

# Qualifying Evidence Matrix standard for verifiable evidence

Swiss Genomics Association

2025

## Version 1.0

**Standard identifier:** SGA-QEM-1.0

This document defines a normative standard published by the Swiss Genomics Association.

## 1 Introduction

The Qualifying Evidence Matrix (QEM) standard for verifiable evidence defines a minimal, interoperable representation of evidence availability derived from rule-based evaluations. It specifies how the outcomes of predefined evidence rules are converted into a binary matrix that records whether verifiable evidence is present or absent for each evaluated item.

In this context, an evidence rule is an externally defined, declarative check that assesses a specific condition for a given item under a declared interpretation context and emits a raw outcome of TRUE, FALSE, or NA. These raw outcomes encode whether a potential contradiction, consistency, or lack of information is observed. The QEM standard does not define the rules themselves, but specifies how their outcomes are interpreted uniformly.

The standard defines how evidence availability is represented, not what constitutes evidence. It is independent of analysis pipelines, domain-specific interpretation logic, and downstream statistical or decision-making methods.

The resulting binary evidence matrix is a lossy abstraction. Raw rule outcomes and their provenance **MUST** be preserved upstream if semantic detail, context, or re-interpretation is required.

## 2 Scope and non-goals

### 2.1 In scope

The QEM standard defines:

- the structure of a binary evidence matrix
- the allowed values of matrix entries
- the normative mapping from raw rule outcomes to matrix entries
- minimal integrity, provenance, and ordering requirements

## 2.2 Out of scope

The QEM standard does not define:

- data generation, other processing, or annotation formats
- domain-specific interpretation rules
- causal classification
- statistical inference or decision support

## 3 Terminology and conventions

The key words **MUST**, **MUST NOT**, **SHOULD**, and **MAY** are to be interpreted as described in RFC 2119.

An **evaluated item** is a uniquely identifiable entity supplied by an upstream process. For instance, in genomic applications, this is typically a genetic variant.

An **evidence rule** is a declarative check that evaluates a specific condition for an evaluated item and emits a raw outcome. Evidence rules **MUST** be defined such that a raw outcome of TRUE indicates a detected contradiction, absence, or weakening of evidence with respect to the evaluated hypothesis, while a raw outcome of FALSE indicates that no such contradiction was detected. A raw outcome of NA indicates that the rule could not be evaluated and does not constitute evidence.

The definition, semantics, and evaluation logic of evidence rules are external to this standard.

A **raw rule outcome** takes one of three values:

- TRUE
- FALSE
- Not Available (NA)

An **evidence matrix** is a two-dimensional array indexed by evaluated items and evidence rules.

## 4 Data model

### 4.1 Matrix definition

Let there be  $n$  evaluated items and  $m$  evidence rules.

The QEM is defined as:

$$X \in \{0, 1\}^{n \times m}$$

where each entry  $X_{ij}$  corresponds to evaluated item  $i$  and evidence rule  $j$ .

### 4.2 Ordering and identifier requirements

Each QEM instance **MUST** define a stable and unambiguous structural association between evaluated items and evidence rules.

Where the matrix is represented as a two-dimensional table, rows **MUST** correspond to evaluated items in a defined and stable order, and columns **MUST** correspond to evidence rules in a defined and stable order.

Where the matrix is represented in a non-tabular form, such as a relational database or graph-based representation, stable and unique identifiers for evaluated items and evidence rules **MUST** be provided, and a deterministic mapping to an ordered matrix view **MUST** be defined and documented.

Reordering of evaluated items or evidence rules, or changes to their identifiers, constitutes a different matrix instance and **MUST** be documented.

## 5 Mapping from raw rule outcomes

For each evaluated item  $i$  and evidence rule  $j$ , an upstream process produces a raw rule outcome in

$$\{\text{TRUE}, \text{FALSE}, \text{NA}\}.$$

The QEM standard defines the following **normative mapping**:

$$X_{ij} = \begin{cases} 1, & \text{if the raw outcome is FALSE,} \\ 0, & \text{if the raw outcome is TRUE,} \\ 0, & \text{if the raw outcome is NA.} \end{cases}$$

A value of  $X_{ij} = 1$  indicates that no contradiction or weakening signal was detected for evidence rule  $j$  and that verifiable evidence is present for evaluated item  $i$  under that rule.

A value of  $X_{ij} = 0$  indicates either a detected contradiction or the absence of evaluable information.

Evidence is accrued only from entries with value 1.

## 6 Integrity and provenance requirements

A QEM instance **MUST** include or reference sufficient metadata to ensure unambiguous interpretation and reproducibility.

At a minimum, this metadata **MUST** specify:

- a unique identifier for each evidence rule corresponding to the matrix columns
- the version or identifier of the evidence rule set used
- the version of the QEM standard applied

Where the matrix rows correspond to identifiable items, stable item identifiers **MUST** be provided or referenced.

Rule identifiers and item identifiers **MAY** be embedded in the matrix representation or supplied as external metadata.

The binary evidence matrix **MUST** be derivable deterministically from the declared raw rule outcomes under the referenced rule set and standard version.

## 7 Example (informative)

The following example illustrates a genomics-specific use of the QEM.

In clinical genetics, some evidence depends on whether parental genotypes were assessed. This reflects the completeness of evidence collection, not the validity of any inheritance model.

An evidence rule named `parent_gt_unavailable` records this at the raw outcome level (TRUE, FALSE, or NA), indicating whether the required information is missing, present, or unknown. This raw outcome is then converted into the binary evidence matrix, where a FALSE indicates that this evidence source is available (1) and a value of 0 indicates that it is not.

This two-step design enforces an unambiguous evidence question: whether verifiable evidence is present. Both detected contradictions (TRUE) and missing information (NA) therefore map to absence of evidence. Situations in which missingness itself is informative should be captured by separate, explicit rules, rather than overloading the binary evidence representation.

rule name	raw outcome	meaning	$X_{ij}$
<code>parent_gt_unavailable</code>	TRUE	parental genotype unavailable	0
<code>parent_gt_unavailable</code>	FALSE	parental genotype available	1
<code>parent_gt_unavailable</code>	NA	availability unknown	0

For a given evaluated item  $i$  and evidence rule  $j$ , exactly one raw outcome  $r_{ij} \in \{\text{TRUE}, \text{FALSE}, \text{NA}\}$  applies. The corresponding matrix entry  $X_{ij}$  is defined uniquely by this realised outcome.

evaluated item ID	parent_gt_unavailable
item_001	1

In this example, the rule does not assert pathogenicity or inheritance correctness. It records only whether a specific evidence source is present. The resulting matrix entry reflects evidence availability and might be combined with other rules to assess the overall completeness of verifiable evidence.

## 8 Non-goals and interpretation

The QEM does not assert causality, or for instance in genetics pathogenicity, or diagnostic correctness. It represents evidence availability only.

Downstream interpretation, ranking, or inference is outside the scope of this standard.

Evidence rules are not required to be independent. Correlation or dependency between rules is expected in practical applications. The QEM standard does not assume rule independence and does not address the detection, modelling, or correction of correlated evidence. Management of rule correlation, including upstream rule design or downstream statistical handling, is outside the scope of this standard.

## 9 Versioning

This document defines **QEM version 1.0**. Future revisions **MUST** preserve backward compatibility of the binary semantics.