf.* 2006;29(5):385-396. [doi: [10.2165/00002018-200629050-00003](https://doi.org/10.2165/00002018-200629050-00003)] [Medline: [16689555](https://pubmed.ncbi.nlm.nih.gov/16689555/)]
16. Matsuda S, Aoki K, Kawamata T, Kimotsuki T, Kobayashi T, Kuriki H, et al. Bias in spontaneous reporting of adverse drug reactions in Japan. *PLoS One.* May 1, 2015;10(5):e0126413. [FREE Full text] [doi: [10.1371/journal.pone.0126413](https://doi.org/10.1371/journal.pone.0126413)] [Medline: [25933226](https://pubmed.ncbi.nlm.nih.gov/25933226/)]
17. Moore N, Hall G, Sturkenboom M, Mann R, Lagnaoui R, Begaud B. Biases affecting the proportional reporting ratio (PPR) in spontaneous reports pharmacovigilance databases: the example of sertindole. *Pharmacoepidemiol Drug Saf.* Jun 2003;12(4):271-281. [doi: [10.1002/pds.848](https://doi.org/10.1002/pds.848)] [Medline: [12812006](https://pubmed.ncbi.nlm.nih.gov/12812006/)]
18. Ferreira-da-Silva R, Cruz-Correia R, Ribeiro I. Beyond black boxes: using explainable causal artificial intelligence to separate signal from noise in pharmacovigilance. *Int J Clin Pharm.* Apr 2026;48(2):677-681. [doi: [10.1007/s11096-025-02004-z](https://doi.org/10.1007/s11096-025-02004-z)] [Medline: [40889016](https://pubmed.ncbi.nlm.nih.gov/40889016/)]
19. Ball R, Talal AH, Dang O, Muñoz M, Markatou M. Trust but verify: lessons learned for the application of AI to case-based clinical decision-making from postmarketing drug safety assessment at the US Food and Drug Administration. *J Med Internet Res.* Jun 06, 2024;26:e50274. [FREE Full text] [doi: [10.2196/50274](https://doi.org/10.2196/50274)] [Medline: [38842929](https://pubmed.ncbi.nlm.nih.gov/38842929/)]
20. Artificial Intelligence in Pharmacovigilance. Geneva, Switzerland. Council for International Organizations of Medical Sciences; 2025.
21. Goddard K, Roudsari A, Wyatt JC. Automation bias - a hidden issue for clinical decision support system use. *Stud Health Technol Inform.* 2011;164:17-22. [Medline: [21335682](https://pubmed.ncbi.nlm.nih.gov/21335682/)]
22. Abdelwanis M, Alarafati HK, Tammam MM, Simsekler MC. Exploring the risks of automation bias in healthcare artificial intelligence applications: a Bowtie analysis. *J Saf Sci Resil.* Dec 2024;5(4):460-469. [doi: [10.1016/j.jnlssr.2024.06.001](https://doi.org/10.1016/j.jnlssr.2024.06.001)]
23. Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. European Medicines Agency. Sep 9, 2024. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf [accessed 2026-05-10]
24. Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products draft guidance for industry and other interested parties. U.S. Food & Drug Administration. Jan 2025. URL: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological> [accessed 2026-05-10]
25. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. *JBIManual for Evidence Synthesis*. North Adelaide, Australia. JBI Global; 2024.
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* Oct 02, 2018;169(7):467-473. [FREE Full text] [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
27. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev.* Jan 26, 2021;10(1):39. [FREE Full text] [doi: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z)] [Medline: [33499930](https://pubmed.ncbi.nlm.nih.gov/33499930/)]
28. Ohta M, Ota M, Ohta M. Artificial intelligence for causality assessment in pharmacovigilance: protocol for a scoping review. *Open Science Framework.* 2025. URL: <https://osf.io/qvf5c/overview> [accessed 2026-06-26]
29. Collins GS, Moons KG, 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.* Apr 16, 2024;385:e078378. [FREE Full text] [doi: [10.1136/bmj-2023-078378](https://doi.org/10.1136/bmj-2023-078378)] [Medline: [38626948](https://pubmed.ncbi.nlm.nih.gov/38626948/)]
30. Moons KG, Damen JA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ.* Mar 24, 2025;388:e082505. [doi: [10.1136/bmj-2024-082505](https://doi.org/10.1136/bmj-2024-082505)] [Medline: [40127903](https://pubmed.ncbi.nlm.nih.gov/40127903/)]
31. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* Jul 21, 2008;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]

Abbreviations

JBIM: Joanna Briggs Institute

PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for reporting literature searches in systematic reviews

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

PROBAST+AI: Prediction Model Risk-of-Bias Assessment Tool for AI studies

TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis for AI studies

Edited by A Schwartz; submitted 18.May.2026; peer-reviewed by T Hammad, S Khan; comments to author 09.Jun.2026; revised version received 26.Jun.2026; accepted 30.Jun.2026; published 16.Jul.2026

Please cite as:

Ohta M, Ota M, Ohta M

AI for Causality Assessment in Pharmacovigilance: Protocol for a Scoping Review

JMIR Res Protoc 2026;15:e101691

URL: <https://www.researchprotocols.org/2026/1/e101691>

doi: [10.2196/101691](https://doi.org/10.2196/101691)

PMID:

©Miki Ohta, Miki Ota, Mikihiro Ohta. Originally published in JMIR Research Protocols (<https://www.researchprotocols.org>), 16.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.researchprotocols.org>, as well as this copyright and license information must be included.