



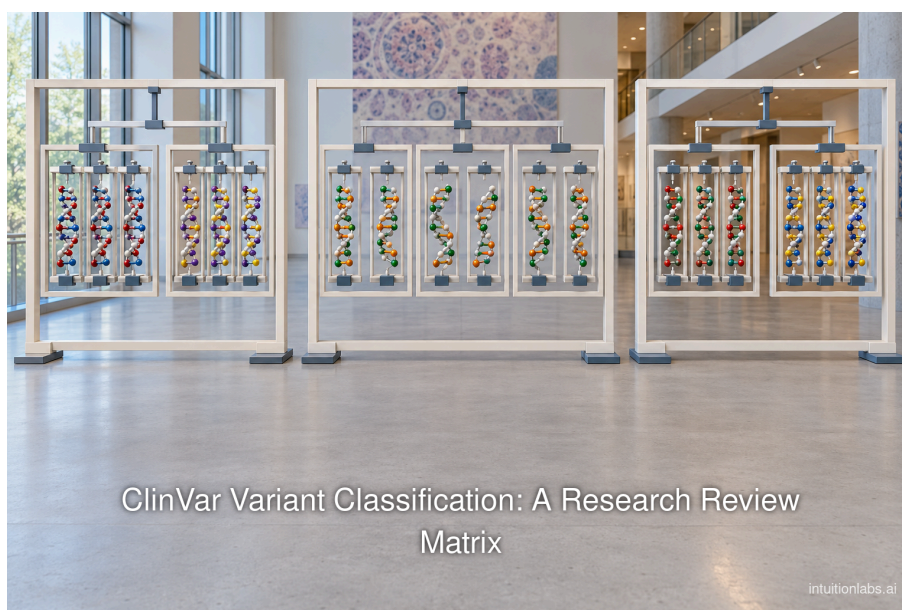
INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

ClinVar Variant Classification: A Research Review Matrix

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 · 2026-10-31 · clinvar variant classification · clinvar review status · clinvar stars · somatic clinical impact · oncogenicity · germline classification · variant curation · genomic data pipelines

Original article: <https://intuitionlabs.ai/articles/clinvar-germline-somatic-oncogenicity-review-matrix>



In this article

1. Executive Summary
 2. Introduction and Background
 3. Key Changes
 4. Assertion Identity and Disease Context
 5. Review Status Acceptance Matrix
 6. Implementation Considerations and Process Changes
 7. Release and Schema Controls
 8. Curator Escalation and Versioned Refresh
 9. Data Analysis and Evidence
 10. Case Studies and Real-World Examples
 11. Implications and Future Directions
 12. Conclusion
-

Executive Summary

ClinVar variant classification should enter a research pipeline as a context-bearing assertion, not a star-filtered label. The database distinguishes **germline classification**, **somatic clinical impact**, and **oncogenicity**, while submitted records (SCV), variant-condition aggregates (RCV), and variant aggregates (VCV) represent different scopes. An RCV narrows the question to a variant-condition pair; a project should preserve that context rather than treating a variant-wide aggregate as universal evidence.

(www.ncbi.nlm.nih.gov) Germline assessment additionally depends on disease and inheritance context, while oncology evidence distinguishes biological effect from tumor-specific clinical implications. (www.nature.com) (faq.oncokb.org)

The decisive review-status distinction is that **two-star somatic clinical impact does not require consensus**. It reflects multiple qualifying submissions, so different tiers or assertion purposes must not be turned into an agreement flag. (www.ncbi.nlm.nih.gov) A review threshold can help users browse, as the Genome Aggregation Database (gnomAD) browser filter demonstrates, but a research acceptance policy also needs an allele match, disease match, assertion purpose, accessible evidence, and a defined intended use. (gnomad.broadinstitute.org) (civic.readthedocs.io) The proposed matrix therefore separates automatic ingestion from curator approval. High-review assertions receive priority for contextual assessment; conflicts, uncertain significance, unsupported mappings, and missing individual-variant classifications require explicit handling.

Implementation should preserve **raw review-status text**, lane-specific classifications, reference-sequence versions, condition mappings, and a reproducible decision history. Human Genome Variation Society (HGVS) nomenclature requires the appropriate reference version, and the Global Alliance for Genomics and Health (GA4GH) Variation Representation Specification (VRS) recommends normalization while warning that computed identifiers need not remain stable across major specification versions. (hgvs-nomenclature.org)

(vrs.ga4gh.org) A publication-linked evidence model can preserve the source behind an assertion instead of exposing only a derived conclusion. (civic.readthedocs.io) The source-linked dictionary and regression checks below make these distinctions testable.

The quantitative evidence supports monitoring change, without supplying a universal acceptance rate. A **2026 accepted historical study** analyzed **3.68 million variants across 148 monthly snapshots**; its scope is historical, not a current ClinVar census. (link.springer.com) A separate hereditary-disease cohort reported **94,453 reclassified unique variants, or 4.60%**, and explicitly limited generalization beyond its laboratory processes. (jamanetwork.com) This report measured no local acceptance rate and inspected no complete compressed release. Its deliverable is a proposed research policy: preserve context and provenance, escalate unresolved applicability, and require reproducible imports and curator confirmation before consequential downstream use.

Introduction and Background

ClinVar variant classification becomes an engineering decision when a platform must decide whether an assertion can support a rare-disease candidate, an oncology biomarker hypothesis, or an automated evidence summary. ClinVar separates germline classification, somatic clinical impact, and oncogenicity; a submitted **SCV** record represents one classification type. These lanes answer different questions and should remain separate in an evidence model. (www.ncbi.nlm.nih.gov) (www.nature.com)

The distinction matters beyond database terminology. The American College of Medical Genetics and Genomics and Association for Molecular Pathology (**ACMG/AMP**) germline framework evaluates pathogenicity in a condition and inheritance context. The Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists (**AMP/ASCO/CAP**) framework organizes somatic clinical significance into tiers. Biological oncogenicity precedes an assessment of actionability in the Clinical Genome Resource, Cancer Genomics Consortium, and Variant Interpretation for Cancer Consortium (**ClinGen/CGC/VICC**) recommendations. (www.nature.com) (www.amp.org) (pmc.ncbi.nlm.nih.gov)

This report proposes a **research acceptance policy**, not an official ClinVar eligibility rule. Its audience is variant-curation leads, translational scientists, and platform architects. For IntuitionLabs' professional research readership, the relevant decision is which assertions may enter a reproducible research workflow and which require a curator's assessment. The proposed policy treats ingestion, contextual matching, and permission to use an assertion as separate operations. Computable variant representation and publication-linked evidence models offer useful architectural precedents. (vrs.ga4gh.org) (civic.readthedocs.io)

The evidence was retrieved on **October 31, 2026**. Research covered official documentation, release listings, classification standards, ontology guidance, and original studies. It did not download and count the complete release or test clinical outcomes. ClinVar documents weekly updates and monthly archives. A reproducible import should distinguish the artifact filename, directory timestamp, and internal release date rather than infer one from another. (www.ncbi.nlm.nih.gov)

Key Changes

Separate the three classification lanes

For **germline classification**, the intended question concerns a variant's relationship to a specified inherited condition. The familiar pathogenic, likely pathogenic, uncertain significance, likely benign, and benign categories belong to the ACMG/AMP framework. Its condition and inheritance context should accompany a research assertion rather than being reconstructed from a gene symbol. ACMG's resource catalog continues to identify the joint guideline as current laboratory guidance. (www.nature.com) (www.acmg.net)

For **somatic clinical impact**, the question concerns therapeutic, diagnostic, or prognostic significance. The AMP/ASCO/CAP framework distinguishes strong, potential, unknown, and benign or likely benign significance. Its evidence levels also distinguish a treatment association in a specified tumor type from evidence involving a different tumor type or clinical-trial eligibility. A downstream evidence object should preserve the original tier and context instead of relabeling every positive oncology assertion as pathogenic. (www.amp.org) (jmd.amjpathol.org)

For **oncogenicity**, the question concerns a variant's biological role in cancer. The ClinGen framework's development included a defined evaluation set, and the recommendations distinguish this assessment from clinical actionability. OncoKB likewise separates mutation effect, oncogenic effect, and tumor-specific clinical implications. The proposed schema should therefore make biological effect and treatment relevance independently queryable. (clinicalgenome.org) (pmc.ncbi.nlm.nih.gov) (faq.oncokb.org)

Interpret review status within its classification type

The consequential ClinVar distinction is that **somatic clinical-impact consensus is not required** for a multiple-submitter review status. Different tumor types and assertion purposes can legitimately yield different tiers. The germline and oncogenicity aggregate rules instead distinguish agreement from conflicting classifications. This makes the same visual star count insufficient to establish the same evidentiary conclusion across lanes. (www.ncbi.nlm.nih.gov)

The gnomAD browser's review-status filter is useful for browsing a variant track, but its threshold is not a disease-matching or intended-use policy. A platform can retain that threshold as a search convenience while requiring additional contextual fields before accepting an assertion into a project. This is the report's proposed engineering interpretation of the filter's documented function. (gnomad.broadinstitute.org) (civic.readthedocs.io) (faq.oncokb.org)

Apply precedence without erasing provenance

ClinVar aggregation uses practice-guideline and expert-panel precedence ahead of ordinary submissions. Other submitted classifications remain present even when excluded from the aggregate calculation. The operational implication is to retain the contributing submissions and distinguish the selected aggregate from the available evidence history. (www.ncbi.nlm.nih.gov)

Precedence should not become an undocumented override inside a research system. Store the selected assertion, selection rationale, original evidence references, and local decision separately. A publication-linked evidence model helps a curator inspect the basis of an assertion, while a disease ontology mapping provides a

separate check on whether the selected context matches the project. (civic.readthedocs.io) (mondo.monarchinitiative.org)

Assertion Identity and Disease Context

SCV, RCV, and VCV are different scopes of information. An SCV is a submitted assertion; an RCV aggregates a variant-condition pair; a VCV aggregates a variant across submitted conditions. For a disease-specific research question, the VCV is a discovery entry point, while the relevant condition-specific aggregate and contributing submissions establish the narrower context. (www.ncbi.nlm.nih.gov)

An allele identifier answers a different question from an assertion identifier. The ClinGen Allele Registry supplies variant identifiers, and the Global Alliance for Genomics and Health (**GA4GH**) Variation Representation Specification (**VRS**) supports computable variant exchange. Neither function by itself establishes that an assertion applies to a particular disease, inheritance model, tumor type, or drug hypothesis. The local data model should keep allele identity and assertion applicability separate. (reg.genome.network) (www.ga4gh.org) (www.nature.com) (faq.oncokb.org)

Preserve reference and disease identity

The Human Genome Variation Society (**HGVS**) nomenclature defines a variant relative to a reference sequence and requires versioned descriptions where reference versions apply. VariantValidator can validate descriptions and map transcript and genomic representations. These capabilities support input reconciliation; they do not replace classification review. (hgvs-nomenclature.org) (variantvalidator.org)

Normalization should therefore retain both the submitted representation and the normalized result. VRS recommends normalization for interoperability, and its computed identifiers depend on the represented object. The Registry's canonical allele identifiers provide another reconciliation route. Record the method, reference sequence, assembly, and software or specification version used for every transformation. (vrs.ga4gh.org) (reg.genome.network)

Disease matching needs equally explicit provenance. ClinVar accepts identifiers from several disease resources, while Mondo distinguishes equivalence from weaker mapping relationships. A cross-reference should not silently authorize transferring an assertion between a broad syndrome, a molecular subtype, and a related phenotype. Keep the source identifier, mapping predicate, target identifier, and mapping release together. (www.ncbi.nlm.nih.gov) (mondo.monarchinitiative.org)

The preferred disease label is a display value. ClinVar and related National Center for Biotechnology Information (NCBI) resources can share preferred terminology, and MedGen documents identifier history including circumstances in which a concept identifier change changes its numeric identifier. Store identifier history and original labels so that a renamed concept does not masquerade as new biological evidence. (www.ncbi.nlm.nih.gov)

The proposed minimum identity envelope is:

- **Assertion scope:** retain submitted, variant-condition, and variant-aggregate identifiers as separate keys; keep biological effect distinct from actionability. (faq.oncokb.org)

- **Reference context:** retain assembly and reference accession with version, following sequence-description requirements. ([hgvs-nomenclature.org](https://hgvsonline.org/))
- **Allele resolution:** retain original description, normalized representation, and any canonical allele mapping. (reg.genome.network)
- **Condition resolution:** retain source condition identifiers and the exact relation used in ontology mapping. (mondo.monarchinitiative.org)
- **Oncology context:** retain tumor type, assertion purpose, therapy, and sensitivity or resistance where supplied. (civic.readthedocs.io)
- **Evidence trace:** retain source publication or abstract identifiers and the passage or evidence object used. (civic.readthedocs.io)

These are proposed acceptance fields, not a claim that each is populated in every ClinVar record. Missing contextual data should produce an explicit applicability gap rather than a guessed value.

Review Status Acceptance Matrix

A research platform needs an **action vocabulary** independent of stars: ingest for discovery, accept for a defined contextual use, escalate to a curator, or retain only as historical provenance. The same raw assertion may be acceptable for documenting uncertainty yet unsuitable as a positive label for model training. The ACMG/AMP caution about uncertain significance and the National Cancer Institute's (NCI) explanation of unknown-effect biomarkers illustrate why uncertainty should remain visible. (www.nature.com) (www.cancer.gov)

Table 1 maps documented review-status distinctions to a proposed research policy. The status definitions come from ClinVar; the actions and escalation triggers are recommendations in this report. The table applies separately to each relevant classification lane and aggregate scope. Practice guidelines carry four stars, expert panels three, qualifying aggregates two, and other documented statuses one or zero. (www.ncbi.nlm.nih.gov)

Lane and record level	Exact status or status family	Minimum provenance	Allowed research action	Escalation trigger
All lanes, SCV; corresponding RCV or VCV aggregate	practice guideline; reviewed by expert panel	Version, condition, lane, criteria, evidence, evaluation date	Prioritize contextual assessment and ingest with provenance	Project disease or assertion purpose differs
Germline or oncogenicity, RCV/VCV	criteria provided, multiple submitters, no conflicts	Aggregate plus contributing SCVs and disease context	Accept as a documented aggregate for the defined research question	Hidden context mismatch or unsupported local conclusion
Somatic clinical impact, RCV/VCV	criteria provided, multiple submitters	Tumor, purpose, tier, therapy if relevant, contributing SCVs	Ingest context-specific evidence; do not infer consensus	A workflow requires agreement on the same clinical claim

Lane and record level	Exact status or status family	Minimum provenance	Allowed research action	Escalation trigger
All lanes, SCV/RCV/VCV where applicable	criteria provided, single submitter	Submitter, criteria, evidence or public contact, date	Ingest as a single-source assertion	Consequential use requires independent corroboration
Germline or oncogenicity, RCV/VCV	criteria provided, conflicting classifications	Conflicting values, relevant SCVs, context and dates	Retain disagreement as evidence; queue review before positive-label use	Any automatic selection of one side
All applicable lanes and levels	no assertion criteria provided	Original submission, evidence availability and contact if present	Discovery and provenance retention	Promotion into a curated evidence set
SCV and applicable germline, somatic clinical-impact, and oncogenicity aggregates	no classification provided	Record identity and submitted context. (NCBI ClinVar review status)	Retain identity and metadata without a classification label	A downstream system attempts to impute a label
Germline, somatic clinical-impact, or oncogenicity aggregate, included variant	no classification for the individual variant	Containing haplotype or genotype and relationship. (NCBI ClinVar review status)	Retain set membership	Transfer of the set's classification to an isolated allele

The matrix deliberately makes **two-star somatic impact** a contextual-ingestion route rather than an agreement route. The absence of a calculated conflict is not evidence that therapeutic, diagnostic, and prognostic assertions agree. For germline or oncogenicity conflicts, the competing values should remain visible. (www.ncbi.nlm.nih.gov)

A criteria-provided status also describes documentation of the classification method, not necessarily a complete public evidence package for the individual variant. ClinVar does not review the detailed variant-scoring criteria. A local curator should therefore distinguish a documented method, accessible supporting evidence, and fitness for the intended project. (www.ncbi.nlm.nih.gov)

Implementation Considerations and Process Changes

Use XML for assertion detail and VCF for scoped annotation

Extensible Markup Language (XML) and **Variant Call Format (VCF)** serve different integration needs. ClinVar describes XML as the complete public dataset and VCF as summary-level data for variants within its location constraints. The documented VCF scope includes simple alleles shorter than 10 kilobases with precise endpoints on GRCh37 or GRCh38. It does not represent every variant class or submission relationship. (www.ncbi.nlm.nih.gov)

Use a rich assertion store as the source of research decisions and expose a derived annotation view where convenient. Ensembl's Variant Effect Predictor (**VEP**) provides allele-specific clinical-significance annotation and an optional somatic-classification output. Its documentation also warns that disabling allele-specific clinical

significance reports known values at the locus. An integration test should distinguish matching an allele from merely matching a position. (static.ensembl.org)

Table 2 is a source-linked field dictionary. It identifies documented lane mappings and adds proposed storage requirements. XML element names describe the classification containers; the documented VCV aggregate germline path is `ClassifiedRecord/Classifications/GermlineClassification/ReviewStatus`, while the RCV path begins `ReferenceClinVarAssertion/Classifications/GermlineClassification/ReviewStatus`. The somatic and oncogenicity containers are siblings of `GermlineClassification`. Submitted-record review status uses `ClassifiedRecord/ClinicalAssertionList/ClinicalAssertion/Classification/ReviewStatus` in VCV XML and `ClinVarAssertion/Classification/ReviewStatus` in RCV XML. Validate each path against the pinned XML format. ([NCBI ClinVar review-status XML mappings](#)) Actual parser paths must be checked against the pinned schema and record level. (www.ncbi.nlm.nih.gov)

Concern	Documented field or representation	Proposed preservation and test
Germline review	CLNREVSTAT; XML <code>GermlineClassification</code> review status	Keep raw text and a separate derived star value.
Oncogenicity review	ONCREVSTAT; XML <code>OncogenicityClassification</code> review status	Prohibit fallback to germline status when absent.
Somatic clinical-impact review	SCIREVSTAT; XML <code>SomaticClinicalImpact</code> review status	Test multiplicity separately from agreement.
Classification value	CLNSIG, ONC, SCI for the respective aggregate lanes	Keep lane and raw classification together. (ftp.ncbi.nlm.nih.gov)
Condition fields	CLNDN/CLNDISDB, ONCDN/ONCDISDB, SCIDN/SCIDISDB	Preserve lane-specific names and identifier sets. (ftp.ncbi.nlm.nih.gov) (mondo.monarchinitiative.org)
Included variants	Germline: CLNSIGINCL, CLNDNINCL, CLNDISDBINCL; oncogenicity: ONCINCL, ONCDNINCL, ONCDISDBINCL; somatic clinical impact: SCIIINCL, SCIDNINCL, SCIDISDBINCL	Preserve each lane's containing haplotype or genotype and its Variation ID:classification pairs separately, with that lane's included-variant condition names and identifiers. Do not copy the containing set's classification onto the isolated allele. (ftp.ncbi.nlm.nih.gov)
Sequence identity	Versioned reference context and normalized allele representation	Store original text, assembly, accession version, and normalization method. (hgvs-nomenclature.org) (vrs.ga4gh.org) (variantvalidator.org)
External evidence	Publication-linked statements, disease and therapy context	Keep provenance and purpose as distinct fields. (civic.readthedocs.io)

The crosswalk should not flatten every classification into a single clinical-significance column. It should also avoid assuming that an external annotation tool preserves every field needed for acceptance. The contrast between VEP's allele and locus options demonstrates a specific configuration dependency; the proposed policy requires that dependency to be captured in the transformation manifest. (static.ensembl.org)

Release and Schema Controls

Pin the release and schema independently

The observed release listing and schema directory are evidence of available artifacts, not proof of a completed import. The fetched directories listed VCV schema 2.6 and RCV schema 2.3; the compressed releases' internal schema linkage was not inspected. Production acceptance should require successful validation of the actual payload against the declared schema. Directory observations alone should not authorize that assertion. (<ftp.ncbi.nlm.nih.gov>)

Current documentation also needs reconciliation with older pages. The XML README states that the older formats have been unsupported since August 7, 2025. The local manifest should name the actual source format rather than inheriting an old parser's label. GA4GH's identifier documentation provides a parallel reason to pin versions: computed identifiers are not guaranteed stable across major specification releases. (<ftp.ncbi.nlm.nih.gov>) (<vrs.ga4gh.org>)

The proposed import manifest should capture:

- **Artifact identity:** source URL, exact filename, retrieval time, byte length, and locally calculated checksum; identify any representation dependency. (<vrs.ga4gh.org>)
- **Release identity:** internal release date if parsed, directory timestamp, and archive cadence as different values.
- **Schema identity:** declared namespace or schema reference, validated schema file, and parser version.
- **Reference identity:** assembly, sequence accession versions, and normalization configuration. (<hgvs-nomenclature.org>) (<vrs.ga4gh.org>)
- **Annotation identity:** VEP options or equivalent enrichment settings, including allele specificity. (<static.ensembl.org>)
- **Decision identity:** policy version, curator outcome, approved use, and reasons for exclusion; preserve tumor-specific therapeutic evidence. (<faq.oncokb.org>) A curator must confirm the intended use.
- **Replay identity:** previous imported release and the reproducible transformation configuration, including allele normalization. (<vrs.ga4gh.org>)

These are proposed operational controls. Their purpose is to let a reviewer reproduce an evidence selection and explain why a later run selected a different assertion. Before promoting an import, require a recorded reviewer disposition for any schema difference that changes field meaning, cardinality, or missing-value behavior. Keep these decisions with the replay manifest.

Curator Escalation and Versioned Refresh

The escalation queue should be based on **what is unresolved**, not merely on a low star count. A high-review assertion can still require context review. Conversely, a low-review assertion can be useful for discovery if its limitations remain explicit. Germline inheritance context, oncology tumor specificity, and ontology relation semantics supply different reasons to escalate. (<www.nature.com>) (<faq.oncokb.org>) (<mondo.monarchinitiative.org>)

A practical queue can include:

- **Context mismatch:** disease, inheritance, tumor subtype, or assertion purpose differs from the project. (www.nature.com) (civic.readthedocs.io)
- **Unresolved disagreement:** relevant germline or oncogenicity assertions point to different classifications; check the disease and inheritance scope. (www.nature.com)
- **Unavailable evidence:** the supporting publication, data, or submitter explanation cannot be inspected. (civic.readthedocs.io)
- **Uncertain significance:** the intended use would turn uncertainty into a positive clinical or biological label. (www.nature.com) (www.cancer.gov)
- **Representation mismatch:** transcript, assembly, reference version, or normalized allele cannot be reconciled. (hgvs-nomenclature.org) (variantvalidator.org)
- **Unsupported disease mapping:** an external cross-reference is being treated as equivalence without a suitable relation. (mondo.monarchinitiative.org)
- **Temporal mismatch:** evaluation predates important evidence or the project requires a new contextual assessment. (jmd.amjpathol.org)
- **Oncology use mismatch:** oncogenicity is being treated as a therapy recommendation without a tumor-specific therapeutic association. (jmg.bmj.com) (faq.oncokb.org)

Keep date meanings and lifecycle events separate

A record version is not a direct freshness score. ClinVar documents that resubmission can increment an SCV version even when only the submission date changes. Certain NCBI annotation edits do not increment aggregate versions. The date a submitter last evaluated a variant is a separate field from these update events. (www.ncbi.nlm.nih.gov)

Deletion and replacement need explicit handling. ClinVar documents removal of deleted submissions from current displays and future files while preserving historical archives. Replaced accessions can remain searchable. A local lifecycle model should distinguish current eligibility, historical retention, replacement links, and a missing record in a particular extract. This research did not verify a universal XML withdrawn flag, so the policy does not assume one. (www.ncbi.nlm.nih.gov)

Use a documented refresh cadence and a separate evidence-review cadence. The upstream release documentation describes weekly updates and monthly archives; AMP/ASCO/CAP calls for ongoing reassessment of significance. The SVIG-UK page describes yearly framework review. These are different activities, and none establishes an automatic project-specific expiry interval. (www.ncbi.nlm.nih.gov) (jmd.amjpathol.org) (www.cangene-canvaruk.org)

The following proposed regression checks belong in every release comparison:

- **Lane isolation:** missing somatic impact never inherits a germline or oncogenicity classification; evidence purposes remain explicit. (civic.readthedocs.io)
- **Status fidelity:** the raw status survives parsing and round-trip export without a star-only substitution; browser thresholds remain a convenience. (gnomad.broadinstitute.org)

- **Allele specificity:** a locus-level annotation cannot silently replace an allele-specific match. (static.ensembl.org)
- **Reference fidelity:** normalization retains the source representation and reference versions. (hgvs-nomenclature.org) (variantvalidator.org)
- **Condition fidelity:** mapping predicates and ontology versions survive transformation. (mondo.monarchinitiative.org)
- **Evidence traceability:** each exported assertion can be traced to source evidence and its context. (civic.readthedocs.io)
- **Lifecycle fidelity:** historical assertions remain auditable while current eligibility can change; measure change using a defined outcome. (link.springer.com)
- **Identifier reproducibility:** computed identifiers can be regenerated under the pinned representation specification. (vrs.ga4gh.org)

Passing these checks would demonstrate the specified software behavior. It would not demonstrate the correctness of an individual variant classification or establish clinical eligibility.

Data Analysis and Evidence

The quantitative evidence supports **versioned review**, but it does not provide a universal acceptance percentage. Historical ClinVar analyses, individual-laboratory cohorts, and framework-development exercises have different denominators. They should be reported separately rather than pooled into a single estimate of database reliability. (link.springer.com) (jamanetwork.com) (clinicalgenome.org)

Table 3 distinguishes the measured objects and the research decision each can inform. None is a newly measured census in this report.

Originating evidence	Measured population and result	Permitted interpretation
Accepted Genome Medicine study published October 23, 2026	3.68 million variants, 148 monthly snapshots spanning 2012 through 2025; 9.3% changed aggregate category. (link.springer.com)	Historical aggregate changes justify longitudinal diffing; these are not present-day error rates.
Same historical analysis, different outcome	0.68%, or 25,105 variants, crossed the authors' pathogenic/likely pathogenic versus uncertain/benign/likely benign boundary; 2.6% of reclassification events were attributed to the RCV-to-VCV transition. (link.springer.com)	Separate biologically consequential boundary changes from aggregate-category changes and format-related artifacts.
Original hereditary-disease testing cohort	2,051,736 unique variants in 3,272,035 individuals; 94,453 variants, or 4.60%, were reclassified. (jamanetwork.com)	Evidence of change in that cohort, not an estimate for all ClinVar lanes.
Oncogenicity framework development	ClinGen describes evaluation on 94 somatic variants in 10 genes; SVIG-UK describes 20 test variants across 13 cancer genes. (clinicalgenome.org) (img.bmj.com)	Defined development sets demonstrate evaluation scope, not a database-wide accuracy benchmark.

The Genome Medicine publication is an accepted early version, and its fetched abstract describes an era-stratified heuristic approach. Its aggregate-category and boundary-change outcomes should retain those definitions. The hereditary-disease study covers testing from January 2015 through June 2023 and cautions that its conclusions come from one laboratory's processes. The denominators are neither interchangeable nor a basis for projecting an oncology pipeline's expected escalation rate. ([link.springer.com](#)) ([jamanetwork.com](#))

A useful local measurement program would calculate rates **after** defining the project subset, release, record scope, and classification lane. Report unique eligible accessions separately from assertions, classification events, and allele representations. An identifier reconciliation study, for example, measures a different property from pathogenicity assessment; the historical Allele Registry work should not be converted into a claim about current clinical-classification accuracy. ([onlinelibrary.wiley.com](#)) ([reg.genome.network](#)) ([www.ga4gh.org](#))

Recommended local measurements include:

- **Context coverage:** proportion of imported assertions with an explicit disease match and recorded mapping relation. ([mondo.monarchinitiative.org](#))
- **Evidence coverage:** proportion with retrievable source evidence, separated from criteria-only documentation. ([civic.readthedocs.io](#))
- **Reconciliation coverage:** proportion with a resolved reference-aware allele identity. ([hgvs-nomenclature.org](#)) ([reg.genome.network](#))
- **Decision reproducibility:** proportion for which the same artifact and policy reproduce the same disposition, with a versioned representation. ([vrs.ga4gh.org](#))
- **Refresh effects:** counts of new, changed, absent, and replaced assertions, with reasons assessed separately.
- **Queue outcomes:** curator dispositions by lane and escalation reason, without generalizing a particular cohort's reclassification rate. ([jamanetwork.com](#))

No values for these local measures were collected here. Publishing an acceptance rate without the underlying accession list and release identity would substitute an unsupported number for a defined research result.

Case Studies and Real-World Examples

Worked acceptance cases (Hypothetical Example)

All identifiers, disease labels, timestamps, and release labels below are **fictional test inputs**, not live ClinVar records. They are included to make the proposed policy executable.

Germline case (Hypothetical Example). Accession SYNTH-RCV-G1, lane germline, disease Fictional Rare Disease A, assembly GRCh38, retrieval date 2026-10-31, release label SYNTH-2026-10-01. The aggregate has a multiple-submitter, no-conflicts status, but the project targets Fictional Rare Disease B. Expected action: retain for discovery and escalate the condition mapping. Agreement does not supply the missing disease match; classification remains conditional on its disease and inheritance context. ([www.nature.com](#)) ([mondo.monarchinitiative.org](#))

Somatic case (Hypothetical Example). Accession SYNTH-VCV-S1, lane somatic clinical impact, disease Fictional Tumor A, assembly GRCh38, retrieval date 2026-10-31, release label SYNTH-2026-10-01. Separate submissions concern therapeutic and diagnostic purposes under a multiple-submitter status. Expected action: create distinct evidence objects and prohibit an inferred consensus flag. Tumor-specific therapeutic association and diagnostic evidence remain different claims. (jmd.amjpathol.org) (faq.oncokb.org) (civic.readthedocs.io)

Included-variant case (Hypothetical Example). Accession SYNTH-VCV-H1, lane germline, disease Fictional Haplotype Disorder, assembly GRCh38, retrieval date 2026-10-31, release label SYNTH-2026-10-01. The individual variant has no direct classification, but its containing set does. Expected action: preserve membership and block copying the set's classification onto the isolated allele. (<ftp.ncbi.nlm.nih.gov>)

The proposed test harness should load each fixture with its complete identity envelope, run the stated acceptance policy, and compare the resulting disposition with the expected action. It should preserve the source record separately from that disposition so that a policy change can be replayed without rewriting the input. A curator should inspect any unexpected promotion into an accepted evidence set. For each fixture, the acceptance result should include an explanation code and a durable link to the preserved input, making disagreements between software and curator visible. This fixture strategy tests transformation and policy behavior; the supporting concepts remain reference-aware variant representation and traceable evidence. (vrs.ga4gh.org) (civic.readthedocs.io)

Implications and Future Directions

The central architectural implication is to treat **ClinVar as an assertion source with context and history**, rather than a ready-made training-label column. Stable allele exchange, validated sequence descriptions, disease mappings, and publication-linked evidence solve different parts of that problem. A robust design composes them while preserving their boundaries. (www.ga4gh.org) (variantvalidator.org) (mondo.monarchinitiative.org) (civic.readthedocs.io)

For **drug-development evidence** and rare-disease discovery, the acceptance policy should preserve the distinction between candidate generation and evidence strong enough for a specified research claim. A variant of uncertain significance can remain a research candidate without becoming a positive disease label. Germline disease and inheritance context remain essential to that distinction. (www.nature.com) (www.acmg.net)

For oncology translation, **oncogenicity and clinical impact should remain independently inspectable**. Clinical actionability also depends on jurisdiction and access context under the SVIG-UK guidance. NCI notes that tumor biomarker testing captures a snapshot and that biomarkers can change. A pipeline should therefore display context and evaluation timing instead of presenting biological relevance as a durable treatment conclusion. (jmg.bmj.com) (www.cancer.gov) (faq.oncokb.org)

For scientific artificial intelligence (**AI**) systems, provenance should be available to the model and to its reviewer. A generated evidence summary should distinguish an individual submission, a disease-specific aggregate, an allele mapping, and a therapeutic association. Clinical Interpretation of Variants in Cancer (CIViC) publication linkage and therapy fields, OncoKB's tumor-specific evidence model, and VRS's computable representation illustrate complementary source structures that an implementation can preserve. (civic.readthedocs.io) (faq.oncokb.org) (vrs.ga4gh.org)

Maintain a **quarterly schema and field-dictionary review** as a proposed local control, with earlier review whenever a release or dependency changes. The review should test field presence, controlled values, mapping behavior, and reference handling. It should also track guideline versions separately from data releases: SVIG-UK's page records version 1.1 approved in May 2026, while ACMG maintains its own guidance catalog. Neither is a substitute for validating the imported payload. (www.cangene-canvaruk.org) (www.acmg.net) (vrs.ga4gh.org)

Conclusion

ClinVar variant classification is useful for evidence pipelines when the system preserves the question each assertion answers. Germline classification, somatic clinical impact, and oncogenicity should remain separate from ingestion through export. Submitted, variant-condition, and variant-aggregate records should likewise retain their different scopes.

The proposed acceptance matrix assigns review status a defined role: it helps prioritize and describe evidence. It does not supply a probability, a disease match, or permission to make an unrelated research inference. A high-review assertion can still require context review. A lower-review submission can still be retained for discovery when its provenance and uncertainty remain visible.

Implementation should preserve the original review-status text, reference-aware allele identity, condition identifiers and mapping relations, source evidence, and versioned decision history. Rich assertion ingestion and scoped annotation exports can coexist, provided transformations do not discard the fields needed to reproduce acceptance decisions.

The immediate deliverable for a research team is a versioned policy, a source-linked field dictionary, a curator queue, and regression checks against pinned artifacts. The historical studies justify monitoring change, but they do not supply a universal local acceptance rate. That rate must be measured on a defined subset with explicit accessions, release identity, and intended use. The result is an auditable research workflow whose conclusions remain proportional to the evidence.

Source links

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

1. <https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig>
2. <https://www.nature.com/articles/gim201530>
3. <https://faq.oncokb.org>
4. https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status
5. <https://gnomad.broadinstitute.org/news/2023-09-clinvar-variants-filter-by-review-status>
6. <https://civic.readthedocs.io/en/latest/model/evidence/overview.html>
7. <https://hgvs-nomenclature.org/stable/background/refseq>
8. https://vrs.ga4gh.org/en/2.0/conventions/computed_identifiers.html
9. <https://link.springer.com/article/10.1186/s13073-026-01805-9>
10. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2825808>

11. https://www.amp.org/AMP/assets/File/pressreleases/2022/AMP_VITAL_Somatic_Press_Release12_13_22.pdf
12. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9081216>
13. <https://www.ga4gh.org/product/variation-representation-specification-vrs>
14. https://www.ncbi.nlm.nih.gov/clinvar/docs/release_cycle
15. <https://www.acmg.net/practice-resource/standards-and-guidelines-for-the-interpretation-of-sequence-variants-a-joint-consensus-recommendation-of-the-american-college-of-medical-genetics-and-genomics-and-the-association-for-molecul>
16. <https://jmd.amjpathol.org/article/S1525157816302239/pdf>
17. <https://clinicalgenome.org/docs/standards-for-the-classification-of-pathogenicity-of-somatic-variants-in-cancer-oncogenicity-joint-recommendations-of-clinical>
18. <https://mondo.monarchinitiative.org/pages/faq>
19. <https://www.ncbi.nlm.nih.gov/clinvar/docs/details>
20. <https://reg.genome.network>
21. <https://variantvalidator.org>
22. <https://www.ncbi.nlm.nih.gov/clinvar/docs/spreadsheet>
23. <https://www.ncbi.nlm.nih.gov/clinvar/nomenclature>
24. <https://www.cancer.gov/about-cancer/treatment/types/biomarker-testing-cancer-treatment>
25. <https://www.ncbi.nlm.nih.gov/clinvar/docs/downloads>
26. https://static.ensembl.org/info/docs/tools/vep/script/vep_options.html
27. https://ftp.ncbi.nlm.nih.gov/pub/clinvar/README_VCF.txt
28. https://ftp.ncbi.nlm.nih.gov/pub/clinvar/xsd_public
29. https://ftp.ncbi.nlm.nih.gov/pub/clinvar/xml/_README
30. <https://jmg.bmj.com/content/63/3/147>
31. <https://www.ncbi.nlm.nih.gov/clinvar/docs/identifiers>
32. <https://www.cangene-canvaruk.org/svg-uk-recommendations>
33. <https://onlinelibrary.wiley.com/doi/full/10.1002/humu.23637>

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.