ClinVar Data Dictionary

Overview

This document defines the data elements represented in the ClinVar database. It supports those using ClinVar's XML products as well as those submitting data to ClinVar. It includes descriptions of how data are managed by ClinVar, the XML used to represent each concept for submission (see https://ftp.ncbi.nlm.nih.gov/pub/clinvar/clinvar_submission.xsd), the field name in the full spreadsheet version of the submission document, and allowed values.

Three XML formats are documented here – XML for submission and two XML formats for reporting, ClinVarFullRelease and ClinVarVariationRelease. Each format is validated by a distinct XSD (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/xsd_public/).

ClinVarFullRelease represents the full ClinVar dataset aggregated by variant and condition; in other words, aggregated by RCV accession number. ClinVarVariationRelease is in a beta release. It is not comprehensive yet but it will represent the full ClinVar dataset aggregated by variant; in other words, aggregated by Variation ID.

ClinVarFullRelease and ClinVarVariationRelease differ by aggregation, but have a similar structure.

- One section of the XML represents data that are aggregated; it also includes data added by NCBI.
    - In ClinVarFullRelease this is the ReferenceClinVarAssertion section.
    - In ClinVarVariationRelease this is the IncludedRecord section, excluding ClinicalAssertionList.
    - Note that not all data are aggregated so users interested in details, particularly the details of the evidence, will need to look at the submitted section as well (see below).
- Another section of the XML represents data as they were submitted to ClinVar.
    - In ClinVarFullRelease this is each ClinVarAssertion section.
    - In ClinVarVariationRelease this is the ClinicalAssertionList.

The structure of each ClinVarAssertion in ClinVarFullRelease is almost identical to the structure of submission XML. Where they are the same, this document refers only to ClinVarFullRelease. Where they are different, Submission XML is also noted. For brevity in this document, ClinVarFullRelease is referred to as FullRelease XML; ClinVarVariationRelease is referred to as VariationRelease XML.

The database has a flexible data model, so submissions may be minimal or very detailed. All submissions include a variant-level interpretation but the evidence may be simple aggregate data or more complex and specific case-level data. Experimental observations generated by the submitter may also be provided as evidence.

Standard terms used by ClinVar (and coordinated with the NIH Genetic Testing Registry (GTR)), are summarized here: https://www.ncbi.nlm.nih.gov/clinvar/docs/authorities/

Status of this document

PublicRelease2. August 4, 2017. Please direct any comments to clinvar@ncbi.nlm.nih.gov

General processing

Data added computationally

Not all values included in this document are supplied from a submitter; some are added based on information in NCBI's databases. These values are marked explicitly as '**from NCBI**'.

Optional and required values

Some elements are hierarchical in representation. If a major category topic is optional, all data elements in that category are optional. But if an optional category is selected, then the data elements listed as required are required for that category.

Validation of submissions

Enforcing the rules presented in this document is not managed only via the XSD provided from our ftp site (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/xsd_submission). Some validation is provided at the database level, by comparison to standard terminologies or known relationships among variants, genes, and conditions, or by validating reported alleles against the stated reference sequence. Standard terms used by ClinVar and coordinated with the NIH Genetic Testing Registry (GTR), are summarized here: https://www.ncbi.nlm.nih.gov/clinvar/docs/authorities/

Conventions used in this document

The following elements are provided in the sections that detail each data element:

- definition (text)
- requirements (format, optional)
- Representation on submission forms (spreadsheet, XML)
- Representation in the relational component of the database

Data representations used in multiple contexts

Source/Status

Most elements in the database are characterized with respect to the source of the information, identifiers used by the submitter or other data sources, date submitted, date modified, status of the record (*e.g.* current/to be deleted/secondary to another record), review status, and whether the data should be public or private. Rather than repeating these elements for each data category defined below, the word **Source/Status** is used as a pointer to the Data source/Status section, where the source and status elements are defined.

AttributeSet

Many concepts in the database are represented by what ClinVar terms an AttributeSet, which is an open-ended structure providing the equivalent of a type of information, the value(s) for that data type, submitter(s), free text comment(s) describing that attribute, identifier(s) for that attribute, and citation(s) related to that attribute. Rather than repeating this description per attribute, the word **AttributeSet** is used to indicate that the data are

stored using this data structure, with the attribute types expected for that database concept. Thus by definition, an **AttributeSet** includes **Source/Status**.

Data source

ClinVar maintains attribution for each data element based on the description of the person and organization providing the information. In the database, these concepts are maintained by identifiers for the organization and identifiers for the individual.

FullRelease XML: XRef
VariationRelease XML: XRef
db: GTR.clinvar.attr_source.extrn_src (organization)
db: GTR.clinvar.attr_source.entered_by (individual)

Identifiers in public database records (optional)

The database cross-reference structure (XRef) is provided to represent pointers to identifiers in other databases for the same concept. For example, if a gene is being described, then XRefs can be provided to NCBI's Gene database, Ensembl, HGNC*, etc.*, including the database, the identifier, and URL.

Spreadsheet: multiple locations
Submission XML: XRef/@db, XRef/@id, XRef/@url
FullRelease XML: XRef/@DB, XRef/@ID, XRef/@URL
VariationRelease XML: XRef/@DB, XRef/@ID, XRef/@URL
db: GTR.clinvar.attr_source

Citations

Citations include published articles and URLs. If a database name and identifier are supplied, the full text is not required.

Type

NCBI classifies citations by type, based in part on curation but also on the publication type provided by PubMed. These types are reported in our public XML, but not expected to be provided by submitters. Values include general, review, practice guideline, Position Statement, Translational/Evidence-based, Suggested Reading, and Recommendation.

The spreadsheet has the following citation columns:

Column	Tab	Purpose
Clinical significance citations	Variant	Citations documenting the clinical significance (by ID)

Citations or URLs for clinical significance without database identifiers	Variant	Text or URL citations documenting the clinical significance
Evidence citations	CaseData	Citations documenting the evidence (by ID)
Citations or URLs for evidence without database identifiers	CaseData	Text or URL citations documenting the evidence
Method citations	CaseData	Citations documenting the method (by ID)
Citation	SubmissionInfo	Citations applicable to the complete submission set.

Spreadsheet: All citations default to citation type "general"
FullRelease XML: /Citation @Type
VariationRelease XML: /Citation @Type
db: GTR.dbo.citation.citation_type

Source

The name of the data service providing an identifier for a citation. This value should not be completed if the citation is a URL or free text. Current options include PubMed, PubMedCentral, DOI, NCBI bookshelf

Spreadsheet: Citation (Source default = PubMed)
FullRelease XML: Citation/ID@Source
VariationRelease XML: Citation/ID@Source
db: GTR.dbo.citation.extrn.src

ID: the identifier provided by that data source for a citation. This value should not be completed if the citation is a URL or a free text.

Spreadsheet: Citation
FullRelease XML: Citation/ID
VariationRelease XML: Citation/ID
db: GTR.dbo.citation.extrn_id

URL: complete URL

Spreadsheet: Citations or URLs without database identifiers
FullRelease XML: Citation/URL

VariationRelease XML: Citation/URL
db: GTR.dbo.citation.url

CitationText: This should be used only when there is no database ID for the publication and no URL.

Spreadsheet: Citations or URLs without database identifiers
FullRelease XML: Citation/CitationText
VariationRelease XML: Citation/CitationText
db: GTR.dbo.citation.citation

Comments

A free text comment can be provided to describe submitted data. Line breaks are retained, but no other formatting.

Spreadsheet: multiple locations (comment or private comment)
FullRelease XML: Comment/CommentText
VariationRelease XML: Comment/CommentText
db: GTR.dbo.comment.comment

Type: (required) public (is rendered on the web) or private (for internal use; to explain a submission and be stored in the database but not rendered on the web). There are subcategories of public comments, differentiated by type.

Spreadsheet: multiple locations (comment or private comment)
FullRelease XML: Comment/Type
VariationRelease XML: Comment/Type
db: GTR.dbo.comment.comment_type

Information describing the submitter and the submission

Note that information describing the submitter and the submission is provided only via the submission spreadsheet for organizations that submit by ftp. For all other submitters, this information is provided in the Submission Portal.

Identification of the submitter

Information in this section applies only to submission XML. An abbreviated version of submitter information is found in ClinVar's public XML products. Note that in ClinVar, "submitter" refers to the submitting organization, not an individual person. ClinVar submissions are attributed to an organization. That organization may have one or more people associated with it, but the people are not directly connected to the submission, only through the organization.

ClinVar uses multiple methods to identify a submitter. One type is the person to contact, listed as 'SubmitterOfRecord' in the XML. The other is the optional official submitter, if the official submitter is different from the contact.

Submitter of record (required)/other official submitters (optional)

Submitter name (choice with SubmitterID/SubmitterIDType)

The first and last name of the person submitting this batch; this includes second party submitters who submit on behalf of other organizations. If you know your identifier in dbSNP (submitter handle) or GTR/ClinVar (Person ID), you can identify the submitter of record by those identifiers.

Spreadsheet: SubmissionInfo.Submitter first name
Spreadsheet:SubmissionInfo.Submitter last name
Submission XML: //Person/Name/First
Submission XML: //Person/Name/Last
db: GTR.dbo.person

Submitter Identifier

NCBI maintains several identifier systems for submitters (*e.g.* the dbSNP submitter handle), and there may be public identifier systems as well. Thus the identifier for a submitter is managed as a database cross-reference. The submitter handle is treated explicitly in the database.

Spreadsheet: SubmissionInfo.SubmitterID
Spreadsheet: SubmissionInfo.SubmitterIDType
Submission XML: SubmitterOfRecord/Person/SubmitterHandle
Submission XML: Submitter.Personnel.PersonRef/@id and @db where db=snp
db: GTR.dbo.person.submitter_handle
db: GTR.dbo.person.submitter_id

Private contact information (required)

The contact information of the person actually making the submission. This is required, but not publicly displayed; it is used by ClinVar staff for contact regarding submissions only.

Spreadsheet: SubmissionInfo.Submitter type
Spreadsheet: SubmissionInfo.Submitter email
Spreadsheet: SubmissionInfo.Submitter phone
Submission XML: Submitter/Personnel/PrivateContact/email
Submission XML: Submitter/Personnel/PrivateContact/phone
Submission XML: Submitter/Personnel/PrivateContact/fax
db: GTR.dbo.contact (identified as private in GTR.dbo.org_person)

Public Contact information (optional, publicly displayed)

The contact information of the person actually making the submission. This is required and publicly displayed; it is also used by ClinVar staff for contact regarding submissions.

Spreadsheet: SubmissionInfo.Submitter type
Spreadsheet: SubmissionInfo.Submitter email
Spreadsheet: SubmissionInfo.Submitter phone
Spreadsheet: SubmissionInfo.Submitter role
Submission XML: Submitter/Personnel/PublicContact/email
Submission XML:S ubmitter/Personnel/PublicContact/phone
Submission XML: Submitter/Personnel/PublicContact/fax
Submission XML: Submitter/Personnel/Title
db: GTR.dbo.contact (identified as public in GTR.dbo.org_person)

Organization (required)

The organization responsible for the submission

Spreadsheet: SubmissionInfo.Organization type
Spreadsheet: SubmissionInfo.Organization
Spreadsheet: SubmissionInfo.OrganizationID
Spreadsheet: SubmissionInfo.Organization abbreviation
Spreadsheet: SubmissionInfo.Institution
Spreadsheet: SubmissionInfo.Street
Spreadsheet: SubmissionInfo.City
Spreadsheet: SubmissionInfo.State/Province
Spreadsheet: SubmissionInfo.Country
Spreadsheet: SubmissionInfo.Postal code
Submission XML: SubmitterOfRecord/Organization/Name
Submission XML: SubmitterOfRecord/Organization/NCBIOrganizationID
Submission XML: Submitter/Organization/NameAcronym
Submission XML: Submitter/Organization/Institution
Submission XML: Submitter/Organization/StreetAddress/Line1
Submission XML: Submitter/Organization/StreetAddress/City
Submission XML: Submitter/Organization/StreetAddress/State
Submission XML: Submitter.Organization/StreetAddress/PostCode
db: GTR.dbo.contact (identified as private in GTR.dbo.org_person)

URLs for the submitter

Organization URL

- optional

The URL for your organization's website/homepage.

Spreadsheet: SubmissionInfo.Organization URL
Submission XML: Organization/URL
db: organization.url

Base URL

- optional

For submitters who maintain a website with variant-specific pages whose URLs can be constructed with a base URL + Local ID.

Spreadsheet: SubmissionInfo.Base URL
Submission XML:
db: external_source.base_url

Submissions from multiple submitting organizations

- required if the submission is from more than one organization

A submission may be attributed to more than one organization, e.g. collaborators on a publication. One organization must be designated the primary submitter; all others are designated additional submitters.

Spreadsheet: SubmissionInfo.Primary organization
Submission XML: OrganizationType
FullRelease XML: ClinVarAssertion/ClinVarAccession/@OrgID
FullRelease XML: ClinVarAssertion/AdditionalSubmitters
VariationRelease XML: ClinicalAssertion/@SubmitterName, @OrgID
VariationRelease XML: ClinicalAssertion/AdditionalSubmitters

Descriptors of the submission

Date submitted

- required

This may be provided explicitly. If not provided, then the date a submission is received is used as the submission date. If the submission is an update of an existing record, the submission date is the date of record of a new version of the submission.
Spreadsheet: SubmitterInfo.Submission date
FullRelease XML: ClinvarSubmissionSet/@Date
VariationRelease XML: ClinicalAssertionList/ClinicalAssertion/@SubmissionDate
db: GTR.clinvar.measure_target.subdate

Review status

- optional

Review status indicates the level of confidence in any assertion. The review status reflects whether an interpretation is available; whether documentation of the assertion criteria provides transparency into the classification process; and whether there is consensus. The allowed values also depend on the type of record. Review status is represented graphically on the web site with gold stars, but the number of stars cannot indicate unambiguously the review status for all cases. Values are:

Review status	Type of record	Description	Number of gold stars
no assertion provided	SCV, RCV, VCV	The allele was included in a submission that did not provide an interpretation.	none
no assertion criteria provided	SCV, RCV, VCV	The allele was included in a submission with an interpretation but without assertion criteria.	none
no assertion for the individual variant	VCV	The allele was not interpreted directly in any submission; it was submitted to ClinVar only as a component of a compound heterozygote or a haplotype.	none
criteria provided, single submitter	SCV, RCV, VCV	One submitter provided an interpretation with assertion criteria.	one
criteria provided, conflicting interpretations	RCV, VCV	Multiple submitters provided assertion criteria but there are conflicting interpretations. The independent values are enumerated for clinical significance.	one
criteria provided, multiple submitters, no conflicts	RCV, VCV	Two or more submitters with assertion criteria provided the same interpretation.	two
reviewed by expert panel	SCV, RCV, VCV	The variant was reviewed by an expert panel.	three
Practice guideline	SCV, RCV, VCV	The variant was reviewed by a professional society that provides practice guidelines.	four

Spreadsheet: SubmissionInfo.Review status
FullRelease XML: MeasureTrait/ClinicalSignificance/ReviewStatus
VariationArchive XML: InterpretedRecord/ReviewStatus
VariationArchive XML: ClinicalAssertionList/ClinicalAssertion/ReviewStatus
db: GTR.clinvar.measure_target.rev_stat

Content note:

A submitter may not self-identify as an expert panel. Please review this document for the application form: https://www.ncbi.nlm.nih.gov/clinvar/docs/expert_panel/

Release status

- required

This field supports a temporary hold on the submission being presented publicly, usually while a manuscript is in the publication process. Allowed values are public or hold until published. If not supplied, public is the default.

Spreadsheet: SubmissionInfo.Release status
FullRelease XML: not represented
VariationRelease XML: not represented
db: GTR.clinvar.measure_target.pubstat (record level submission)
db: GTR.clinvar.mset.pubstat

Study name

- optional

The public name of a study submitting these data and providing the sample. Used for a large-scale study funded by a granting agency for this study name. Can be used to indicate the name of a study population or cohort. Examples: Framingham, CSER, CMG.

Spreadsheet: SubmissionInfo.Study name
FullRelease XML: ClinvarSubmissionSet/StudyName
VariationRelease XML: ClinicalAssertion/StudyName
db: GTR.clinvar.mt_set_attr where attr_type = 731

Study description

- optional

Description of the study generating the submission. Study description is only provided if there is a study name.

Spreadsheet: SubmissionInfo.Study description
FullRelease XML: ClinvarSubmissionSet/Comment
VariationRelease XML: ClinicalAssertion/StudyDescription
db: GTR.dbo.comment

Submission name

- optional

Use this field to identify your submission. If you supply one, we index that value so that the records from this submission can be retrieved from the ClinVar interface in batch by that name.

Spreadsheet: SubmissionInfo.Submission name
FullRelease XML: ClinVarAssertion/@SubmissionName
VariationRelease XML: ClinicalsAssertion/@SubmissionName
db: GTR.clinvar.mt_set.set_key

Record

Accessions and versions

SCV

The accession for each submitted assertion. This is provided by NCBI, and it is required in a submission that is an update.

Spreadsheet: Variant.ClinVarAccession
Submission XML: ClinvarSubmissionAcc /@Acc
Submission XML: ClinvarSubmissionAcc/ClinvarSubmissionAccType/[@val_type=name] SCV
FullRelease XML: ClinVarAssertion/ClinVarAccession/@Acc
FullRelease XML: ClinVarAssertion/ClinVarAccession/@Version
VariationRelease XML: ClinicalAssertionList/ClinicalAssertion/ClinVarAccession/@Acc
VariationRelease XML: ClinicalAssertionList/ClinicalAssertion/ClinVarAccession/@Version
db: GTR.clinvar.measure_target.accession

RCV

The accession for the aggregate record for a variant-condition pair. This accession is **not** included in submissions.

Spreadsheet: not represented in spreadsheet
FullRelease XML: ReferenceClinVarAssertion/ClinVarAccession /@Acc
FullRelease XML: ReferenceClinVarAssertion/ClinVarAccession /@Version
VariationRelease XML: InterpretedRecord/RCVList/ClinVarAccession/@Acc
VariationRelease XML: InterpretedRecord/RCVList/ClinVarAccession/@Version
db: GTR.clinvar.measure_target.accession

VCV

The accession for the aggregate record for a variant (or set of variants). This accession is **not** included in submissions.

Spreadsheet: not represented in spreadsheet
FullRelease XML: not represented until production release of ClinVarVariationRelease
VariationRelease XML: VariationArchive/@Acc
VariationRelease XML: VariationArchive/@Version
db: GTR.clinvar.measure_target.accession

Record Status

- optional

If ClinVar accessions are included in a submission, record status indicates if the submission is novel and accessions were reserved prior to submission, or an update to existing SCV records, or to delete an existing SCV record.

Spreadsheet: Variant.Novel or Update
Spreadsheet: Deletes.ClinVarAccession
FullRelease XML: ClinvarSubmission/RecordStatus
FullRelease XML: ClinvarSubmissionAcc/ClinvarSubmissionAccType/[@val_type="name"] RCV
VariationRelease XML: VariationArchive/RecordStatus
db: GTR.clinvar.measure_target.accession

Replaces ClinVarAccessions

- optional

For updates in which one or more ClinVar submitted records (SCVs) are being merged into another existing ClinVar submitted record.

Spreadsheet: Variant.Replaces ClinVarAccessions
FullRelease XML: ClinvarSubmission/ReplacesAccession
VariationRelease XML: not represented yet
db: GTR.dbo.id_hist

Descriptors of the record

Date created

Date the record was created in ClinVar.

Spreadsheet: not represented
FullRelease XML: ReferenceClinVarAssertion/@DateCreated
VariationRelease XML: VariationArchive/@DateCreated

Date last updated

Date the record was last updated in ClinVar. Note that the record may be updated when underlying submissions are updated and also when NCBI updates data, such as HGVS expressions.

Spreadsheet: not represented
FullRelease XML: ReferenceClinVarAssertion/@DateLastUpdated
VariationRelease XML: VariationArchive/@DateLastUpdated

InterpretedRecord vs IncludedRecord

VariationFullRelease comprises two types of VCV records, interpreted and included records. An interpreted record represents a variant or set of variants for which there is an explicit submission in ClinVar. An included record

represents a variant or set of variants for which there is only an interpretation for that variant or set of variants in combination with other variants. For example, a SNP may have no explicit interpretation for the SNP alone, but ClinVar may have an interpretation for a haplotype that includes the SNP. For that case, the individual SNP is represented by an included record; the haplotype is represented by an interpreted record.

Spreadsheet: not represented
FullRelease XML: not represented
VariationRelease XML: VariationArchive/InterpretedRecord
VariationRelease XML: VariationArchive/IncludedRecord

Interpreted Condition

Information about the interpreted condition (disease or phenotype) is represented by the combination of type of term, the term value, and the relationship of that term to other terms in the submission. The condition information must be connected at the variant level, but can also be represented as part of each set of observations.

When the condition is defined as part of the condition-variant relationship, that condition should be the one about which the variant interpretation is being made. For phenotypes *observed* in one or more patients, those data should be provided as Clinical features. The relationship between the observed phenotypes and the interpreted variant may be unknown.

Please note that ClinVar retains submitted terms, but maps these to controlled terms whenever possible. To facilitate that mapping, we encourage submitters to provide their condition descriptions as the combination of a database name and a database identifier, rather than free text.

ClinVar requires categorization of condition information. This may be done at the level of a set of conditions, a relationship between conditions, or a single condition. If the combination of term and type provided in the submission is not consistent with ClinVar's representation, ClinVar may re-assign the condition category. Current options for condition categories are:

Controlled term	Usage	Data elements
Disease	Use for a diagnostic name.	TraitSetType TraitRelationship TraitType
Drug response	Usually written as drug name + response. This includes pharmacodynamic and pharmacokinetic differences.	TraitSetType TraitRelationship TraitType
Subphenotype	Use to submit a disease hierarchy.	TraitRelationship TraitType

Blood group	For the name of a blood group system. If an allele of a blood group is manifested as an additional phenotype, include in the trait set.	TraitRelationship TraitType
Finding	Use for clinical features or phenotypic measures.	TraitRelationship TraitType
Infection resistance	Corresponds to "genetic resistance to infectious agent", **IDO_0000587**	TraitRelationship TraitType

Description of one condition (trait)

Names

Preferred name

The name of the condition used for reporting from ClinVar by default.

When available, this is a preferred term from SNOMED CT. Other sources may include Office of Rare Diseases Research (ORDR), Human Phenotype Ontology (HPO), OMIM®, and MeSH. The name for the condition that the submitter provides is retained, but is mapped to controlled vocabularies when possible. Because testing laboratories may know only the name of the ordered test (e.g. "deafness"), the indication for testing or the test name may be provided (see below), with condition submitted as "not provided". A list of disorder names used by ClinVar/GTR is provided from ClinVar's ftp site in the file disease_names.

When the condition is a drug response, the condition name should be constructed as the drug name + response.

Spreadsheet: Variant.Condition ID type, Condition ID value
Spreadsheet: Variant.Preferred condition name
Spreadsheet: Variant.Condition descriptions
FullRelease XML: MeasureTrait/TraitSet/Trait/Name/ElementValueType/[@val_type = "name"] Preferred
FullRelease XML: MeasureTrait/TraitSet/Trait/Name/ElementValue
VariationRelease XML: ClinicalAssertion/TraitSet/Trait/Name
db: GTR.clinvar.target_attr where attr_type = 17 (AttributeSet)

Content note: When an allele is asserted to be globally benign

When making a clinical assertion of "benign", the condition can either be listed as benign relative to a specific condition, or as globally benign for all highly penetrant Mendelian diseases which are not specified by name. For the latter, please submit the name of the condition as "not specified".

Alternate name(s)

- Optional, multiple allowed

Other names used for this condition. These are added to the set of search terms used in ClinVar, MedGen, and GTR.

Spreadsheet: Variant.Condition description
Spreadsheet: Variant.Preferred condition name
FullRelease XML: Trait/Name/ElementValueType/[@val_type = “name”] Alternate
FullRelease XML: Trait/Name/ElementValue
VariationRelease XML: InterpretedRecord/TraitSet/Trait/Name
db: GTR.clinvar.target_attr where attr_type = 18 (AttributeSet)

Preferred symbol

- Optional, only one allowed.

The preferred symbol for the condition. The final value used for display may be recalculated by NCBI.
Spreadsheet: not represented
FullRelease XML: Trait/Symbol/ElementValueType/[@val_type = “name”] Preferred
FullRelease XML: Trait/Symbol/ElementValue
VariationRelease XML: InterpretedRecord/TraitSet/Trait/Symbol/ElementValue/[@val_type = “name”] Preferred
VariationRelease XML: Interpreted Record/Trait/Symbol/ElementValue
db: GTR.clinvar.target_attr where attr_type = 19

Alternate symbols(s)

- Optional, multiple allowed.

Alternate symbols for the condition.

Spreadsheet: not represented
FullRelease XML: Trait/Symbol/ElementValueType/[@val_type = “name”] Alternate
FullRelease XML: Trait/Symbol/ElementValue
VariationRelease XML: InterpretedRecord/TraitSet/Trait/Symbol/ElementValue/[@val_type = “name”] Alternate
VariationRelease XML: Interpreted Record/Trait/Symbol/ElementValue
db: GTR.clinvar.target_attr where attr_type = 20

Attributes

These are based on the AttributeSet structure, and thus can be used to capture values assigned to defined information categories, along with supporting documentation. The values can be words, integers, decimals, and/or dates. Types are restricted by an enumerated list of allowed values per major information set. These restrictions may be applied in the XSD, or only in the underlying relational database. If you wish to suggest a new attribute, please contact us at clinvar@ncbi.nlm.nih.gov.

- Optional, multiple allowed

Concept	attr_type	XML	column in spreadsheet

Usual age of onset	257	MeasureTrait/AttributeSet/MeasureTraitAttribute Type/[@val_type = “name”] AgeOfOnset	not represented
Reported penetrance	353	MeasureTrait/AttributeSet/MeasureTraitAttribute Type/[@val_type = “name”] Penetrance	not represented
Mode of inheritance*	162	MeasureTrait/AttributeSet/MeasureTraitAttribute Type/[@val_type = “name”] ModeOfInheritance	Variant.Mode of inheritance
Severity **	468,740	MeasureTrait/Severity MeasureTrait/ObservedIn/ObservedData/Severity	not represented
Activity level **	258	MeasureTrait/ObservedIn/ObservedData/[ObsAttributeType= “ActivityLevel”][Attribute]	not represented

* **NOTE**: Mode of inheritance is stored as an attribute of the variant/condition relationship (i.e. the MeasureTrait). Mode of inheritance of a condition as provided by authoritative sources is available in MedGen.

*** NOTE:** Severity and activity level, which may qualify the condition, can be submitted either at the level of the relationship between a set of conditions and a set of variations (*e.g*. for this variation, the condition is severe or the activity level of the product is decreased), or at the level of the observations themselves (*e.g*. for this sample, the condition is severe or the activity level of the product is decreased).

Category/Type of condition term

Category of the condition term. Only one classification is allowed per condition identifier. If the submission provides the name of a trait with a type different from ClinVar’s categorization, ClinVar retains what is submitted but continues to report the ClinVar categorization and reviews the discrepancy with the submitter.

The options and usage are tabulated at the beginning of the Condition section.

Spreadsheet: Variant.Condition category
FullRelease XML: TraitSet/Trait/TraitType
VariationRelease XML: TraitSet/Trait/@Type
db: GTR.clinvar.target.id_type

Sets of conditions

Category/type of condition set

More than one term may be used to describe condition, e.g. a set of clinical findings in addition to a diagnosis; a set of clinical findings in the absence of a named disease; or multiple disease terms if it is the combination of diseases about which an assertion is being made. Each condition term is captured explicitly [see the next section, Description of one condition (trait)], and the co-occurring conditions are represented as a set. Options for type of condition sets are Disease, Drug Response, and Finding.

Content note (should condition names be submitted together?)

To report that a variant is pathogenic for disease1 and also pathogenic for disease2, submit the data on multiple rows, with one row specific to the variant-disease1 combination and one row specific to variant-disease2. If, on the other hand, to report that a variant is pathogenic only in the context of the combination of disease1+disease2, submit on the same row in the spreadsheet or in the same TraitSet in the XML.

To report that a variant is pathogenic for either disease1 or disease2, but you are not sure which is correct, submit the data on one row and use the Condition uncertainty column, or TraitSetType = TraitChoice in XML. This case is expected to be very rare.

Content note (no diagnostic term, but observed features)

When the name of a condition is not known, but multiple clinical features are being reported at the case level, please submit Preferred condition name as 'see cases'. 'See cases' will then indicate to users of the data that a diagnosis has not been made, but the clinical features are provided. This case is used for structural variants where the relationship between one or more variants and one or more phenotypes is unclear.

Spreadsheet: Variant.Condition category
FullRelease XML: TraitSet/TraitSetType[@val_type="name"]
VariationRelease XML: TraitSet/@Type
db: GTR.clinvar.tset.type

Relationships among members of a set of conditions

The relationship among a set of conditions may be described. Some examples:

- a set of a clinical features and a diagnostic name may be represented as several Findings and a Disease.
- sickle cell anemia and resistance to malaria may be represented as Disease and Infection resistance.
- subphenotype may be used for hierarchical relationships, for example Usher Syndrome, type 1B may be represented as a subphenotype of Usher Syndrome, type 1. The ClinVar/GTR staff curates some hierarchical relationships, but usually uses those provided by external authorities. If you wish to suggest a revision of current hierarchies, or suggest new ones, please contact us. Note that severity should be represented as an attribute of the phenotype (see below), rather than as a subphenotype.

Spreadsheet: not represented
FullRelease XML: Trait/TraitRelationship/TraitRelationshipType/@val_type = "name"
VariationRelease XML: Trait/TraitRelationship/@Type
VariationRelease XML: Trait/TraitRelationship/Name
db: GTR.clinvar.tsubset.relat_type

Comment about condition

- Optional.

As with most other data elements, a free text comment about the set of interpreted conditions may be submitted. This comment is used to provide additional information about the interpreted condition and should apply to any use of this set of conditions, not information that is specific to a submission or an individual. For example, it may be used to describe a new syndrome only defined by a set of clinical features or a rare form of a known disease.

Spreadsheet: Condition comment
FullRelease XML: MeasureTrait/TraitSet/Comment
VariationRelease XML: ClinicalAssertion/TraitSet/Comment

Indication for testing

- Optional.

Testing labs may only know the indication for testing, not the actual condition of the tested individual. Consistent with UMLS, ClinVar treats these as Findings.

Spreadsheet: Variant.Indication
Spreadsheet: CaseData.Indication
FullRelease XML: ObservedIn/Sample/Indication
VariationRelease XML: ObservedIn/Sample/Indication
db: GTR.clinvar.version (as XML)
db: GTR.clinvar.target_attr where attr_type = 17 (AttributeSet)

Variant allele(s)

ClinVar maintains information about sequence changes by representing location on an explicit reference sequence and the nucleotide or amino acid observed at that location. The allele may be the same as reference, for example polymorphic sites in which the reference sequence matches the allele about which information is being submitted. Sequence changes may be single or multiple. Because it is possible to submit information about multiple sequence changes (e.g. a haplotype or a genotype) with relationship to condition, variants are submitted as a set (MeasureSet or GenotypeSet), even if the size of the set is one. Each individual variant within the set is submitted as a Measure. The same representation is used in ClinVarFullRelease. In ClinVarVariationRelease, each individual variant is represented as SimpleAllele, and combinations of variants are represented as Haplotype or Genotype.

Content notes

Names (XML)

Please submit official designations of alleles (e.g. CYP2D6*2) as Name of type=”Preferred”.

HGVS expressions (XML)

Please submit HGVS expressions as attributes of Type=”HGVS”. ClinVar extracts each expression submitted as an HGVS name, validates it (can that allele be identified on the referenced sequence), compute other HGVS expressions, and returns all values for public display. If a gene is known to have multiple splice variants, or legacy numbering systems, or more than one nucleotide change resulting in the same protein change, it is critical that the HGVS value contain the reference nucleotide sequence, the version of the reference sequence, the location, and the change. Submissions with insufficient specificity are returned for review or reported as non-validated. Submissions go more smoothly if HGVS expressions are validated by tools such as Mutalyzer (https://mutalyzer.nl/) in advance.

Each variant allele

Allele ID

The identifier for each individual variant, which may be interpreted as a single variant or as part of a haplotype or genotype.

Spreadsheet: not represented
FullRelease XML: MeasureTrait/MeasureSet/Measure/@ID
VariationRelease XML: SimpleAllele/@AlleleID

Location

Each allele needs to be described unambiguously as the location of the variant and the sequence at that location. There are multiple options to specify the location of a variant: cytogenetic, chromosome location, or nucleotide or protein change as an HGVS expression.

To permit unambiguous mapping to the genome, a submission in nucleotide coordinates as accession.version+location is highly preferred. If an LRG sequence is used, the version is not applicable. If the description of the variation is provided via an HGVS expression which includes the explicit reference sequence and its version, then chromosome location need not be reported as a separate value.

Cytogenetic location

This is required for large structural variations defined only cytogenetically.
For variations defined by sequence, cytogenetic location is optional; it is computed by NCBI.
Spreadsheet: Variant.Chromosome
FullRelease XML: Measure/CytogeneticLocation
VariationRelease XML: SimpleAllele/Location/CytogeneticLocation
db: GTR.clinvar.seq_loc.cytogenetic + GTR.clinvar.seq_loc.chr

Chromosome Location

The location of the variant defined by assembly, chromosome, and location. The location may be a point or a range, with or without defined end points. If a point, only start needs to be provided, and ClinVar computes stop based on the value reported as start. For variants without exact locations defined, multiple values are provided to represent the boundaries of what is known (e.g. outer start and outer stop, inner start and inner stop). These are defined as documented here: https://www.ncbi.nlm.nih.gov/dbvar/content/overview/

Spreadsheet: SubmissionInfo.Assembly
Spreadsheet: Variant.Start/Stop/ReferenceAllele/AlternateAllele, Outer start, Outer stop, etc.
FullRelease XML: Measure/SequenceLocation with multiple attributes to define the assembly, sequence, and position/boundaries of the variation's location
VariationRelease XML: SimpleAllele/Location/SequenceLocation with multiple attributes to define the assembly, sequence, and position/boundaries of the variation's location

db: GTR.clinvar.seq_loc (multiple columns)

Nucleotide change as HGVS

The nucleotide change for a variation represented as an HGVS expression. For more details about how NCBI and ClinVar manage HGVS expressions, please see https://www.ncbi.nlm.nih.gov/clinvar/docs/hgvs_types.

The Reference sequence must include accession and version, such as NM_000492.3, NG_016465.3, NC_000007.13. The Variation name is the c., g., m., n. or r. portion of the full HGVS expression and it must be in agreement with the reference sequence type.

Spreadsheet: Variant.Reference sequence and Variant.HGVS
FullRelease XML: Measure/AttributeSet/Attribute/MeasureAttributeType = 'HGVS'
FullRelease XML: Measure/AttributeSet/Attribute/
VariationRelease XML: SimpleAllele/AttributeSet/Attribute/@Type='HGVS'
VariationRelease XML: SimpleAllele/AttributeSet/Attribute
db: GTR.clinvar.measure_attr

Protein change (HGVS, single letter or 3 letter amino acid abbreviations)

- optional

Although submitting the definition of a variation *only* in protein coordinates is accepted, this format is not recommended. It is our goal to map sequence variation to the genome, and protein coordinates are not always sufficient. That said, submission of a variation as both the nucleotide change and protein change is desirable, to support confirmation of location.

Spreadsheet: Variant.Alternate designations
FullRelease XML: Measure/AttributeSet/Attribute/@Type="HGVS"
VariationRelease XML: SimpleAllele/AttributeSet/Attribute/@Type='HGVS'
VariationRelease XML: SimpleAllele/AttributeSet/Attribute
db: GTR.clinvar.measure_attr

Nucleotide change for structural/cytogenetic variants

Spreadsheet: Variant.Variant type
FullRelease XML: Measure/MeasureType
VariationRelease XML: SimpleAllele/VariantType

Variant length

- required for structural variants if outer start/stop if provided but inner start/stop is not provided.

This value is generally calculated by NCBI but should be submitted for the case above.

Spreadsheet: Variant.Variant length
FullRelease XML: Measure/SequenceLocation/@variantLength
VariationRelease XML: SimpleAllele/Location/SequenceLocation/@variantLength

Copy number

- required for structural variants described as copy number gain or loss

The observed copy number of the variant region.

Spreadsheet: Variant.Copy number
FullRelease XML: Measure/AttributeSet/ MeasureAttributeType = AbsoluteCopyNumber
FullRelease XML: Measure/AttributeSet/Attribute/
VariationRelease XML: ClinicalAssertion/SimpleAllele/AttributeSet/Attribute/@Type="AbsoluteCopyNumber"

Reference copy number

- required for structural variants described as copy number gain or loss

The copy number of the variant region in the reference genome; i.e. the expected copy number.
Spreadsheet: Variant.Reference copy number
FullRelease XML: Measure/AttributeSet/ MeasureAttributeType = ReferenceCopyNumber
FullRelease XML: Measure/AttributeSet/Attribute/
VariationRelease XML: ClinicalAssertion/SimpleAllele/AttributeSet/Attribute/@Type="ReferenceCopyNumber"

Official variant name

- optional

This must be an official allele name. For variants that are assigned official allele names, e.g. CYP3A4*18 for one of the cytochrome P450 gene CYP3A4; or HLA-DRA*0102 for the MHC gene HLA-DRA.
Spreadsheet: Variant.Official allele name
FullRelease XML: Measure/Name@type=preferred
VariationRelease XML: SimpleAllele/Name
db: GTR.clinvar.measure_attr.attr_char where attr_type = 17 (AttributeSet)

Alternate names

- optional

Other names in common use for an allele, including legacy names with alternate numbering systems.
Spreadsheet: Variant.Alternate designations
FullRelease XML: Measure/Name/ElementValueType=alternate
FullRelease XML: Measure/Name/ElementValue
VariationRelease XML: SimpleAllele/OtherNameList
db: GTR.clinvar.measure_attr.attr_char where attr_type = 18 (AttributeSet)

Identifiers in public databases

- optional

Identifiers in dbSNP/dbVar/OMIM, locus-specific databases, *etc.* Special handling is provided for identifiers generated by NCBI, namely rs#, nsv, nssv, in that they have dedicated attribute types and are stored in measure_attr. Other non-NCBI public identifiers are stored in attr_source. At times, a submission may include information that the location of a variation can be identified by an rs# or an nsv# or some other public identifier.

Spreadsheet: Variant. Variation identifiers
FullRelease XML: Measure.AttributeSet.Attribute.rsNumber
FullRelease XML: Measure.AttributeSet.Attribute.nsv
Measure/XRef/@DB
db: GTR.clinvar.attr_source
or
db: GTR.clinvar.measure_attr

Location relative to a gene, protein, or other genomic location

- optional

Some of these values are based on sequence ontology terms and computed per transcript. Content can be computed by NCBI and/or provided by submitter. This category includes exon and intron numbers, position relative to splicing or regulatory regions, position in conserved protein domains, *etc*. The sequence ontology terms used by NCBI include:

- UTR (SO:0000203)
    - 5_prime_UTR (SO:0000204)
    - 3_prime_UTR (SO:0000205)
- Upstream location
    - Upstream variant (SO:0001631)
        - Within 5kb (SO:0001635)
        - Within 2kb (SO:0001636)
- Downstream location
    - downstream_gene_variant (SO:0001632)
        - 5KB_downstream_variant (SO:0001633)
        - 500B_downstream_variant (SO:0001634)
- Splice site
    - splice_site (SO:0000162)
- Distance from nearer splice junction
  (can be calculated if not provided)
- Regulatory site (yes/no or name of promoter/locus control region)
    - Promoter: SO:0000167

Intron or exon number

- optional

Submitters may provide an intron or exon designation and Arabic numeral (*e.g.*, exon 4, intron 3, not IVS 3 or Exon IV). The sequence used to define the numbering system for the intron/exon organization must also be included, e.g. an NM RefSeq.

Spreadsheet: Variant.Location
FullRelease XML: Measure/AttributeSet/Attribute Type='Location'
db: GTR.clinvar.measure_attr where attr_type = 472

Region name (active site, conserved domain, unspecified, etc)

- optional

Submitters may provide a domain name in which the variation is found. NCBI will also report when the variation lies within a known domain.

Spreadsheet: Not represented
FullRelease XML: Measure/AttributeSet/Attribute@type='Domain'
db: GTR.clinvar.mset_attr where attr_type = 473

Total exons in transcript

This optional concept is included in the dictionary because the value may be included in our public displays. The data are not to be submitted however, and will be provided based on the sequence used to define any gene annotation.

Other regions with similar sequence which may confound interpretation

Submitters may describe other regions in the genome with sequence highly similar to the context of the reported variant, and which may affect variation calls. This attribute may describe a gene or a variant.

Spreadsheet: Not represented
FullRelease XML: MeasureSet/AttributeSet/Attribute@type='RelatedSequence'
db: GTR.clinvar.mset_attr

Molecular consequence

- optional

Molecular consequence is reported from Sequence Ontology terms, and, when possible, are computed per transcript by NCBI. These terms are in this group because they can be calculated explicitly from the type and location of the variation, unlike the functional consequence which must be established experimentally (or predicted).

Spreadsheet: Not represented
FullRelease XML: Measure/AttributeSet/Attribute type='MolecularConsequence'
VariationRelease XML: SimpleAllele/MolecularConsequenceList/MolecularConsequence
db. GTR.clinvar.measure_attr

Comment about molecular consequence

A free text comment may be submitted about the molecular consequence. The comment structure should be used if the consequence being submitted is not defined by the Sequence Ontology group. We strongly recommend, however, that an SO term be requested if current terms are insufficient.

Spreadsheet: not represented
FullRelease XML: Measure/AttributeSet/Attribute Type='MolecularConsequence'
Measure/AttributeSet/Comment/CommentText
VariationRelease XML: SimpleAllele/MolecularConsequenceList/MolecularConsequence/Comment

Functional consequence

- optional

These attributes are provided by the submitter since they require determination of the consequences of the molecular change. Each is qualified by whether the submitter predicted the consequence or established it experimentally. Options include terms used by the LOVD databases:

- affects function
- probably affects function
- probably does not affect function
- does not affect function
- unknown

and terms established by VariO (variationontology.org) and Sequence Ontology (SO, www.sequenceontology.org/browser/obo.cgi). These terms include but are not limited to:

- effect on protein activity
- exon loss
- sequence_variant_affecting_splicing
- effect on RNA splicing
- protein truncation
- protein loss of function
- protein gain of function
- effect on RNA abundance
- cryptic splice acceptor activation
- cryptic splice donor activation
- Variation affecting splicing function of RNA
- RNA degradation by nonsense-mediated decay

Spreadsheet: Variant.Functional consequence
FullRelease XML: Measure/AttributeSet/Attribute@Type='FunctionalConsequence'
VariationRelease XML: SimpleAllele/FunctionalConsequence
db: GTR.clinvar.measure_attr where attr_type = 474

Method for determining functional consequence

Spreadsheet: FunctionalEvidence. Method
FullRelease XML: Method/Description
db: GTR.clinvar.method.description

Functional consequence comment

- optional

A free text comment may be submitted about the functional consequence. The comment structure should be used if the consequence being submitted is not defined by VariO, SO, or LOVD. We strongly recommend, however, that a term be requested if current terms are insufficient.

Spreadsheet: Variant.Comment on functional consequence
FullRelease XML: Measure/AttributeSet/Attribute Type='FunctionalConsequence'
MeasureAttributeSet/Comment/CommentText
VariationRelease XML: SimpleAllele/FunctionalConsequence/Comment
db: GTR.clinvar.measure_attr

Type of variation

Description of the type of variation, using terms from the Sequence Ontology as appropriate. Note that the option *undefined* exists as a default value. NCBI reassigns the type when necessary. For the list of allowed values, please refer to the XSD or the authorities document (https://www.ncbi.nlm.nih.gov/clinvar/docs/authorities/).

Spreadsheet: Variant.Variant type (required only for structural variants)
FullRelease XML: Measure/MeasureType
VariationRelease XML: SimpleAllele/VariantType
db: GTR.clinvar.measure.id_type

Sets of variants

A condition may be interpreted for a haplotype with more than one sequence change or for a genotype (this is rare). For consistent representation, all interpreted variants in ClinVar are represented as a set. For most variants in the database, the set has a single member; haplotypes and genotypes have multiple members. This representation is to be distinguished from co-occurrence, which is used to report rare alleles in genes thought to contribute to a condition, but for which the alleles are not thought to be pathogenic in the reported context. Sets of variants can have many of the same attributes as single variants, *e.g.* names, identifiers in other databases, allele frequencies, *etc*.

Spreadsheet: Variant.HGVS
Submission XML: ClinVarSubmission/MeasureSet or GenotypeSet
FullRelease XML: ReferenceClinVarAssertion/MeasureSet or GenotypeSet
VariationRelease XML: InterpretedRecord/SimpleAllele or Haplotype or Genotype

VariationRelease XML: IncludedRecord/SimpleAllele or Haplotype

Variation ID

The identifier for the variant or set of variants that were interpreted.

Spreadsheet: not represented
FullRelease XML: MeasureTrait/MeasureSet/@ID
VariationRelease XML: VariationArchive/@ID
VariationRelease XML: SimpleAllele/@VariationID

Variation name

This is a name provided by NCBI for the variant or set of variant. Many names are calculated; some may be curated or designed by official nomenclature groups.

Spreadsheet: Variant.Official allele name
FullRelease XML: MeasureTrait/MeasureSet/Name
VariationRelease XML: VariationArchive/@VariationName

Submitter's identifier for the allele being described in the submission

- optional

ClinVar uses this value, plus the condition, to construct a unique key for the clinical assertion being submitted. This key is included in reports of assertions that we were unable to process.

Spreadsheet: Variant.Local ID
FullRelease XML: ClinvarSubmission/ClinvarSubmissionID/@localKey
FullRelease XML: ClinvarSubmission/MeasureTrait/ExternalID/@id
VariationRelease XML: ClinicalAssertionList/ClinicalAssertion/SubmissionID/@localKey
db:GTR.clinvar.measure_target.local_key

URL to submitter's record

- optional

A URL that points to the submitted variant on the submitter's website and that does not match the pattern of Base URL + Local ID (see Base URL, above). ClinVar uses a mapping of database (DB) and base URLs to construct links; this mapping is not currently publicly available.

Spreadsheet: Variant.URL
FullRelease XML: XRef/@url (The Xref structure can be provided at many levels in the submission, to indicate what URL the submitter has for that object).
FullRelease XML: XRef/@DB, @ID, @Type
VariationRelease XML: XRef/@url (The Xref structure can be provided at many levels in the submission, to indicate what URL the submitter has for that object).

VariationRelease XML: XRef/@DB, @ID, @Type
db: GTR.clinvar.attr_source

Definition of the variant by locations and sequence changes

- required

The reference sequence and version, such as NM_000492.3, NG_016465.3, NC_000007.13, LRG_76t1.

Spreadsheet: Variant.Reference sequence, HGVS
FullRelease XML: ClinvarSubmission/MeasureSet/Measure/AttributeSet/MeasureAttributeType='HGVS'
FullRelease XML: ClinvarSubmission/MeasureSet/Measure/AttributeSet/Attribute
VariationRelease XML: SimpleAllele/AttributeSet/Attribute/@Type='HGVS'
VariationRelease XML: SimpleAllele/AttributeSet/Attribute/
db: GTR.clinvar.mset + GTR.clinvar.msubset+GTR.clinvar.measure

OMIM allelic variant ID

- optional

An OMIM allelic variant ID is reported for a set of variations as appropriate. If an allele occurs in more than one gene, and has multiple allelic variant ids, then both are reported. If curation determines that there are multiple allelic variant identifiers for the same allele, both identifiers are reported in that case as well.
Spreadsheet Variant.Variation identifiers
FullRelease XML: //ClinvarSubmission/MeasureSet/XRef
VariationRelease XML: //SimpleAllele/XRef
db: GTR.clinvar.mset_attr

Additional descriptors

Names and other attributes of a set of alleles can be submitted, similar to the names and attributes of single alleles.

Description of the asserted relationship between a set of variants and a set of conditions

Mode of inheritance

- optional

The mode of inheritance is reported as an attribute of the relationship between a variant (or a set of variants) and the condition. The list of allowed values is maintained for ClinVar and GTR on GTR's ftp site:

https://ftp.ncbi.nlm.nih.gov/pub/GTR/standard_terms/Mode_of_inheritance.txt

If you provide other, please specify, e.g. other:'new mode of inheritance'.
Spreadsheet: Variant.Mode of inheritance
FullRelease XML: MeasureTrait/AttributeSet/MeasureTraitAttributeType = "ModeOfInheritance"
FullRelease XML: MeasureTrait/AttributeSet/Attribute
VariationRelease XML: ClinicalAssertion/AttributeSet/Attribute/@Type = "ModeOfInheritance"
VariationRelease XML: ClinicalAssertion/AttributeSet/Attribute/
db: GTR.clinvar.mt_attr where attr_type = 163

Clinical significance

Description

- Required

Clinical significance is required for public reporting. The list of allowed values is maintained for ClinVar and GTR on GTR's ftp site:

https://ftp.ncbi.nlm.nih.gov/pub/GTR/standard_terms/Clinical_significance.txt

Guidance for use of these terms in ClinVar submissions is found on the ClinVar website:
https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/#clinsig_options_scv

If you submit "other" or if you have a value that you think should be included in this list, please contact us at clinvar@ncbi.nlm.nih.gov.

One option is 'not provided'; however, this value is only allowed in a few cases, documented in the link above for guidance on using terms for clinical significance.
Spreadsheet: Variant.Clinical significance
FullRelease XML: MeasureTrait/ClinicalSignificance/Description
VariationRelease XML: Interpretation/Description
db: GTR.clinvar.mt_attr where attr_type = 151

Date last evaluated

- Required if available

Date the clinical significance of the variant was last evaluated (not the date that a patient was evaluated). This may be different from the submission date, but it cannot be after the submission date.
Spreadsheet: Variant.Date last evaluated
FullRelease XML: MeasureTrait/ClinicalSignificance/DateLastEvaluated
VariationRelease XML: Interpretation/@DateLastEvaluated
db: GTR.clinvar.mt_attr.attr_date where attr_type = 151

Assertion method

- required

Assertion criteria (or assertion method) is a document describing the set of categories used to classify variants and the criteria needed to classify a variant in each category. This is the general process that the submitter uses to classify variants, not the specific evidence used to classify each variant. A name for the method and a document (a file, a URL, or a citation) are provided for assertion criteria. The inclusion of assertion criteria affects the review status of a submission; see the Review status section.

Spreadsheet: Variant.Assertion method
FullRelease XML: MeasureTrait/AttributeSet/MeasureTraitAttributeType/ = AssertionMethod
FullRelease XML: MeasureTrait/AttributeSet/Attribute/
VariationRelease XML: AttributeSet/Attribute/@Type = AssertionMethod/
db:

Spreadsheet: Variant.Assertion method citation
FullRelease XML: MeasureTrait/AttributeSet/Citation/ID/@Source = (type of citation identifier)
FullRelease XML: MeasureTrait/AttributeSet/Citation/ID/URL/
VariationRelease XML: AttributeSet/Citation/URL

Citations

- optional

Citations documenting the assertion of clinical significance. Any of PubMed, PubMedCentral, DOI, NCBI Bookshelf combined with the id in that database (e.g. PMID:123456). See Citations.
Spreadsheet: Variant.Clinical significance citations
Spreadsheet: Variant. Citations or URLs for clinical significance without database identifiers
FullRelease XML: MeasureTrait/ClinicalSignificance/Citation (CitationType)
VariationRelease XML: Interpretation/Citation
db: GTR.dbo.citation; GTR.dbo.citation_many

Comment

- Optional, but highly recommended

Free text describing the rationale for the assertion of clinical significance. This should be a general comment about the interpretation. Specific comments about each observation may be provided as part of the evidence (see Evidence/ObservedIn). See Comments.

Spreadsheet: Variant.Comment on clinical significance
FullRelease XML: MeasureTrait/ClinicalSignificance/Comment
VariationRelease XML: Interpretation/Comment
db GTR.dbo.comment

Custom Assertion Score

- optional

Submitter-specific scoring method names and the values obtained for each, (where submitter has alternate system/nomenclature for clinical significance). These are not standardized, not stored in a normalized fashion in relational columns, but are being retained for the submitter's use.

Spreadsheet: Not represented in spreadsheet
FullRelease XML: MeasureTrait.CustomAssertionScore[@Value= "*string*"]
FullRelease XML: MeasureTrait/CustomAssertionScore/CustomAssertionScoreType
VariationRelease XML: ClinicalAssertion/CustomAssertionScore
db: GTR.clinvar.version.xml_object

Citations

Citations describing the custom assertion score. See Citations.

Spreadsheet: Not represented in spreadsheet
FullRelease XML: MeasureTrait/CustomAssertionScore/Citation
VariationRelease XML: ClinicalAssertion/CustomAssertionScore/Citation

XRef

Database cross-references for the custom assertion score. See XRefs.

Spreadsheet: Not represented in spreadsheet
FullRelease XML: MeasureTrait/CustomAssertionScore/XRef
VariationRelease XML: ClinicalAssertion/CustomAssertionScore/XRef

Evidence/ObservedIn

The evidence section maintains the details necessary to review the assertion of clinical significance. This evidence may be computational, based on experimental testing, or observations in human subjects. (Note that ClinVar does not accept submissions where the evidence is only computational.) A submission may contain multiple observations for the same allele/condition combination. In the XML, these are represented by multiple //MeasureTrait/ObservedIn elements; in the spreadsheet these are represented by multiple lines in the CaseData or FunctionalEvidence tab with the same value in the Linking ID column as the Linking ID column of the Variant tab.

Spreadsheet: Variant, CaseData, FunctionalEvidence
FullRelease XML: ObservedIn
VariationRelease XML: ObservedIn
db: GTR.clinvar.observations

Sample

This section is used to describe the sample that was studied to generate the submission and its supporting evidence.

Allele origin

- required

The genetic origin of the variant for each observation. Allowed values are germline, de novo, somatic, maternal, paternal, inherited, unknown, uniparental, biparental. Uniparental and biparental are intended for the context of uniparental disomy, not to indicate zygosity.

Spreadsheet: Variant.Allele origin
Spreadsheet: CaseData.Allele origin
Spreadsheet: FunctionalEvidence.Allele origin
FullRelease XML: ObservedIn/Sample/Origin
VariationRelease XML: ObservedIn/Sample/Origin
db: clinvar.sample.origin

Species

- required

Defaults to human if not supplied. Although ClinVar only accepts submissions for human variants, functional evidence from other species may be provided as evidence.

Spreadsheet: FunctionalEvidence.Species
FullRelease XML: ObservedIn/Sample/Species/@TaxonomyId=9606 (“human”)
VariationRelease XML: ObservedIn/Sample/Species/@TaxonomyId=9606 (“human”)
db: clinvar.sample.txid

Affected status

- required

Indicates whether the individual(s) in which the variant was identified had the condition for which an assertion is being made. Allowed values are yes, no, unknown.
Spreadsheet: Variant.Affected status
Spreadsheet: CaseData.Affected status
Spreadsheet: FunctionalEvidence.Affected status
FullRelease XML: ObservedIn/Sample/AffectedStatus
VariationRelease XML: ObservedIn/Sample/AffectedStatus
db: clinvar.sample.affected_status

Structural variant method/analysis type

Sex

- optional

If explicit numbers are known in a sample set, they should be specified in XML as NumberMales and /or NumberFemales. Otherwise, use Gender.

Spreadsheet: Variant.Sex
Spreadsheet: CaseData.Sex
Spreadsheet: FunctionalEvidence.Sex
FullRelease XML: ObservedIn/Sample/NumberMales
FullRelease XML: ObservedIn/Sample/NumberFemales
FullRelease XML: ObservedIn/Sample/Gender
VariationRelease XML: ObservedIn/Sample/NumberMales
VariationRelease XML: ObservedIn/Sample/NumberFemales
VariationRelease XML: ObservedIn/Sample/Gender
db: clinvar.sample.males
db: clinvar.sample.females
db: clinvar.obs_attr

Age/age range

- optional

The age of the individual or the range of ages for individuals included in this aggregate observation. If age range is an important variable in your submission, with different observations based on the age, please submit each observation separately, rather than lumping into one summary observation with one sample description

Spreadsheet: Variant.Age range
Spreadsheet: CaseData.Age
FullRelease XML: ObservedIn/Sample/Age
VariationRelease XML: ObservedIn/Sample/Age
db: clinvar.sample.min_age
db: clinvar.sample.max_age
db: clinvar.sample.age_units

Geographic origin

- optional

Can be used to indicate country, continent, or a region in which this allele was reported.

Spreadsheet: Variant.Geographic origin
Spreadsheet: CaseData.Geographic origin
FullRelease XML: ObservedIn/Sample/GeographicOrigin
VariationRelease XML: ObservedIn/Sample/GeographicOrigin
db: GTR.clinvar.sample.geographic_origin

Population Group/**Ethnicity**

- optional

Name or description of the ethnicities of the individual in which the allele was reported.

Spreadsheet: Variant.Ethnicity
Spreadsheet: CaseData.Ethnicity
FullRelease XML: ObservedIn/Sample/Ethnicity

VariationRelease XML: ObservedIn/Sample/Ethnicity
db: clinvar.sample.ethnicity

Tissue

- optional

Name or description of the tissue that was assayed. Highly recommended if the origin is somatic or for experimental observations.
Spreadsheet: Variant.Tissue
Spreadsheet: CaseData.Tissue
Spreadsheet: FunctionalEvidence.Tissue
FullRelease XML: ObservedIn/Sample/Tissue
VariationRelease XML: ObservedIn/Sample/Tissue
db: clinvar.sample.tissue

Fraction of sample which is tumor-containing

- optional

Free text description of the fraction of the sample that has tumor cells. Applicable only if origin is somatic.
Spreadsheet: Not represented in spreadsheet
FullRelease XML: ObservedIn/Sample/FractionTumor
VariationRelease XML: Not represented in ClinVarVariationRelease
db: clinvar.sample.fraction_tumor

Number of chromosomes tested

- optional

Used as the denominator when “Number of chromosomes with variant” is reported for Observed data.
Spreadsheet: Not represented in spreadsheet
FullRelease XML: ObservedIn/Sample/NumberChrTested
VariationRelease XML: ObservedIn/Sample/NumberChrTested
db: clinvar.sample.chr_tested

Number of individuals tested

- optional

The number of subjects on which this submission is based. Used as the denominator when “Number of individuals with variant” is reported for Observed data.
Spreadsheet: Variant.Total number of individuals tested
FullRelease XML: ObservedIn/Sample/NumberTested
VariationRelease XML: ObservedIn/Sample/NumberTested
db: clinvar.sample.individuals_tested

Number families tested

- optional

Used as the denominator when "Number of families with variant" is reported for Observed data.

Spreadsheet: Variant.Number of families tested
FullRelease XML: ObservedIn/Sample/FamilyData/FamilyHistory/@NumFamilies
VariationRelease XML: ObservedIn/Sample/FamilyData/FamilyHistory/@NumFamilies
db: clinvar.sample.families_tested

Number of families with segregation observed

- optional

Number of independent families where the variant segregates with the condition among two or more family members.

Spreadsheet: Variant. Number of families with Segregation observed
FullRelease XML:
ObservedIn/Sample/FamilyData/FamilyHistory/@NumFamiliesWithSegregationObserved
VariationRelease XML:
ObservedIn/Sample/FamilyData/FamilyHistory/@NumFamiliesWithSegregationObserved
db: clinvar.obs_attr

Family history

- optional

Used to indicate that at least one other member of a family has the reported condition. It does not require that other family members were included in the observation set. Allowed values are yes, no.
Spreadsheet: Variant.Family history
FullRelease XML: ObservedIn/Sample/FamilyData/FamilyHistory
VariationRelease XML: ObservedIn/Sample/FamilyData/FamilyHistory
db: clinvar.sample.positive_family_history

Number of Independent Affected Subjects tested

- optional but highly encouraged

ClinVar computes how many times a variant has been seen in affected individuals from independent families by data aggregated by the submitter (Variant tab on the spreadsheet), or by summing data submitted as cases.
Spreadsheet: CaseData.Proband, Family ID (CaseData)
Spreadsheet: Variant.Number of families with variant, Affected status=yes
FullRelease XML: ObservedIn/ObservedData/Attribute/@ type = IndependentObservations
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
FullRelease XML: ObservedIn.Sample.AffectedStatus = "yes"

VariationRelease XML: ObservedIn/ObservedData/Attribute/@ type = IndependentObservations
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn.Sample.AffectedStatus = "yes"
db: GTR.clinvar.obs_attr where attr_type =IndependentObservations
db: GTR.clinvar.sample.affected_status='yes'

Secondary finding

- optional

To indicate variants that were identified as secondary findings, i.e. variants unrelated to the indication for testing but are known pathogenic or expected pathogenic variants in genes recommended by ACMG for reporting of secondary findings.

Spreadsheet: Variant.Secondary finding
Spreadsheet: CaseData.Secondary finding
FullRelease XML: ObservedIn/ObservedData/Attribute/@ type = SecondaryFinding
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@ type = SecondaryFinding
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue

Family ID

- Optional

An anonymous identifier used to indicate individual observations that are from the same family.

Spreadsheet: CaseData.Family ID
FullRelease XML: ObservedIn/Sample/FamilyData/@PedigreeID
VariationRelease XML: ObservedIn/Sample/FamilyData/@PedigreeID
db: clinvar.sample.family_id

Citations

- optional

Citations documenting the evidence. Any of PubMed, PubMedCentral, DOI, NCBI Bookshelf combined with the id in that database (e.g. PMID:123456). See Citations.
Spreadsheet: Variant.Evidence citations
Spreadsheet: CaseData.Evidence citations
Spreadsheet: FunctionalEvidence.Evidence citations
FullRelease XML: ObservedIn/Citation
VariationRelease XML: ObservedIn/Citation
db: GTR.dbo.citation and GTR.dbo.citation_many

Comment

Free text describing the observation. A general comment about the interpretation can be provided with clinical significance (see Clinical significance). See Comments.

Spreadsheet: Variant.Comment on evidence
Spreadsheet: CaseData.Comment on evidence
Spreadsheet: FunctionalEvidence.Comment on evidence
FullRelease XML: ObservedIn/Citation
VariationRelease XML: ObservedIn/Comment

Cell line

- optional (required if evidence was generated in a cell line)

Name of the cell line. To be used only when experimental evidence is being reported.
Spreadsheet: FunctionalEvidence.Cell line
FullRelease XML: ObservedIn/Sample/CellLine
VariationRelease XML: ObservedIn/Sample/CellLine
db: clinvar.sample.cell_line

Strain/breed

- optional

Name of the strain or breed that was analyzed in this experimental study.
Spreadsheet: FunctionalEvidence.Strain/breed
FullRelease XML: ObservedIn/Sample/Strain
VariationRelease XML: ObservedIn/Sample/Strain
db: clinvar.sample.strain

Observed phenotypes

- optional

A list of clinical features observed in the sample (not all the clinical features that are associated with a named condition). If provided for an aggregate observation, the set of features should apply to all individuals in the aggregate. Clinical features may be provided for each individual observation as well. HPO terms may be used to indicate the clinical features.
Spreadsheet: Variant.Clinical features
Spreadsheet: CaseData.Clinical features
FullRelease XML: ObservedIn/TraitSet
VariationRelease XML: ObservedIn/TraitSet

Comment

- optional

A free text comment may be submitted to expand on the terms provided as observed phenotypes/clinical features for an observation. The comment should supplement, not replace, HPO terms. The comment may provide more

detail, such as qualifying the phenotypes or explaining the progression of presentation of phenotypes. See Comments.

Spreadsheet: Variant.Comment on clinical features
Spreadsheet: CaseData.Comment on clinical features
FullRelease XML: ObservedIn/TraitSet/Comment
VariationRelease XML: ObservedIn/TraitSet/Comment

Method

Test name or type

- optional

Because testing laboratories may know only the name of the ordered test (e.g. "deafness"), the test name may be submitted with condition submitted as "not provided". A GTR id maybe submitted instead of a free text name.

Spreadsheet: Variant.Test name or type
Spreadsheet: CaseData.Test name or type
FullRelease XML: ObservedIn/Method/MethodAttribute/Attribute/@Type = TestName
FullRelease XML: ObservedIn/Method/XRef for GTR ids
VariationRelease XML: ObservedIn/Method/MethodAttribute/Attribute/@Type = TestName
VariationRelease XML: ObservedIn/Method/XRef for GTR ids
db: clinvar.method.extrn_src (for GTR ids)
db: clinvar.method.extrn_id (for GTR ids)

Platform type

- optional

The type of platform used for data capture, e.g. next-gen sequencing, microarray.
Spreadsheet: Variant.Platform type
Spreadsheet: CaseData. Platform type
Spreadsheet: FunctionalEvidence.Platform type
FullRelease XML: ObservedIn/Method/TypePlatform
VariationRelease XML: ObservedIn/Method/ TypePlatform

Platform name

- optional

The name of the platform used for data capture, e.g. HiSeq, MiSeq.
Spreadsheet: Variant.Platform name
Spreadsheet: CaseData. Platform name
Spreadsheet: FunctionalEvidence.Platform name
FullRelease XML: ObservedIn/Method/NamePlatform
VariationRelease XML: ObservedIn/Method/ NamePlatform

Method

- optional

Free text describing the experimental method used to generate support. This is particularly relevant for submissions of experimental evidence.

Spreadsheet: Variant.Method
Spreadsheet: CaseData.Method
Spreadsheet: FunctionalEvidence.Method
FullRelease XML: ObservedIn/Method/Description
VariationRelease XML: ObservedIn/Method/Description

Result

- optional

Free text describing the result of the test or experimental assay.

Spreadsheet: FunctionalEvidence.Method
FullRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute/@Type = MethodResult
FullRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute/
VariationRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute/@Type = MethodResult
VariationRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute/

Method purpose

- optional

The primary intent of the method used to generate the data provided in the submission. Allowed values are discovery, validation, assert pathogenicity.

Spreadsheet: Variant.Method purpose
Spreadsheet: CaseData.Method purpose
Spreadsheet: FunctionalEvidence.Method purpose
FullRelease XML: ObservedIn/Method/Purpose
VariationRelease XML: ObservedIn/Method/Purpose

Method citations

- optional

Citations documenting the method. Any of PubMed, PubMedCentral, DOI, NCBI Bookshelf combined with the id in that database (e.g. PMID:123456). See Citations.
Spreadsheet: Variant.Method citations

Spreadsheet: CaseData.Method citations
Spreadsheet: FunctionalEvidence.Method citations
FullRelease XML: ObservedIn/Method/Citation
VariationRelease XML: ObservedIn/Method/Citation

Software name and version

- optional

Description of key software with explicit representation of name and version.

Spreadsheet: Variant.Software name and version
Spreadsheet: CaseData.Software name and version
Spreadsheet: FunctionalEvidence.Software name and version
FullRelease XML: ObservedIn/Method/Software/@name
FullRelease XML: ObservedIn/Method/Software/@version
VariationRelease XML: ObservedIn/Method/ Software/@name
VariationRelease XML: ObservedIn/Method/Software/@version

Software purpose

- optional

Purpose of the software. e.g. alignment, variant calling.

Spreadsheet: Variant.Software purpose
Spreadsheet: CaseData.Software purpose
Spreadsheet: FunctionalEvidence.Software purpose
FullRelease XML: ObservedIn/Method/Software/@purpose
VariationRelease XML: ObservedIn/Method/Software/@purpose

Testing laboratory

- optional

Used when the variant was tested by a clinical testing laboratory other than the submitter. This may apply to submissions from clinicians who provide their own interpretation of a variant reported to them, or to submissions from patient registries or clinicians with a focus on phenotype rather than interpretation.

Spreadsheet: Variant.Testing laboratory
Spreadsheet: CaseData.Testing laboratory
FullRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute/@Type = TestingLaboratory
FullRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute
VariationRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute/@Type = TestingLaboratory

VariationRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute

Date variant was reported to submitter

- optional

Used when the variant was tested by a clinical testing laboratory other than the submitter. This may apply to submissions from clinicians who provide their own interpretation of a variant reported to them, or to submissions from patient registries or clinicians with a focus on phenotype rather than interpretation.

Spreadsheet: Variant.Date variant was reported to submitter
Spreadsheet: CaseData. Date variant was reported to submitter
FullRelease XML: ObservedIn/Method/ObsMethodAttribute/Attribute/@dateValue
VariationRelease XML: ObservedIn/Method/ ObsMethodAttribute/Attribute/@dateValue

Observations/Observed Data

ClinVar uses 'observations' to store evidence generated from a combination of methods applied to a sample.

Methods for data collection

The method used to collect data for a submission is reported as a method type (collection method on the spreadsheet and website). The options are used to support evaluation of the submission, as well as to allow representation of clear distinctions between primary data and data culled from the literature.

Spreadsheet: Variant.Collection method
Spreadsheet: CaseData.Collection method
Spreadsheet: FunctionalEvidence.Collection method
FullRelease XML: ObservedIn/Method
VariationRelease XML: ObservedIn/Method

Option	Explanation
clinical testing	For variants that were interpreted as part of clinical genetic testing, or as part of a large volume research study in which results compliant with CLIA, ISO, GLP, or an equivalent accreditation body are routinely returned to research subjects. Interpretation may be guided from the literature, but the number of individuals tested are reported only from the direct testing.
literature only	For variants extracted from published literature with interpretation as reported in the citation.
curation	For variants that were not directly observed by the submitter, but were interpreted by curation of multiple sources, such as publications, public databases, and unpublished case data.

research	For variants that were interpreted as part of a research project but results are not routinely returned to research subjects and do not meet the requirements for clinical testing above. This is a general term to use when other more specific methods to not apply.
reference population	For variants gathered in a research setting but results are not routinely returned to research subjects and do not meet the requirements for clinical testing above. This term is used for baseline studies of a population group of apparently unaffected individuals to assess allele frequencies.
case-control	For variants gathered in a research setting but results are not routinely returned to research subjects and do not meet the requirements for clinical testing above. This term is for research projects specifically to compare alleles observed in cases and controls (without data about segregation).
in vivo	For variants that were interpreted as part of an *in vivo* research project but results are not routinely returned to research subjects and do not meet the requirements for clinical testing above.
in vitro	For variants that were interpreted as part of an *in vitro* research project but results are not routinely returned to research subjects and do not meet the requirements for clinical testing above.
provider interpretation	For variants that were interpreted by a clinical provider.
phenotyping only	For variants that are submitted to ClinVar to provide individual observations with detailed phenotype data, such as submissions from clinicians or patient registries, without an interpretation from the submitter. The interpretation from the testing laboratory may be provided in a separate field.
Not provided	This value should not be submitted.

Number of families with variant

- optional

Spreadsheet: Variant.Number of families with variant
FullRelease XML: ObservedIn/FamilyData/FamilyHistory/@NumFamiliesWithVariant
VariationRelease XML: ObservedIn/FamilyData/FamilyHistory/@NumFamiliesWithVariant
db: clinvar.obs_attr

Number of individuals with variant

- optional

Spreadsheet: Variant.Number of individuals with variant
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type=VariantAlleles
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type=VariantAlleles
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: GTR.clinvar.obs_attr where attr_type ='VariantAlleles'

Number of independent individuals with variant

- optional

Spreadsheet: not represented
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type=IndependentObservations
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type=IndependentObservations
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: clinvar.obs_attr

Number of chromosomes with variant

- optional

Spreadsheet: Variant.Number of chromosomes with variant
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type=VariantChromosomes
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type=VariantChromosomes
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: clinvar.obs_attr

Number of homozygotes

- optional

Spreadsheet: Variant.Number of homozygotes
Spreadsheet: CaseData.Zygosity
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = Homozygote
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = Homozygote
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: clinvar.obs_attr

Number of hemizygotes

- optional

Spreadsheet: Variant.Number of homozygotes
Spreadsheet: CaseData.Zygosity
FullRelease XML: ObservedIn/ObservedData/ObsAttributeType[@val_type = Hemizygote
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/ObsAttributeType[@val_type = Hemizygote
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: GTR.clinvar.obs_attr

Number of single heterozygotes

- optional

This count includes single heterozygotes reported in the context of dominant mode of inheritance, and single heterozygotes observed (in a recessive context) but where no other pathogenic variant was identified to classify as a compound heterozygote.

Spreadsheet: Variant.Number of single heterozygotes
Spreadsheet: CaseData.Zygosity
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = SingleHeterozygote
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = SingleHeterozygote
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: GTR.clinvar.obs_attr where attr_type ='SingleHeterozygote'

Number of compound heterozygotes

- optional

This is a count of heterozygotes where another heterozygous pathogenic variant WAS identified. Both variant alleles must be submitted.

Spreadsheet: Variant.Number of compound heterozygotes
Spreadsheet: CaseData.Zygosity
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = CompoundHeterozygote
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = CompoundHeterozygote
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: GTR.clinvar.obs_attr where attr_type =CompoundHeterozygote

Mosaicism

- optional

Spreadsheet: Variant.Mosaicism
Spreadsheet: CaseData.Mosaicism
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = NumberMosaic
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = NumberMosaic

VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: clinvar.obs_attr

Number of affected subjects with genotype consistent with mode of inheritance

- optional

The sum of single heterozygotes, compound heterozygotes, homozygotes, and hemizygotes for the reported allele with a condition consistent with the asserted mode of inheritance.

Spreadsheet: not represented

FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = GenotypeAndMOIConsistent
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
FullRelease XML: ObservedIn.Sample.AffectedStatus = “yes”
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = GenotypeAndMOIConsistent
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn.Sample.AffectedStatus = “yes”
db: GTR.clinvar.obs_attr where attr_type =GenotypeAndMOIConsistent

Number of affected subjects with this variant who also have another variant thought to be responsible for condition

- optional

This information is captured to evaluate pathogenicity, based on the logic that if another allele may account for the observed condition, this one has unknown pathogenicity.

Spreadsheet: not represented
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = SubjectsWithDifferentCausativeAllele
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
FullRelease XML: ObservedIn.Sample.AffectedStatus = “yes”
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = SubjectsWithDifferentCausativeAllele
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn.Sample.AffectedStatus = “yes”
db: GTR.clinvar.obs_attr where attr_type=’ SubjectsWithDifferentCausativeAllele’

Number of instances observed of heterozygous parent transmitting normal allele to an affected child

- optional

One of several observations that support evaluation of penetrance.

Spreadsheet: not represented
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = HetParentTransmitNormalAllele
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
FullRelease XML: ObservedIn.Sample.AffectedStatus = “yes”
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = HetParentTransmitNormalAllele
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn.Sample.AffectedStatus = “yes”
db: GTR.clinvar.obs_attr where attr_type =’HetParentTransmitNormalAllele ‘
db: GTR.clinvar.sample.affected_status=’yes’

Number of independent families demonstrating co-segregation

- optional

Spreadsheet: Not represented

FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = CosegregatingFamilies
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
FullRelease XML: ObservedIn.Sample.AffectedStatus = “yes”
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = CosegregatingFamilies
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn.Sample.AffectedStatus = “yes”
db: GTR.clinvar.obs_attr where attr_type =’CosegregatingFamilies’
db: GTR.clinvar.sample.affected_status=’yes’

Number of informative meioses

- optional

Spreadsheet: Not represented
FullRelease XML: ObservedIn/ObservedData/Attribute/@Type = InformativeMeioses
FullRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
VariationRelease XML: ObservedIn/ObservedData/Attribute/@Type = InformativeMeioses
VariationRelease XML: ObservedIn/ObservedData/Attribute/@integerValue
db: GTR.clinvar.obs_attr where attr_type =’InformativeMeioses’

Co-occurrence

- optional

Submitters may have identified other alleles which may contribute to the condition and were observed in the individuals studied, but which are not included in the interpretation. The values provided in each co-occurrence set describe the distinct sets of co-occurring alleles and the number of times each was observed.

Zygosity

- optional

The zygosity of the co-occurring variant, not the interpreted variant.

Spreadsheet: Not represented
FullRelease XML: ObservedIn/Co-occurrenceSet/Zygosity
VariationRelease XML: ObservedIn/Co-occurrenceSet/Zygosity
db: GTR.clinvar.co_occurs.zygosity_type

Count of individuals with this co-occurrence set

- optional

Spreadsheet: Not represented
FullRelease XML: ObservedIn/Co-occurrenceSet/Count
VariationRelease XML: ObservedIn/Co-occurrenceSet/Count

db: GTR.clinvar.co_occurs.num_reported

Description of the alleles observed (co-occurring genotypes)

Spreadsheet: CaseData.Co-occurrences, same gene
Spreadsheet: CaseData.Co-occurrences, other genes
FullRelease XML: ObservedIn/Co-occurrenceSet/AlleleDescrSet/Name
VariationRelease XML: ObservedIn/Co-occurrenceSet/AlleleDescrSet/Name
db: GTR.clinvar.alelle_set.name

- Orientation

FullRelease XML: ObservedIn/Co-occurrenceSet/AlleleDescrSet/RelativeOrientation
db: GTR.clinvar.alelle_set.orientation

- Zygosity

FullRelease XML: ObservedIn/Co-occurrenceSet/AlleleDescrSet/Zygosity
db: GTR.clinvar.alelle_set.zygosity_type

- Clinical Significance

FullRelease XML: ObservedIn/Co-occurrenceSet/AlleleDescrSet/ClinicalSignificance
db: GTR.clinvar.alelle_set.clinical_significance

Description of a gene

- optional

A gene, if provided, must be unambiguously defined. That definition may be supplied either by the official symbol (see the Names section), or an identifier in a public database: GeneID, HGNC id or MIM number (see the Attributes section). The gene is considered optional because the location of the variation should be sufficient to define the gene.

Genes are represented in ClinVar in two major contexts:

1. The gene in which a variation has been described
2. An explicit representation of gene-condition relationship.

The former case is represented in the XML as a MeasureRelationship of type 'variant in gene'.

The latter case is used internally, but is not currently processed as a publicly reported accession.

Names

ClinVar coordinates the representation of names of genes with NCBI's Gene database. In other words, the name is defined primarily by the nomenclature established by the HUGO Gene Nomenclature Committee (HGNC).

Preferred name

- optional, only one allowed

The preferred full name as reported by NCBI's Gene database.

Spreadsheet: not represented
FullRelease XML: MeasureRelationship/Name/ElementValueType[@val_type="name"] = Preferred
FullRelease XML: MeasureRelationship/Name/ElementValue
VariationRelease XML: SimpleAllele/GeneList/Gene/@FullName
db: GTR.clinvar.measure_attr where attr_type = 17 (AttributeSet)

Alternate name(s)

- optional, multiple allowed

The previous official name from HGNC.
Spreadsheet: not represented
FullRelease XML: MeasureRelationship/Name/ElementValueType[@val_type="name"] = Alternate
FullRelease XML: MeasureRelationship/Name/ElementValue
VariationRelease XML: not represented
db: GTR.clinvar.measure_attr where attr_type = 18 (AttributeSet)

Preferred symbol

- optional, only one allowed

The official symbol from HGNC. This may be used by submitters to indicate the gene.

Spreadsheet: Variant.Gene symbol
FullRelease XML: MeasureRelationship/Symbol/ElementValueType[@val_type="name"] = Preferred
FullRelease XML: MeasureRelationship/Symbol/ElementValue
VariationRelease XML: SimpleAllele/GeneList/Gene/@Symbol
db: GTR.clinvar.measure_attr where attr_type = 19 (AttributeSet)

Alternate symbols(s)

- optional, multiple allowed

The previous official symbol from HGNC.

Spreadsheet: not represented
FullRelease XML: MeasureRelationship/Symbol/ElementValueType[@val_type="name"] = Alternate
FullRelease XML: MeasureRelationship/Symbol/ElementValue
VariationRelease XML: not represented

db: GTR.clinvar.measure_attr where attr_type = 20 (AttributeSet)

Attributes

- Examples are GeneID, HGNC id, MIM number, chromosome, cytogenetic band, chromosome sequence location, and whether there are pseudogenes or paralogs. The set of optional attributes is designed to capture information necessary to set the framework for interpretation of variation.
- Many of these attributes are not duplicated in the ClinVar database but are provided by NCBI as imports from the Gene database or defined by the sequence used to define the gene structure.
- GeneID, HGNC id, and MIM number may be used by XML submitters to indicate the gene. The spreadsheet only supports submission of the HGNC-approved gene symbol.

Overview of Gene related concepts reported by ClinVar

Concept	FullRelease XML	VariationRelease XML	column in spreadsheet
GeneID	XRef/@DB=Gene	SimpleAllele/GeneList/Gene/@GeneID	Not represented
HGNC id	XRef/@DB=HGNC	SimpleAllele/GeneList/Gene/@HGNCID	Not represented
MIM number	XRef/@DB=OMIM	SimpleAllele/GeneList/Gene/OMIM	Not represented
HGNC symbol		SimpleAllele/GeneList/Gene/@Symbol	Variant.Gene
Location of the gene on the GRC assembly; chromosome	Measure/SequenceLocation/@Assembly, @Chr, @start, @stop	SimpleAllele/GeneList/Gene/Location/SequenceLocation/@Assembly, @Chr, @start, @stop	SubmissionInfo.Assembly; Variant.Chromosome coordinates
Gene relationships	Measure/MeasureRelationship/AttributeSet/@Type=”gene relationship”	SimpleAllele/GeneList/Gene/@RelationshipType	Not represented