

Large language models for patient-centred evidence generation and use

Author preprint | Not peer reviewed | 19 September 2026

Chuan Yin, MD, PhD^{1,*}; Zhicheng Zhang, PhD²

1 Department of Orthopaedic Surgery, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

2 JancsiLab, JancsiTech, Hong Kong, China

Chuan Yin ORCID: [0009-0005-2830-8167](https://orcid.org/0009-0005-2830-8167)

Zhicheng Zhang ORCID: [0000-0002-5333-1394](https://orcid.org/0000-0002-5333-1394)

* Correspondence: Chuan Yin, Department of Orthopaedic Surgery, Shanghai Ninth People's Hospital, 639 Zhizaoju Road, Huangpu District, Shanghai 200011, China; yinchuan@sjtu.edu.cn

Author perspectives

Chuan Yin, the first and corresponding author, is a practising orthopaedic surgeon with 12 years of clinical experience. Zhicheng Zhang, the final author, is an AI scientist working in medical AI, particularly imaging diagnosis and prognostic evidence.

Main text: 2305 words | Abstract: 141 words

32 references | 1 figure

Main-text count excludes headings, citations, summary, figure legend, declarations, and references.

Abstract

For patients with several health conditions, evidence for one disease may leave important questions about care unanswered. Large language models (LLMs) offer ways to help formulate these questions, examine the literature, and prepare requests for research. We propose an evidence operating system (Evidence OS) that follows a patient's question from clinical goals through evidence synthesis and research to the return of findings. The framework applies across specialties, with multimorbidity illustrating the difficulties of judging what evidence means for an individual. LLM assistance would require source checks and professional review. Similarity between patients and fluent explanations cannot establish treatment benefit. Prospective evaluation should examine supported answers, clinical use, workload, patient experience, and health outcomes. Comparisons with and without LLM assistance are needed to distinguish model effects from better organisation of evidence services. Whether the proposed framework improves care remains to be tested.

The question in the consultation

For a person living with several health conditions, a recommendation for one disease may be difficult to reconcile with treatment for another, or with daily life.¹ Evidence-based medicine recognises that decisions depend on clinical expertise and individual circumstances as well as research.² Bringing these together in a consultation remains difficult. Patients and clinicians need to know which findings apply, where the evidence runs out, and what can be done about the uncertainty.

We propose a framework for following a patient's question through evidence synthesis and, when needed, research, with a team responsible for returning the findings to care (figure 1). Large language models (LLMs) could assist this work. Whether they make it more useful is an empirical question: experience with conversational medical AI shows that benchmark performance alone does not establish clinical usefulness or patient trust.³

Consider a hypothetical patient with a recent fragility fracture and stage G4 chronic kidney disease (CKD), who is not receiving dialysis. The patient's priorities are mobility and independent living. Choosing a fracture-prevention treatment requires consideration of kidney-related bone abnormalities, previous treatment, fracture risk, and the monitoring that can be arranged. A European consensus describes limited evidence for drug treatment in advanced CKD.⁴ Guidance on bone-turnover assessment identifies further uncertainties about treatment targets and strategies.⁵ A response that lists treatment options without addressing these constraints would leave much of the clinical question unanswered.

Multimorbidity makes this problem particularly clear, but the proposed framework extends across clinical specialties and questions about diagnosis, prognosis, prevention, treatment, rehabilitation, and monitoring. Participating services would share ways of framing questions, tracing sources, arranging research, and delivering updates. Each would still need the expertise and methods appropriate to its questions. This is a proposal for how services might work together, with clinical value yet to be established.

What language models might contribute

Learning from other patients is a longstanding ambition. PatientsLikeMe demonstrated online patient matching, and the green-button proposal connected clinical questions with aggregate records and research.^{6,7} PCORnet supports collaborative research and the return of results.⁸ Living systematic reviews address the need to keep syntheses current,⁹ while the CLEAR framework connects research priorities, study design, guidelines, evaluation, and shared decisions.¹⁰ Our proposal brings these activities around a question raised in care and assigns responsibility for following it through.

Much of that work involves moving between forms of information: a patient's account, a searchable question, a research report, a request for analysis, and an explanation that can be discussed in the consultation. LLMs may help with these transitions, including questions that do not fit a fixed template. The quality of the answer will nevertheless depend on the sources and analytical methods.

In the fracture example, a model might draft questions about prevention, treatment harms, kidney stage, and monitoring for the patient and clinician to check. Retrieval and extraction tools could then locate relevant passages and identify study populations and outcomes. The response would need to separate dialysis from non-dialysis evidence and establish whether studies measured mobility, independence, and treatment burden. Similar language in two reports is no assurance that their findings can be combined.

Early evaluations offer reasons to test this approach, although they do not establish its clinical benefit. In a small pilot comparison with unaided experts, LEADS reported faster screening and extraction, with higher aggregate extraction accuracy.¹¹ The otto-SR preprint examined screening, extraction, and risk-of-bias assessment, including effects on primary meta-analyses in 12 Cochrane reviews. Discrepancies underwent dual human review.¹² The relevant next question is how such workflows perform when the starting point is a decision in care.

Results also depend on what is being tested. A two-topic hepatology evaluation reported shorter screening times, but drafts produced with retrieval support contained numerical errors and unsupported results.¹³ Approximating the findings of a published meta-analysis differs from reconstructing its evidence base.¹⁴ Agreement with its conclusion, or with a reporting checklist, cannot replace source and analysis checks. Reports of performance need to retain excluded questions, failed tasks, and human corrections. Otherwise, difficult cases can disappear from the assessment.

Some systems extend to preparing analyses. THESEUS translated study descriptions into structured specifications and generated scripts using rules. Its evaluation addressed four components of those specifications, rather than the causal validity of resulting estimates.¹⁵ TRIALSCOPE combined language-model extraction from records with observational trial emulation in oncology, using data from one US healthcare network. Unmeasured confounding remained a limitation.¹⁶ These studies suggest ways to prepare requests for approved analyses and explain their results, subject to methodological review. Successful execution of code, even with an estimate resembling that of a trial, does not establish a valid causal comparison.

A model can lose a patient's goal in translation, misread a source, invent support for a claim, or propose an unsuitable analysis. Guidance from evidence-synthesis organisations therefore

emphasises responsible use and evaluation.¹⁷ Sources and missing information must be visible, and transformations that could alter a conclusion require professional review.

Deciding what applies to this patient

A useful request begins with the decision and its context: the alternatives, outcomes important to the patient, time horizon, and practical constraints. Function, disease severity, concurrent treatment, and available support can matter more than a shared diagnosis. Evidence-to-decision frameworks provide a way to consider benefits, harms, values, and feasibility together.¹⁸ Translating the account into research terms must preserve these features.

The answer then needs to distinguish direct evidence from extrapolation and from conclusions the evidence cannot support. For each estimate, the record should retain its study and source version, population, outcome definition, denominator, time point, and analytical transformations. Reports from overlapping cohorts cannot serve as independent confirmation. Nor will one synthesis necessarily answer the whole question: diagnosis requires suitable reference standards, prognosis requires validated predictions, and treatment effects require valid comparisons.

What counts as a similar patient depends on the inference being made. A factor that predicts prognosis need not predict a difference in treatment effect.¹⁹ Very close matching may leave no adequate comparator. Broader trial evidence can still be informative when the assumptions needed to apply it are explicit, taking account of disease stage, previous treatment, and care setting.

Improvement after treatment is also compatible with explanations other than a treatment effect. Observational comparisons require suitable comparators, measured confounders, sufficient overlap, and defensible assumptions. Target trial emulation helps specify the intended comparison, but cannot supply missing measurements or remove unmeasured confounding.²⁰ Evidence about separate treatments leaves further questions about their combination and sequence.

For example, an observational study of women aged at least 65 years receiving dialysis associated denosumab initiation with a substantially higher risk of severe hypocalcaemia than oral bisphosphonate initiation.²¹ Our hypothetical patient is not receiving dialysis, so those estimates cannot be transferred directly. They do not establish that the comparator is suitable for every patient either. The FDA emphasises assessment and monitoring when denosumab is considered in advanced CKD.²² An answer to the clinical question should explain these differences in population and identify any monitoring information still needed.

Even direct trial evidence can leave a difficult choice. In PRESTIGE-AF, direct oral anticoagulants reduced ischaemic stroke but increased recurrent intracerebral haemorrhage among survivors of intracerebral haemorrhage with atrial fibrillation. The haemorrhage safety endpoint did not meet the prespecified non-inferiority criterion.²³ Patients and clinicians must still judge applicability and weigh the outcomes against personal goals. These examples illustrate the work the framework might support, rather than define its clinical scope or demonstrate its effectiveness.

When uncertainty calls for research

Before proposing a study, the team needs to establish why the question remains unanswered. Relevant patients or outcomes may be missing from the literature, or the available comparison may not match the clinical choice. Sometimes the obstacle is missing clinical information, unavailable care, unmet monitoring needs, or an update that never reached the team. These require different responses. Another study will not resolve every uncertainty encountered in a consultation.

Where research is warranted, an LLM could help set out the question, population, outcomes, existing evidence, possible design, and unresolved assumptions. For fracture prevention, the first step might be to establish whether existing trials contain adequate individual participant data (IPD) on non-dialysis CKD G4, previous fractures, and relevant outcomes. Only then would a subgroup analysis be worth considering. The question and available data would guide prespecified reanalysis, prospective measurement, an observational comparison, or randomisation. Investigators would assess methods and feasibility before taking a generated proposal forward.

Authorised follow-up could record clinical assessments, treatment, timing, context, outcomes, and missingness. Model extraction might help prepare records, provided reviewers check their clinical meaning and temporal relationships. Inclusion in a study would depend on eligibility, compatible measurements, and dependence between records.²⁴ Suitable studies and updates could subsequently inform living reviews and, where appropriate, aggregate-data or IPD meta-analysis.

For the illustrative patient, outcomes include subsequent fractures, treatment harms, mobility, independence, death, and treatment burden at specified times. Follow-up observations may raise a safety concern without establishing its cause. Any resulting change to a treatment-effect claim requires appropriate analysis and review, whether the findings strengthen, weaken, or reverse an earlier conclusion.

Returning findings to care

Following a question over time requires a maintained record of patient goals, sources and their limits, the responsible team, the next task, and the answer returned. Maintenance deserves evaluation in its own right. In a prospective oncology-guideline case study, models anticipated only some subsequent substantive revisions despite strong performance on a separate extraction task.²⁵

When evidence changes, version and dependency records should identify which conclusions are affected and when the information became available, was checked, and was released. An LLM might help explain the implications for the original decision, after review. The clinical team would record receipt and discussion of the findings, including reasons for acting or not acting. Patients still need care while research proceeds. Answering from the evidence available today and arranging research to improve tomorrow's answer are distinct responsibilities.

Participation and responsibility

Patients need a role in defining the decisions, outcomes, and burdens worth studying, with assistance where necessary. Clinicians assess applicability, record care, identify errors, and help develop research questions. Research teams design analyses, data custodians oversee access and quality, and

synthesis teams maintain the evidence. Language models could assist with expressing questions, finding related work, and explaining responses across languages. Review must check for lost meaning and unequal performance.

Methods such as those of the James Lind Alliance can support priority setting with patients and clinicians.²⁶ Priorities should reflect the consequences of uncertainty, feasibility, equity, and the value of further information.²⁷ Online popularity offers a poor measure of importance when some groups cannot readily participate or their outcomes are seldom measured. Clinical and community routes would allow people with limited digital access to contribute.

Collaborators could share questions, protocols, data dictionaries, code, findings, and corrections linked to archived versions. GitHub is one possible venue, alongside accessible routes for patient participation. Public literature submissions must remain separate from clinical data accessed under controlled research agreements. Existing standards, including the Observational Medical Outcomes Partnership common data model²⁸ and the HL7 Fast Healthcare Interoperability Resources Evidence resource,²⁹ provide ways to represent information consistently. Model-generated mappings still need checks of clinical meaning.

Participation also creates an obligation to report back. Contributors should be told what happened to their questions and how their input was used, including null findings, harms, and corrections. LLMs might help prepare these accounts, but whether understandable feedback sustains participation needs testing. Funding must cover maintenance, professional review, and participation as well as new analyses. Changes to rules and algorithms require review, versioned release, and monitoring. Data stewardship would govern access, reuse, retention, and unnecessary copying while preserving the records needed to audit published claims.

How to test the proposal

Our related programme has produced a clinical evidence synthesis and an evaluation of evidence-production workflows. These unpublished studies inform the proposal. They do not show whether the complete service improves care or patient outcomes.

An initial study could follow consecutive questions of a defined type in one clinical service, using a preregistered protocol and independent review. Existing studies and authorised updates might suffice without creating a new patient registry. Evaluation would begin without influencing care, then progress to supervised feasibility work and appropriate comparative studies, with prespecified escalation and stopping criteria. Selection of the service would depend on expertise, data access, patient participation, and the capacity to return findings, rather than on the illustrative cases.

Two comparisons would help establish what the approach contributes: existing evidence services versus the proposed service, and the proposed service with versus without LLM assistance, using comparable sources and expertise. Component studies could examine question formulation, extraction, analytical requests, and explanations of updates. Model identifiers, prompts, permissions, versions, and human corrections need to be recorded so that better organisation or additional reviewer effort is not mistaken for a model effect. Clinical studies should use fixed versions or prespecified change procedures, documenting interruptions and switches.

Measures should capture whether patient goals were preserved, answers were supported, important gaps were missed, appropriate research was initiated, and findings were returned and used. Reports must include consequential unsupported claims and keep unanswered questions, failed tasks, and losses to follow-up in their denominators. A service that avoids most questions can have few errors yet offer little help.

Workload includes supervision, verification, rework, and follow-up, alongside patient burden and computing and storage resources. Corrected sources and new findings could test missed updates, delivery delays, dependency tracking, and explanation accuracy. Performance and participation burden need assessment across populations and languages. Each new specialty or question type brings sources, methods of inference, and clinical workflows that require evaluation.

Comparative studies should examine safety, understanding, care burden, attainment of patient goals, and health outcomes, together with trust over time and appropriate reliance on the service. Safety claims must specify supervision and exclusions of high-risk patients. DECIDE-AI and CONSORT-AI offer reporting guidance.^{30,31} The 3D multimorbidity trial found no improvement in its primary quality-of-life outcome.³² Patient-centred intentions are therefore insufficient grounds for expecting benefit. The test is whether patients and clinicians receive better-supported answers, whether useful research follows from unresolved questions, and whether returning the findings improves decisions and care.

Contributors

CY developed the concept and clinical framing, synthesised the literature, and prepared the manuscript. ZZ contributed the AI methodological perspective, drawing on expertise in medical imaging, clinical prediction, and prognostic evidence. CY is the first and corresponding author; ZZ is the final author. The authors are responsible for the scientific content.

Patient and public involvement

Eight people with experience as surgical patients, carers, or people awaiting surgery contributed perspectives through consultation during development of the framework. Their contributions informed its emphasis on patient-important outcomes, understandable uncertainty, source traceability, and further evidence needs.

Declaration of interests

CY is a founder of Beijing Surface Medical Technology Co., Ltd. and is developing an evidence-synthesis platform related to concepts discussed in this Viewpoint. ZZ is Chief Scientific Officer of JancsiTech/JancsiLab, Hong Kong, an AI research entity.

Funding

No specific funding was received for this work.

Data availability

This Viewpoint reports no new patient-level clinical analysis and generates no patient-level research dataset. The sources supporting its arguments are cited in the reference list. The clinical scenarios are hypothetical.

Acknowledgments

We thank the patient and public contributors whose perspectives informed the patient-centred framing of this work.

Declaration of generative AI and AI assisted technologies in manuscript preparation

OpenAI ChatGPT through the Codex interface was used to assist literature scoping and source checking, manuscript structure and language editing, conceptual figure preparation, document formatting, and consistency checks. The service was accessed on September 10–19, 2026; an exact model snapshot identifier was not available. Figure 1 was rendered programmatically using Python 3.12.14 and Matplotlib 3.10.8 and contains no patient images or empirical results. The authors retain responsibility for the accuracy, interpretation, originality, and final content of the manuscript.

References

1. Boyd CM, Darer J, Boult C, Fried LP, Boult L, Wu AW. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. *JAMA* 2005; 294: 716–24. doi:10.1001/jama.294.6.716
2. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71–72. doi:10.1136/bmj.312.7023.71
3. Schaekermann M, Palepu A, Rodman A, et al. Prospective evidence for conversational medical AI is hard, but non-negotiable. *Nat Med* 2026; published online Sept 14. doi:10.1038/s41591-026-04639-5
4. Evenepoel P, Cunningham J, Ferrari S, et al. European Consensus Statement on the diagnosis and management of osteoporosis in chronic kidney disease stages G4–G5D. *Nephrol Dial Transplant* 2021; 36: 42–59. doi:10.1093/ndt/gfaa192
5. Haarhaus M, Jørgensen HS, Bacchetta J, et al. Bone turnover markers in the management of CKD-associated osteoporosis—a European consensus. *Nephrol Dial Transplant* 2026; 41: 1357–75. doi:10.1093/ndt/gfag033
6. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotechnol* 2011; 29: 411–14. doi:10.1038/nbt.1837
7. Longhurst CA, Harrington RA, Shah NH. A 'green button' for using aggregate patient data at the point of care. *Health Aff (Millwood)* 2014; 33: 1229–35. doi:10.1377/hlthaff.2014.0099
8. PCORnet. PCORnet Playbook: getting started with PCORnet. Updated June 10, 2025. Accessed September 10, 2026. <https://pcornet.org/module-2-playbook/>
9. Elliott JH, Turner T, Clavisi O, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med* 2014; 11: e1001603. doi:10.1371/journal.pmed.1001603
10. Fang Y, Zhang G, Chen F, et al. CLEAR: A vision to support clinical evidence lifecycle with continuous learning. *J Biomed Inform* 2025; 169: 104884. doi:10.1016/j.jbi.2025.104884
11. Wang Z, Cao L, Jin Q, et al. A foundation model for human-AI collaboration in medical literature mining. *Nat Commun* 2025; 16: 8361. doi:10.1038/s41467-025-62058-5

12. Cao C, Arora R, Cento P, et al. Automation of Systematic Reviews with Large Language Models. medRxiv 2026; version 4, posted May 4. doi:10.1101/2025.06.13.25329541 (preprint). <https://www.medrxiv.org/content/10.1101/2025.06.13.25329541v4>
13. Zou Y, Kim I, Gao N, Li M, Kim MO, Ge J. Closing the screening gap but not the writing gap: a two-topic evaluation of LLMs for systematic reviews and meta-analyses in hepatology. NPJ Gut Liver 2026; 3: 21. doi:10.1038/s44355-026-00068-w
14. Pratte M, Thirukumar S, Zhang C, Slessarev M, Basmaji J, Prager R. Can large language models approximate the results of meta-analyses in critical care? A meta-research study. J Crit Care 2026; 92: 155358. doi:10.1016/j.jcrc.2025.155358
15. Kim H, Kim M, Kim S, You SC. From study design to executable code: automating target trial emulation with large language models. JAMIA Open 2026; 9: ooag131. doi:10.1093/jamiaopen/ooag131
16. González J, Ueno R, Wong C, et al. TRIALSCOPE — A Framework for Clinical Trial Simulation from Real-World Data. NEJM AI 2025; 2: AIoa2400859. doi:10.1056/AIoa2400859
17. Flemyng E, Noel-Storr A, Macura B, et al. Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell Collaboration, JBI and the Collaboration for Environmental Evidence 2025. Environ Evid 2025; 14: 20. doi:10.1186/s13750-025-00374-5
18. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ 2016; 353: i2016. doi:10.1136/bmj.i2016
19. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. Ann Intern Med 2020; 172: 35–45. doi:10.7326/M18-3667
20. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016; 183: 758–64. doi:10.1093/aje/kwv254
21. Bird ST, Smith ER, Gelperin K, et al. Severe hypocalcemia with denosumab among older female dialysis-dependent patients. JAMA 2024; 331: 491–99. doi:10.1001/jama.2023.28239. <https://jamanetwork.com/journals/jama/fullarticle/2814249>
22. US Food and Drug Administration. FDA adds Boxed Warning for increased risk of severe hypocalcemia in patients with advanced chronic kidney disease taking osteoporosis medicine Prolia (denosumab). Jan 19, 2024. (accessed Sept 10, 2026). <https://www.fda.gov/drugs/drug-safety-communications/fda-adds-boxed-warning-increased-risk-severe-hypocalcemia-patients-advanced-chronic-kidney-disease>
23. Veltkamp R, Korompoki E, Harvey KH, et al. Direct oral anticoagulants versus no anticoagulation for the prevention of stroke in survivors of intracerebral haemorrhage with atrial fibrillation (PRESTIGE-AF): a multicentre, open-label, randomised, phase 3 trial. Lancet 2025; 405: 927–36. doi:10.1016/S0140-6736(25)00333-2. <https://pubmed.ncbi.nlm.nih.gov/40023176/>
24. Tierney JF, Stewart LA, Clarke M. Chapter 26: Individual participant data. In: Higgins JPT, Thomas J, Chandler J, et al, eds. Cochrane Handbook for Systematic Reviews of Interventions, version 6.5. Cochrane, 2024. Chapter updated October 2019. Accessed September 10, 2026. <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-26>
25. Knauer M, Greß J, Kather JN, May P. Large Language Models for Oncology Guideline Maintenance: Prospective Case Study. JMIR AI 2026; 5: e93239. doi:10.2196/93239
26. James Lind Alliance. The James Lind Alliance Guidebook. Chapter 10: Translating priorities into research. Accessed September 10, 2026. <https://www.jla.nihr.ac.uk/jla-guidebook-translating-priorities-into-research>
27. Claxton K, Griffin S, Koffijberg H, McKenna C. How to estimate the health benefits of additional research and changing clinical practice. BMJ 2015; 351: h5987. doi:10.1136/bmj.h5987

28. Observational Health Data Sciences and Informatics. Data standardization: the OMOP Common Data Model. Accessed September 10, 2026. <https://www.ohdsi.org/data-standardization/>
29. HL7 International. FHIR Release 5, version 5.0.0. Resource Evidence. Accessed September 10, 2026. <https://hl7.org/fhir/R5/evidence.html>
30. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* 2022; 28: 924–33. doi:10.1038/s41591-022-01772-9
31. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020; 26: 1364–74. doi:10.1038/s41591-020-1034-x
32. Salisbury C, Man MS, Bower P, et al. Management of multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3D approach. *Lancet* 2018; 392: 41–50. doi:10.1016/S0140-6736(18)31308-4

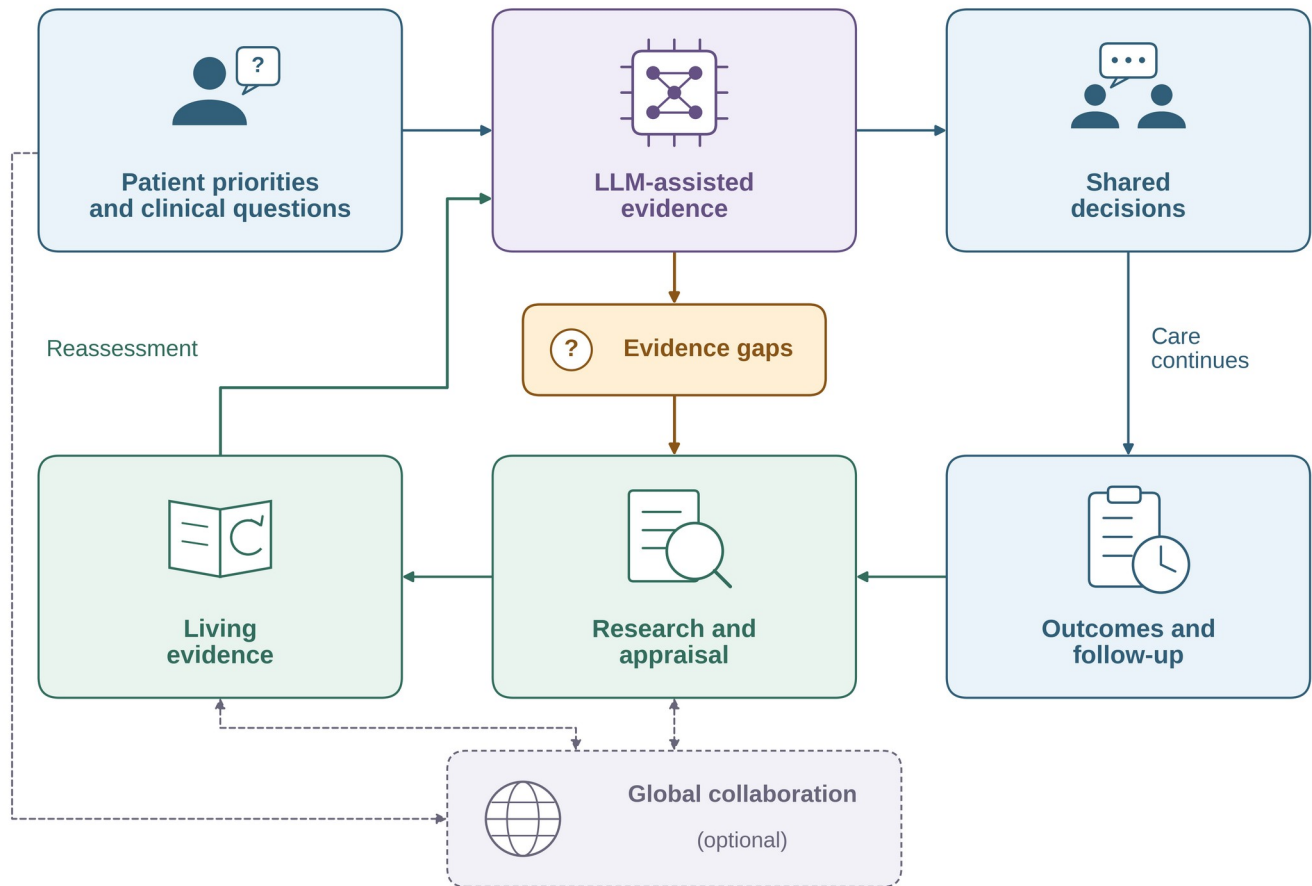


Figure 1 Following a patient question through evidence generation and use. Patient priorities and clinical questions enter a large language model (LLM)-assisted evidence workflow that supports shared decisions. Where research is warranted, unresolved questions can prompt studies or prespecified reanalysis. Eligible findings can update living syntheses and return for clinical reassessment while care continues. Blue regions show clinical processes, purple shows LLM assistance, green shows research and evidence synthesis, and amber highlights evidence gaps. The globe and dashed links indicate optional global collaboration. All icons are schematic. Arrows represent proposed information flows, not automatic clinical decisions or eligibility for research or pooling. Source and methods checks and professional review remain necessary. Research and data use require appropriate authorisation and eligibility checks. Public literature contributions remain separate from clinical data accessed under controlled research agreements. Clinical benefit, safety, patient experience, and workload require prospective evaluation.