



INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Claude for Rare Disease Drug Development: Evidence Gates

Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-10-20 · claude for rare disease drug development · claude rare disease research · variant interpretation · evidence synthesis · natural history studies · regulatory scientific writing · dismech · ind application drafting

Original article: <https://intuitionlabs.ai/articles/claude-rare-disease-drug-development>



In this article

1. Executive Summary
 2. Introduction and Background
 3. Key Changes
 4. Five Workflow Lanes and Their Review Gates
 5. Evidence-State Architecture
 6. Regulatory Expectations and Sparse-Data Controls
 7. Implementation Considerations and Process Changes
 8. Data Analysis and Evidence
 9. Implications and Future Directions
 10. Conclusion
-

Executive Summary

Claude for rare disease drug development is best approached as a supervised evidence-processing workflow: retrieve, extract, reconcile, and draft, while qualified specialists retain consequential decisions. Anthropic's **July 20, 2026** rare-genetic-disease call offered **up to \$50,000 in credits over six months** across basic-science and early-biotech tracks; its application deadline was **August 2, 2026**. (www.anthropic.com) These were proposed research workflows, not demonstrated development-time reductions. The **Life Sciences Verification Program (LSVP)** subsequently expanded an access pathway, but its **September 30, 2026** clarification describes the linked form as an expression of interest. (www.anthropic.com) Access is not evidence of scientific fitness.

The useful division is between assistance and authority. Claude can prepare variant-evidence summaries, literature maps, natural-history extraction sheets, modality comparisons, and regulatory drafts for review. A **variant of uncertain significance (VUS)** cannot become a clinical decision merely because a model gives it a compelling explanation: the American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG/AMP) guideline explicitly restricts that use. (pmc.ncbi.nlm.nih.gov) Likewise, the United States Food and Drug Administration (FDA) describes an investigational new drug (IND) application as requiring pharmacology and toxicology, manufacturing, and clinical-investigation information; fluent writing cannot supply missing experiments. (www.fda.gov)

DisMech offers a concrete engineering pattern: structured records, ontology checks, reference-snippet validation, and a human pull-request gate. Its project documentation explicitly warns that these checks do not guarantee scientific correctness. (github.com) Its newer evidence model goes further, distinguishing citation grounding from evidence appraisal. (github.com) Accordingly, every material assertion should carry its source passage, identifier, ontology version, data cutoff, transformation, uncertainty, reviewer, disposition, and permitted downstream use. This proposed architecture extends the project's technical checks into a program-specific decision process.

The regulatory distinction is equally important. FDA's rare-disease development guidance is **final, December 2023**, while its dedicated natural-history guidance remains labeled **draft, March 2019**, on the retrieved pages. (www.fda.gov) Its **January 2025** artificial intelligence (AI) credibility guidance is also labeled draft and frames

credibility around context of use. (www.fda.gov) A **60 to 90 day pilot**, proposed here rather than prescribed by a regulator, should therefore start with retrospective, public or synthetic material. Advancement should depend on prospectively justified evidence coverage, omission detection, expert agreement, and review effort, not writing speed alone. International Council for Harmonisation (ICH) E8(R1) supplies the underlying principle: study quality means fitness for purpose. (database.ich.org)

Introduction and Background

Rare-disease program teams must connect heterogeneous evidence before they can make development decisions. A disease name may denote a broad group in one resource and a subtype in another. **Mondo** exists to harmonize disease definitions, but its documentation warns that mappings across resources are not necessarily one-to-one. (mondo.monarchinitiative.org) **Orphanet** similarly distinguishes groups of disorders, disorders, and subtypes. A successful search or apparent label match therefore does not establish that two cohorts, mechanisms, or development programs concern the same biological entity. (www.orpha.net)

Claude for rare disease research becomes useful when its output is a reviewable intermediate artifact. A scientist should be able to move from a proposed mechanism to the exact supporting experiment, inspect the population and model system, and identify the inference that connects the observation to the development decision. DisMech's evidence-model documentation makes this distinction explicit: a validated literature pointer is useful, but it is not a complete appraisal of experimental strength or mechanistic inference. (github.com)

This report takes the perspective of an **adjacent life-sciences and AI consultancy**: IntuitionLabs is the publisher, not a competing model or research platform. The practical question is which work products a team can inspect and approve, and where genetics, clinical, biostatistical, chemistry, manufacturing and controls (**CMC**), or regulatory expertise must determine the outcome. The Clinical Genome Resource (ClinGen) requires genetics-professional review before direct diagnostic or medical use. (www.clinicalgenome.org)

The recommendations below distinguish documented capabilities, source-specific observations, and proposed operating controls. The retrieved sources do not establish current Claude performance across the full path from variant interpretation to IND preparation. The available clinical-genetics benchmark discussed later tests specific tasks and an older model, rather than an end-to-end development program. (arxiv.org) The goal is consequently an **evidence architecture** that makes uncertainty visible and restricts reuse of unresolved claims. No patient examples are needed: a pilot can begin with governed public material and clearly labeled synthetic records.

Key Changes

A closed grant call became a workflow map

Anthropic's **July 20, 2026** call offered **up to \$50,000 in Claude credits for six months**, with applications accepted through **August 2, 2026, at 11:59 PM PST**. (www.anthropic.com) It is now a closed call, not an application opportunity. Its basic-science and early-biotech tracks proposed mechanism discovery, natural-history curation, phenotype matching, modality assessment, endpoints, and regulatory drafting; the explicit

public-output commitment covers **track one**, through Monarch. (www.anthropic.com) Anthropic also acknowledges that insufficient or poorly organized data can limit usefulness. The program should not be read as evidence that missing data can be reconstructed reliably. (www.anthropic.com)

Anthropic names **Every Cure** for repurposing, the **Centre for Population Genomics** for expert-reviewed variant drafts, and **Violet Research Institute** for bioinformatics and regulatory work. These are attributed use-case descriptions, not comparative performance measurements. (www.anthropic.com) Every Cure's own process describes expert assessment followed by laboratory and clinical evidence generation, reinforcing the distinction between computational prioritization and therapeutic substantiation. (everycure.org)

LSVP changes access conditions, not evidence standards

The **September 17, 2026** LSVP announcement describes a beta for verified institutions and teams, with review of credentials, security, and ethical oversight. (www.anthropic.com) Its published scope includes first-party application programming interface (**API**), Enterprise, and Team access, while excluding individual plans, third-party platforms, and organizations enabled for a business associate agreement (**BAA**). (www.anthropic.com)

Standard Use grants renew annually; High-risk Use grants apply to individual research projects and renew every six months. (www.anthropic.com) The page also specifies **30-day retention** for monitoring, states that these data cannot train models or be accessed by its life-sciences research teams, and clarifies that the linked form records interest rather than constituting an application. (www.anthropic.com) These published conditions should be checked before a data steward designs a deployment.

Live validation is becoming more explicit

DisMech's current evidence documentation separates **support direction**, **inferential directness**, and **quotation role**. These identify how a source relates to a claim; they do not grade evidentiary strength. (github.com) Its live disclaimer also describes pre-alpha development, uneven coverage, frequent changes, and insufficient identifier stability for citation as a fixed release. A pilot should pin the repository revision and dependencies itself, rather than treating a live page as an immutable scientific record. (github.com)

Five Workflow Lanes and Their Review Gates

Table 1 proposes an operating matrix. Its prohibitions are governance recommendations; the underlying decision responsibilities follow the cited variant, registry, and IND sources.

Workflow and input class	Intended decision and required evidence	Required reviewer	Prohibited autonomous action
Variant and phenotype: governed genomic records or public examples	Prepare an evidence packet with transcript, sequence version, classification criteria, and conflicting submissions. (hgvs-nomenclature.org) (www.ncbi.nlm.nih.gov)	Clinical molecular geneticist or appropriately qualified laboratory specialist	Issue a clinical classification or convert a VUS into a treatment decision. (pmc.ncbi.nlm.nih.gov)

Workflow and input class	Intended decision and required evidence	Required reviewer	Prohibited autonomous action
Mechanism and literature: publications and curated annotations	Rank hypotheses with exact passages, experimental systems, evidence codes, and alternative explanations. (www.geneontology.org)	Disease biologist and relevant experimental specialist	Represent cross-disease similarity as demonstrated causation.
Natural history and endpoints: governed registry extracts and reports	Assemble an extraction sheet and endpoint rationale tied to population, timing, and scientific objective. (www.fda.gov)	Clinician, registry steward, biostatistician, and outcome specialist	Select the definitive endpoint or declare a cohort fit as a control.
Target and modality: public studies and approved internal experiments	Compare mechanistic rationale, tissue access, assay limitations, and unresolved translational assumptions. (database.ich.org)	Translational scientist, pharmacologist, and modality specialist	Set a starting dose or authorize an experimental treatment.
IND, investigator's brochure (IB), and CMC: controlled reports and source documents	Draft source-linked sections, discrepancy lists, and missing-evidence inventories. (www.fda.gov)	Regulatory writer plus accountable clinical, nonclinical, and CMC owners	Invent results, release a dossier, or submit without authorized review.

The matrix assigns review by consequence rather than document type. A literature summary used for exploration needs less assurance than the same passage used to justify a protocol. FDA's draft AI framework makes context of use central to credibility; a team should define that context before selecting its evaluation tasks. (www.fda.gov)

Variant classification and phenotype matching

Claude for genetic variant interpretation should mean evidence organization under established rules. ACMG/AMP describes five categories, from pathogenic through benign, including uncertain significance. (pmc.ncbi.nlm.nih.gov) ClinGen supplies criterion-specific guidance, including functional, splicing, population, segregation, and computational evidence. The model should retrieve the applicable specification and assemble candidate evidence, leaving criterion application and classification to the qualified reviewer. (www.clinicalgenome.org)

Require explicit transcript and reference-sequence versions. Human Genome Variation Society (HGVS) nomenclature treats missing required sequence versions as invalid, while Ensembl explains that an allele can have different effects in different transcripts. (hgvs-nomenclature.org) (static.ensembl.org) Phenotype matching should retain observed, absent, and unknown findings separately; **Human Phenotype Ontology (HPO)** terms describe abnormalities, not a complete diagnosis. (HPO)

Mechanistic hypotheses and literature synthesis

Claude for rare disease literature review can produce a structured map of experiments and candidate relationships. Preserve gene-product, species, tissue, perturbation, and assay context. **Gene Ontology (GO)** annotations already connect gene products to terms with references and evidence codes, providing a useful

model for traceable extraction. (www.geneontology.org) An unannotated process is not proof of biological absence, and a shared pathway remains a hypothesis until its disease-specific relevance is evaluated. (www.geneontology.org)

Natural history, targets, and regulatory writing

Natural-history curation should preserve what happened, when, and under which observation conditions. FDA describes natural history as disease information collected without an intervention. A model-generated summary must not silently combine treated follow-up with untreated progression. (www.fda.gov) Modality comparison should likewise keep evidence generation separate from inference, consistent with DisMech's distinction between grounding and appraisal. (github.com)

Claude for IND application drafting can populate an approved outline, retrieve precedents, and flag unsupported statements. FDA permits appropriate references to published CMC literature, but the information needed varies with investigation phase, formulation, and duration. A precedent is therefore a candidate rationale for expert assessment, not permission to reuse another program's assumptions. (www.fda.gov)

Evidence-State Architecture

AI evidence synthesis for rare disease drug development needs a durable claim unit. The proposed unit is an atomic assertion with a source passage and explicit downstream permissions. LinkML supplies a schema language and validation framework; DisMech demonstrates schema, ontology, reference, and coverage checks in a concrete project. Neither a valid schema nor a matching quotation establishes the biological conclusion. (linkml.io) (github.com)

Table 2 specifies a proposed claim ledger and its identifier crosswalk. Fields beyond the source projects are recommendations for a pilot, not claims that those projects already implement a regulatory system.

Ledger element	Required record and handling rule
Claim and source	Atomic assertion; PubMed identifier (PMID), accession, or persistent document identifier (ID); exact excerpt; section, table, or figure location; retrieval date. DisMech requires verbatim snippets. (github.com)
Support and uncertainty	Supporting, contradicting, or unclear; direct observation versus inference; primary result versus background or review; unresolved alternatives. These dimensions do not substitute for a strength assessment. (github.com)
Disease identity	Mondo ID and label plus source disease identifiers; preserve exact, broader, narrower, or related mapping. Cross-resource equivalence must be checked. (mondo.monarchinitiative.org)
Phenotype identity	HPO ID, label, release, observation status, onset, and ascertainment method. HPO provides standardized abnormality terms. (HPO documentation)
Mechanism identity	GO ID, aspect, species, gene product, evidence code, and reference. GO covers molecular function, biological process, and cellular component. (www.geneontology.org)
Intervention identity	National Cancer Institute (NCI) Thesaurus (NCIT) ID and release where applicable; retain original intervention wording. NCI publishes the terminology monthly. (www.cancer.gov)

Ledger element	Required record and handling rule
Gene and variant identity	Stable gene identifier, genome assembly, chromosome/position where relevant, alleles, transcript and sequence versions, and HGVS expression. Keep original and normalized forms. (hgvs-nomenclature.org)
Version and transformation	Data cutoff; source version; exact model identifier; prompt, retrieval configuration, code revision, and transformation log. OBO Foundry requires marked, stored ontology releases. (obofoundry.org)
Review and disposition	Reviewer discipline, decision date, rationale, unresolved items, and permitted use: exploration, internal decision support, protocol development, or submission preparation.

This ledger makes **AI evidence traceability in drug development** actionable: reviewers can inspect the assertion and its transformations without reconstructing a chat. Store enough source location detail to distinguish an abstract statement from a full experimental result. DisMech's own evidence model warns that citation grounding does not fully describe experimental design or inference. (github.com)

Resolve mappings without erasing ambiguity

Treat ontology collisions as review events. Orphanet provides stable ORPHAcodes, while its nomenclature pack supplies changes relative to the previous version. Retain both the original identifier and the reviewed successor mapping. (www.orpha.net) (www.orphacode.org) GO obsolete terms retain identifiers while losing relationships; a successful ID lookup therefore needs an obsolescence check. (www.geneontology.org)

Use curated mappings where available, but preserve their provenance and scope. NCI describes paired terminology mappings; Mondo warns against assuming one-to-one correspondence. (www.cancer.gov) (mondo.monarchinitiative.org) A proposed crosswalk should record whether two labels are exact equivalents, differently granular concepts, or unresolved. Global Alliance for Genomics and Health (GA4GH)

Phenopackets supplies a human-readable and machine-readable structure for clinical and phenotypic information, useful for interoperability, but adoption does not itself authorize data sharing or establish a diagnostic conclusion. (www.ga4gh.org)

Regulatory Expectations and Sparse-Data Controls

Match evidence to the question

FDA's **December 2023 final rare-disease guidance** recommends assessing the depth and quality of natural-history knowledge early, while stating that natural-history studies are not universally required. (www.fda.gov) The decision is what evidence the development question needs. Its **May 13, 2024** workshop emphasized registry and natural-history data fit for regulatory purposes. (www.fda.gov) The European Medicines Agency (EMA) registry guideline similarly addresses methodology, data quality, operations, and analysis. (www.ema.europa.eu)

For every extracted cohort, retain the following minimum review fields. This is a proposed extraction specification aligned with the agencies' fitness-for-purpose emphasis, not a universal mandated template. (www.fda.gov) (www.ema.europa.eu)

- **Population:** diagnostic criteria, genotype, phenotype, age range, and recruitment setting.

- **Size:** participants, families where relevant, analysis population, and excluded records.
- **Observation:** baseline, follow-up duration, assessment schedule, and censoring.
- **Exposure:** treatment history, supportive care, and changes during observation.
- **Missingness:** missing variables, reasons where known, and denominator by outcome.
- **Endpoints:** definition, instrument, measurement conditions, and change interpretation.
- **Overlap:** shared centers, recruitment periods, authors, identifiers, and possible duplicate cohorts.
- **Fitness limits:** region, ascertainment, disease stage, and restrictions on reuse.

Endpoint selection needs its own scientific argument. FDA states that clinical endpoint and clinical outcome assessment suitability is context dependent and encourages early discussion. (www.fda.gov) An **estimand**, the precise treatment effect to be estimated, should remain connected to trial design, data collection, and interpretation under ICH E9(R1); a plausible summary measure alone does not settle that question. (www.ema.europa.eu)

Keep sparse-data failure modes visible

The following are review controls, not quantified claims about Claude's error frequency. Each asks whether the evidence supports the proposed use.

- **Publication selection:** search beyond convenient positive reports; log excluded sources and unavailable material.
- **Cohort overlap:** reconcile participant provenance before pooling counts or treating studies as independent.
- **Phenotype mismatch:** distinguish unrecorded findings from documented absence; retain original descriptions alongside HPO terms. ([HPO documentation](#))
- **Ancestry gaps:** test representativeness rather than assuming reference data transfer; NHGRI documents gaps associated with historically European-heavy genomic research. (www.genome.gov)
- **Correlated evidence:** inspect shared training data and overlapping predictors. ACMG/AMP cautions against overestimating computational evidence. (pmc.ncbi.nlm.nih.gov)
- **Unsupported causality:** separate association, perturbation results, animal findings, and human therapeutic evidence.
- **Citation errors:** verify the passage and its relevance, rather than accepting a plausible title or valid identifier. DisMech requires failed reference checks to be corrected or removed. (github.com)
- **Cross-disease transfer:** specify the shared mechanism and the differences that could invalidate transfer.

ClinVar illustrates why retrieval must preserve provenance. It retains prior submission versions and reports review status, but it does not independently curate submitted classifications. A database match therefore supplies a record to appraise, not an automatic resolution of conflicting evidence. (www.ncbi.nlm.nih.gov)

Implementation Considerations and Process Changes

A proposed 60 to 90 day pilot

These Claude workflows for rare disease research and development (R&D) should begin with retrospective, non-sensitive tasks whose reference answers and source documents experts can inspect. Define the intended use narrowly: for example, draft an evidence packet from a frozen publication set, rather than make a development recommendation. FDA's draft AI credibility framework is organized around a specific context of use, supporting this evaluation logic without constituting approval of a particular Claude workflow.

(www.fda.gov)

- **Days 1 to 15:** appoint accountable owners, define material claims, select public or synthetic records, and freeze the source and model configurations.
- **Days 16 to 30:** create an expert-adjudicated reference set, including missing evidence, contradictory sources, obsolete terms, and ambiguous disease labels.
- **Days 31 to 60:** compare assisted and existing workflows on the same task definitions; measure total review time and corrections, not generation time alone.
- **Days 61 to 90:** adjudicate disagreements, analyze error patterns by consequence, and decide whether to stop, narrow scope, or repeat a targeted evaluation.

These intervals are a planning proposal. No fixed accuracy threshold or sample size can be inferred from the access programs. ICH E8(R1)'s fitness-for-purpose principle and the official E6(R3) training material's emphasis on feasible designs support choosing controls proportionate to the work. (database.ich.org)

(admin.ich.org)

Separate review gates by consequence

For **exploration**, permit unconfirmed hypotheses only with visible uncertainty and links to their evidence. For **internal decision support**, require source verification, contradiction review, and an accountable subject-matter owner. For **protocol and endpoint design**, add clinical, outcome-measurement, and biostatistical review and discuss material novelty with the regulator. FDA explicitly encourages early endpoint engagement.

(www.fda.gov)

For **submission preparation**, require controlled sources, reconciliation against approved reports, versioned changes, and authorization by responsible document owners. FDA's IND overview supports early sponsor-agency communication about needed data, while ICH's E6(R3) training material emphasizes procedures across the data life cycle. The model should have no independent submission authority. (www.fda.gov)

(admin.ich.org)

Stop automation when evidence cannot support reuse

Validating Claude for regulatory scientific writing means validating the defined workflow, including retrieval, transformation, and review. Escalation should occur when any of the following proposed stop conditions is met:

- **Genetics:** transcript ambiguity, unresolved classification conflicts, or a VUS promoted beyond its permitted use. (www.clinicalgenome.org)

- **Clinical:** phenotype mismatch, uncertain eligibility, or treatment decisions embedded in a summary.
- **Biostatistics:** unexplained cohort overlap, changing denominators, missingness assumptions, or endpoint sensitivity questions.
- **CMC:** unsupported process parameters, absent batch evidence, or extrapolation across formulations. Required information depends on the investigation context. (www.fda.gov)
- **Regulatory:** source guidance status is unclear, precedent differs materially, or the draft implies agreement that has not been documented.
- **Data stewardship:** provenance, permitted processing, retention, or re-identification constraints are unresolved.
- **Program governance:** a material claim lacks a source, a reviewer, or an approved disposition.

Failures should enter a correction queue with an owner and reuse restriction. A corrected output should preserve the original error and adjudication history so the next evaluation can distinguish repaired evidence from a repeated model failure. This operational recommendation follows the same separation of structured data and scientific appraisal emphasized by DisMech. (github.com)

Data Analysis and Evidence

What the quantitative evidence establishes

The **CGBench preprint**, dated **October 13, 2025**, evaluated **eight language models**, including **Claude Sonnet 3.7**, on clinical-genetics literature tasks. (arxiv.org) Its variant evidence-scoring evaluation used **205 samples**, drawn from **239 total samples** spanning **120 publications, 40 diseases, and 191 variants**. (arxiv.org) These are task-set characteristics, not patient-outcome denominators.

For binary variant evidence verification, the study used **242 evaluation examples** and reported **F1 of 0.576**, a precision-and-recall summary score, for Claude Sonnet 3.7. (arxiv.org) That metric concerns evidence-code verification, not clinical classification accuracy, contemporary Claude performance, or IND readiness. The authors also frame explanation assessment as correspondence with human interpretation rather than absolute correctness. (arxiv.org) A program should therefore use this work to inform evaluation design, not transfer its score into a deployment acceptance threshold.

Operational scale is a different measure. The Centre for Population Genomics reports data from **over 11,600 individuals across 5,900 families, as of October 2026**, and describes collaborative interpretation through CaRDinal. (populationgenomics.org.au) These figures characterize the research resource; they do not quantify Claude's contribution, accuracy, or development-time savings.

Measure the evidence package and the review burden

Table 3 proposes a pilot scorecard. Every threshold must be set prospectively, justified by consequence, and tested on the locked evaluation set. The metrics below are operational definitions proposed for this report, not published benchmark results.

Measure	Proposed calculation	Decision-relevant interpretation
Verified evidence coverage	Verified material claims / all material claims	Count only claims whose passage, interpretation, and permitted use were reviewed.
Contradiction rate	Claims with unresolved material counterevidence / all material claims	Show unresolved and adjudicated contradictions separately.
Unresolved-citation rate	Material claims with missing, inaccessible, or mismatched support / all material claims	A resolving identifier alone is insufficient.
Expert agreement	Matching judgments / adjudicated judgments	Report criterion-level and final-disposition agreement separately.
Material omission rate	Required evidence items missed / required items in the expert reference set	Detects apparently complete packets with important gaps.
Correction rate	Outputs needing material correction / reviewed outputs	Classify corrections by scientific consequence and workflow stage.
Review effort	Median and spread of reviewer time per accepted packet	Include source inspection, reconciliation, correction, and escalation.
False-confidence incidents	Count of high-confidence assertions later rejected as unsupported	Review every incident and its possible downstream reuse.

The denominators matter more than the appearance of precision. Freeze the definition of a material claim, count unverified claims in the denominator, and report performance by workflow. Evidence-code verification should not be pooled with regulatory-writing completeness. Even GO's evidence codes describe evidence type and review status, rather than supplying a universal confidence score. (www.geneontology.org)

Implications and Future Directions

The strongest near-term opportunity is a reusable evidence layer whose outputs remain inspectable as models and sources change. A new model should trigger a targeted re-evaluation of affected tasks, rather than automatic inheritance of a previous model's acceptance. The historical scope of CGBench makes the danger of transferring results across model generations particularly clear. (arxiv.org)

Ontology and database maintenance deserve the same discipline. Store ontology release identifiers, preserve obsolete terms for audit, and review successor mappings. OBO Foundry requires versioned releases with unique version identifiers; an unversioned label match cannot reproduce the knowledge state on which a decision depended. (obofoundry.org) Similarly, preserve ClinVar submission versions and distinguish a record's review status from the underlying evidence for that individual variant. (www.ncbi.nlm.nih.gov)

Patient-foundation teams can contribute useful curation and registry expertise without treating a language model as a clinical authority. The National Center for Advancing Translational Sciences (**NCATS**) RaDaR resource offers registry guidance based on best practices and data standards. Its role illustrates the value of explicit registry design before downstream analysis. (registries.ncats.nih.gov) For developers, the complementary opportunity is to connect these governed sources to transparent extraction and review processes.

Three development priorities follow from the evidence reviewed:

- **Measure transfer:** test whether improvements in retrieval actually improve expert-accepted decisions in the intended disease and modality.
- **Evaluate omissions:** include negative findings, unavailable evidence, and contradictory reports in reference tasks.
- **Preserve context:** keep the population, measurement schedule, intervention history, and estimand attached to every reused endpoint argument. EMA's E9(R1) discussion connects design and interpretation and keeps sensitivity analyses aligned to the same estimand. (www.ema.europa.eu)

Early regulator interaction should resolve material endpoint or evidentiary questions before a polished draft creates an impression of consensus. ICH E8(R1) encourages early engagement when novel elements matter to quality. Better drafting can make that conversation clearer; the conversation still needs the underlying scientific argument. (database.ich.org)

Conclusion

Claude-assisted rare-disease development is most defensible when the output is a bounded, reviewable work product: a variant-evidence packet, a mechanism map, a natural-history extraction sheet, a modality rationale, or a source-linked regulatory draft. The useful distinction is the decision each artifact may inform and the evidence required before it can be reused.

The proposed claim ledger supplies that discipline. It connects an assertion to an exact source passage, a stable identifier, a versioned knowledge state, an explicit transformation, uncertainty, an accountable reviewer, and a permitted downstream use. Ontology checks, valid references, and schema conformity make the record inspectable. Scientific appraisal determines whether the record supports the conclusion.

A retrospective pilot should expose the review burden as carefully as it measures generation speed. Teams should compare evidence coverage, omissions, contradictions, corrections, and false confidence under a prospectively defined evaluation plan. They should retain the ability to stop a workflow or narrow its scope when unresolved evidence prevents reliable reuse.

For program leaders, the investment decision is therefore specific: select a recurring task with inspectable inputs and an accountable expert owner, define its evidence package, and test it against an existing process. Expand only when the evidence supports the intended use. Access programs and research announcements can motivate that experiment; they do not replace its results.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. <https://www.anthropic.com/news/rare-disease-research-grants>
2. <https://www.anthropic.com/news/life-sciences-verification-program>
3. <https://pmc.ncbi.nlm.nih.gov/articles/PMC25741868>
4. <https://www.fda.gov/drugs/types-applications/investigational-new-drug-application-ind>

5. <https://github.com/monarch-initiative/dismech>
6. <https://github.com/monarch-initiative/dismech/blob/main/docs/explanation/evidence-model.md>
7. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-considerations-development-drugs-and-biological-products>
8. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological>
9. https://database.ich.org/sites/default/files/E8-R1_Guideline_Step4_2021_1006.pdf
10. <https://mondo.monarchinitiative.org>
11. <https://www.orpha.net/en/disease>
12. <https://www.clinicalgenome.org/tools/clingen-variant-classification-guidance>
13. <https://arxiv.org/html/2510.11985v1>
14. <https://everycure.org/about>
15. <https://github.com/monarch-initiative/dismech/raw/refs/heads/main/docs/disclaimer.md>
16. <https://hgvs-nomenclature.org/stable/background/refseq>
17. https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status
18. <https://www.geneontology.org/docs/go-annotations>
19. <https://www.fda.gov/drugs/news-events-human-drugs/natural-history-studies-and-registries-development-rare-disease-treatments-05132024>
20. https://static.ensembl.org/info/genome/variation/prediction/predicted_data.html
21. <https://human-phenotype-ontology.github.io/about.html>
22. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-natural-history-studies-drug-development>
23. <https://www.fda.gov/drugs/investigational-new-drug-application-ind/ind-applications-clinical-investigations-chemistry-manufacturing-and-control-cmc-information>
24. <https://linkml.io/linkml>
25. <https://human-phenotype-ontology.github.io/documentation.html>
26. <https://www.geneontology.org/docs/ontology-documentation>
27. <https://www.cancer.gov/about-nci/organization/cbiit/vocabulary>
28. <https://obofoundry.org/principles/fp-004-versioning.html>
29. <https://www.orphacode.org/pack-nomenclature>
30. <https://www.ga4gh.org/product/phenopackets>
31. <https://www.fda.gov/media/119757/download>
32. <https://www.ema.europa.eu/en/guideline-registry-based-studies-scientific-guideline>
33. <https://www.ema.europa.eu/en/scientific-guidelines/ich-e9-statistical-principles-clinical-trials>
34. <https://www.genome.gov/about-genomics/fact-sheets/Participation-in-Genomic-Research>
35. <https://www.ncbi.nlm.nih.gov/clinvar/intro>
36. https://admin.ich.org/sites/default/files/inline-files/ICH_E6%28R3%29_Step%204_Presentation_2025_0123.pdf
37. <https://populationgenomics.org.au/our-impact/rare-disease>

38. <https://registries.ncats.nih.gov/about-radar>

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore [AI enablement](#), [pharma and biotech software development](#), [data engineering and BI](#), and [Veeva services](#). [Contact IntuitionLabs](#) to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: <https://intuitionlabs.ai>

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.