Opinion Paper

Hikmet Can Çubukçu*, Oswald Sonntag, Nathan Lorde, Charles R. Lefèvre, Chiara Cosma, Annelien Van Dalem, Panagiotis N. Kanellopoulos, Marija Kocijancic, Guilaine Boursier, Pika Meško Brguljan, Neda Milinkovic, Marith van Schrojenstein Lantman, Marc Thelen, Amitava Dasgupta, Giuseppe Lippi and Mario Plebani, on behalf of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Committee Laboratory Error Database and Committee Accreditation and ISO/CEN Standards (C: A/ISO)

The severity of harm score database: an evidence-based foundation for laboratory test risk analysis

https://doi.org/10.1515/cclm-2026-0875
Received June 2, 2026; accepted July 16, 2026;
published online July 29, 2026

Abstract

Background: Patient risk management has become central to clinical laboratory quality standards, including the Clinical and Laboratory Standards Institute (CLSI) EP23 2nd edition (CLSI EP23-ED2), International Organization for Standardization (ISO) 15189:2022, ISO 22367:2026, and ISO 14971:2019. A key component of risk analysis is estimating the severity of harm (SoH) that may result from erroneous laboratory results. We present here the Severity of Harm Evidence Database, a structured, multi-study resource that integrates survey-based SoH assessments from the published independent studies. The database provides ordinal distributions and pooled SoH for 196 unique laboratory tests.
Methods: A canonical five-level scale (negligible, minor, serious, critical, catastrophic) was adopted, aligned with ISO 14971:2019. For the 19 tests common to both studies, direct count pooling was performed by summing category-level respondent counts. Wilson score confidence intervals were calculated for all proportions. The database is organized into nine relational sheets and is accompanied by an interactive web application.
Results: Out of the 196 tests covered, 19 have pooled estimates from both studies (total n=781 respondents). Pooled analysis confirms that cardiac biomarkers, blood gases, coagulation parameters, electrolytes (potassium, sodium),

***Corresponding author: Hikmet Can Çubukçu**, MD, PhD, MSc, EuSpLM, Department of Medical Biochemistry, Sincan Training and Research Hospital, Gökçek Mah. 250 Cad. 2/A Sincan, Ankara, Türkiye, E-mail: hikmetcancubukcu@gmail.com. https://orcid.org/0000-0001-5321-9354
Oswald Sonntag, Independent Consultant, Munich, Germany. https://orcid.org/0009-0003-1579-0412
Nathan Lorde, Department of Clinical Chemistry, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. https://orcid.org/0009-0006-3501-7218
Charles R. Lefèvre, Laboratoire de Biochimie-Toxicologie, CHU de Rennes, Rennes, France. https://orcid.org/0000-0001-8887-376X
Chiara Cosma, Department of Medicine, University of Padova, Padua, Italy. https://orcid.org/0000-0002-0678-0277
Annelien Van Dalem, Department of Clinical Biology, Universitair Ziekenhuis Brussel, Brussels, Belgium. https://orcid.org/0000-0003-2304-5298
Panagiotis N. Kanellopoulos, Central Laboratory, General Hospital of Athens, "Laiko" Athens, Greece. https://orcid.org/0000-0002-1040-7176
Marija Kocijancic, Institute of Clinical Chemistry and Pathobiochemistry, University Hospital Tübingen, Tübingen, Germany. https://orcid.org/0000-0003-1502-5705
Guilaine Boursier, CHU Montpellier, Department of Molecular Genetics and Cytogenomics, National Reference Laboratory of Autoinflammatory Diseases, French National Reference Center of Autoinflammatory Diseases and Amyloidosis (CeRéMAIA), IRMB, INSERM U1183, Univ Montpellier, Montpellier, France. https://orcid.org/0000-0002-2903-3135
Pika Meško Brguljan, Department of Clinical Chemistry, University Clinic for Respiratory and Allergic Diseases, Golnik, Slovenia. https://orcid.org/0000-0002-4945-6637
Neda Milinkovic, Department of Medical Biochemistry, Faculty of Pharmacy, University of Belgrade, Belgrade, Serbia. https://orcid.org/0000-0002-2641-9817
Marith van Schrojenstein Lantman, SKML, Foundation for Quality Assessment in Medical Laboratory Diagnostics, Nijmegen, The Netherlands. https://orcid.org/0000-0002-5454-990X
Marc Thelen, SKML, Nijmegen, The Netherlands; and Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands. https://orcid.org/0000-0003-1771-669X
Amitava Dasgupta, Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, USA. https://orcid.org/0000-0003-3916-1357
Giuseppe Lippi, Section of Clinical Biochemistry, University of Verona, Verona, Italy. https://orcid.org/0000-0001-9523-9054
Mario Plebani, Department of Laboratory Medicine, University Hospital – Padova, University of Padova, Padova, Italy. https://orcid.org/0000-0002-0270-1711

and therapeutic drug concentrations (digoxin, lithium) consistently receive the highest severity ratings, while certain endocrine markers receive lower scores.

Conclusions: The SoH evidence database offers the first publicly available, structured repository of SoH assessments for clinical laboratory tests. It supports compliance with risk management requirements, informs risk-based strategies per CLSI EP23-ED2, and provides an evidence-based foundation for laboratory risk analysis.

Keywords: severity of harm; risk management; laboratory errors; patient safety; quality

Introduction

Clinical laboratories play a critical role in patient care, with many clinical decisions influenced by laboratory test results [1]. When laboratory errors translate into incorrect test results, the consequences can range from trivial inconvenience up to patient death, depending on the analyte, the magnitude and direction of the error, and the clinical context in which the result is interpreted [2]. Severity of harm (SoH) categories capture this 'impact' dimension of risk by being a practical simple overall risk metric of erroneous potential severity of harm associated with erroneous results of measurands.

Recent revisions to international quality standards have placed patient risk at the center of laboratory quality management [3]. In its introduction, International Organization for Standardization (ISO) 15189:2022 identifies the reduction of potential patient harm as an important benefit of the risk-based approach adopted in the latest revision of the standard. Clause 5.6 requires laboratories to identify, evaluate, and mitigate risks of harm to patients arising from the total testing process [4]. ISO 22367:2026 mandates that foreseeable harms be identified and classified according to the severity of their impact [5].

ISO 15189 paragraph 7.1 states that 'The identified risks and effectiveness of the mitigation processes shall be monitored and evaluated according to the potential harm to the patient.' In paragraph 7.3.7.2 clause d states that IQC shall be performed at a frequency that is based on the stability and robustness of the examination method and the risk of harm to the patient from an erroneous result. Finally, paragraph 7.5. states in about non-conforming work in its clause c that risk of potential harm shall be considered when deciding whether examinations are halted, and reports withheld and in clause f of that paragraph whether requester shall be informed on previously released results.

In addition, the Clinical and Laboratory Standards Institute (CLSI) EP23 (2nd edition) (CLSI EP23-ED2) framework for risk-based quality control explicitly uses SoH categories to determine acceptable probabilities of patient harm and to design quality control strategies accordingly [6].

In all these requirements, the probability that a risk of harm will arise needs to be distinguished from the severity of that harm. When probability and severity are multiplied, the result represents the risk of patient harm. Because the probability component of this equation is quantitative, it is tempting to also quantify the severity component, thereby enabling numerical calculations.

Despite these normative requirements, the task of assigning SoH categories to individual laboratory tests remains challenging. SoH determination is by nature subjective, and there has been virtually no published guidance to support this process until recently. Two landmark studies have addressed this gap: Çubukçu et al. assessed SoH for 195 tests via a survey of 514 specialists in Turkey [7] (hereafter referred to as CUB2024), and Peltier et al. surveyed 267 laboratory professionals from 43 countries on 20 routine analytes (hereafter referred to as PEL2025) [8]. Both approaches categorize individual tests into a canonical five-level ordinal scale aligned with the ISO 14971:2019 [9], ISO 24971:2020 [10], and CLSI EP23-ED2 [6] (Table 1). Both studies represent important individual contributions, but their data exist in separate publications with differing respondent populations, and test panels. Laboratories seeking to use these data for risk management must currently navigate these differences manually. Moreover, the scientific community lacks a consolidated, structured resource that enables direct comparison, pooled analysis, and systematic exploration of SoH evidence across studies.

To address this need, we released the Severity of Harm Evidence Database (version 1.1), an open relational database that integrates data from both studies into a unified framework. The database applies a harmonized canonical scale, performs direct count pooling for the 19 tests common to

Table 1: Definitions of severity of harm categories.

Scale	Category	Definition
1	Negligible	Results in inconvenience or temporary discomfort
2	Minor	Results in temporary injury or impairment not requiring medical or surgical intervention
3	Serious/major	Results in injury or impairment that requires medical or surgical intervention
4	Critical	Results in permanent impairment or irreversible injury
5	Catastrophic/fatal	Results in death

both studies, and provides derived statistics including Wilson score confidence intervals. An accompanying interactive web application enables users to explore, compare, and export data. This paper describes the database design, methodology, and content.

Methods

Data sources

The database integrates two independent survey-based studies (Table 2). Çubukçu et al. surveyed 514 specialist physicians in Turkey including 75 medical biochemistry specialists and 439 clinicians across 30 specialties, who rated 195 laboratory tests on a numeric 1–5 severity scale aligned with the ISO 22367:2026 and ISO 14971:2019 standards, with individual-level raw responses yielding full category counts per test and a reported 91 % agreement between medical biochemists and clinicians [7]. Peltier et al. surveyed 267 clinical laboratory professionals from 43 countries (68.2 % European), who rated 20 routine blood analytes using the same five-level categorical scale (negligible through catastrophic), with data reported as percentage distributions from which category counts were derived by multiplying each percentage by the effective number of valid respondents [8].

Database design and architecture

The SoH Evidence Database (v1.1) is implemented as a multi-sheet Excel workbook (Supplementary Material, Appendix) containing nine relational sheets. The information from the database is rendered in a web application, which provides interactive visualization and export capabilities.

Table 2: Comparison of source studies integrated in the SoH evidence database.

Characteristic	Çubukçu et al. [20]	Peltier et al. [8]
Study ID	CUB2024	PEL2025
Country/region	Turkey (multi-center)	International (43 countries)
Respondents, n	514	267
Respondent profile	Specialist medical doctors (including medical biochemists)	Clinical laboratory professionals
Number of tests	195	20
Scale type	Categorical scale	Categorical scale

Schema overview

The database is organized into nine interconnected sheets (Supplementary Material 1). The *study* sheet contains study-level metadata including the digital object identifier (DOI), publication year, country or region, survey design, respondent group description, total respondent count, number of analytes assessed, scenario framing (worst-case), the verbatim question text, scale identifier, data granularity indicator, and notes. The *scale* sheet defines the labels for each of the five severity levels (1 through 5) as used in each source study and in the canonical harmonized scale. The *scale_map* sheet provides the crosswalk mappings from each study's original scale to the canonical 1–5 scale; all current mappings are direct one-to-one correspondences reflecting the conceptual alignment between the two studies' severity definitions.

The *test_registry* sheet serves as the canonical list of 196 unique laboratory tests. Each entry has a unique test identifier, a canonical name, specimen matrix, clinical category, and Boolean flags indicating whether the test is present in each source study. The *study_test_result* sheet contains the per-study, per-test SoH distributions. Each row records the five category counts (c1 through c5), five proportions (p1 through p5, expressed as percentages), and derived statistics including the mean, median, mode, interquartile range boundaries, critical/catastrophic risk probability, and catastrophic risk probability.

The *pooled_analysis* sheet presents the SoH results for each of the 196 tests. For the 19 tests present in both studies, pooled counts and proportions are calculated via direct count pooling. For the 177 tests appearing in only one study, single-study values are carried forward. The *stratum* sheet defines respondent subgroups within each study: CUB2024 includes all respondents (n=514), clinicians (n=439), and medical biochemistry specialists (n=75) [7]; PEL2025 includes all respondents (n=267) [8].

The *risk_of_bias* sheet contains a structured risk-of-bias assessment for each study across four domains: respondent representativeness, scale validity, scenario clarity, and data granularity. Ratings use a three-level system (Low, Moderate, High) with narrative justifications. Finally, the *changelog* sheet documents all corrections and updates applied to the database, including affected sheets and the nature of each fix.

Test registry and harmonization

A critical step in constructing the database was harmonizing test identities across studies. The two source studies used different naming conventions in some cases. For example,

Çubukçu et al. referred to “PT (plasma)” [7] while Peltier et al. used “Prothrombin time” [8]. These were mapped to a single canonical entry.

Of the 196 tests in the registry, 195 originate from CUB2024 [7], 20 from PEL2025 [8], and 19 are common to both studies [7, 8]. The 19 overlapping tests span key clinical domains including renal function (creatinine), electrolytes (sodium, potassium, phosphorus), cardiac biomarkers (cardiac troponin), coagulation (prothrombin time [PT], D-dimer, fibrinogen), liver enzymes (alanine aminotransferase [ALT], pancreatic enzymes (lipase), proteins (albumin, total protein), hormones (thyroid stimulating hormone [TSH], tumor markers (cancer antigen 19-9 [CA 19-9]), therapeutic drug monitoring (digoxin), iron studies (iron), and hematology (hemoglobin, platelet count). One additional test, glucose, bridges the diabetes/glucose category. Table 3 presents these 19 overlapping tests with their pooled statistics.

Table 3: The 19 laboratory tests common to both source studies, with pooled mode SoH and total respondent counts.

Analyte (specimen)	Category	n (pooled)	Mode SoH	P(≥4) %
hs-Troponin I&T (high sensitivity) (serum)	Cardiac biomarkers	653	5	84.84
Digoxin (serum)	Therapeutic drug monitoring	532	5	74.81
Potassium (serum)	Electrolytes	741	5	74.49
Platelet count (blood)	Hematology CBC	744	5	65.46
PT (prothrombin time) (plasma)	Coagulation	716	4	65.08
Hemoglobin (blood)	Hematology CBC	745	5	63.36
D-Dimer (plasma)	Coagulation	668	4	62.28
Glucose (serum/plasma)	Diabetes/Glucose	763	5	58.85
Sodium (serum)	Electrolytes	738	4	56.5
Creatinine (serum)	Renal function	761	3	50.99
Fibrinogen (plasma)	Coagulation	612	3	45.42
CA 19-9 (cancer antigen 19-9) (serum)	Tumor markers	562	3	39.68
Lipase (serum)	Pancreatic	668	3	35.93
TSH (thyroid stimulating hormone) (serum)	Thyroid	705	3	34.18
ALT (alanine aminotransferase) (serum)	Liver/hepatic	753	3	29.61
Phosphorus (inorganic) (serum)	Electrolytes	672	3	20.24
Albumin (serum)	Liver/hepatic	713	3	19.07
Total protein (serum)	Liver/hepatic	675	3	13.78
Iron (serum)	Iron studies	674	2	8.31

Mode SoH, most frequent severity category in pooled data; P(≥4), combined proportion of respondents rating the test as critical or catastrophic; n, total number of valid responses across both studies.

Pooling methodology

Canonical scale

The database adopts a canonical five-level ordinal scale aligned with the ISO 14971:2019 [9], ISO 24971:2020 [10], and CLSI EP23-ED2 [6]. Level 1 (Negligible) corresponds to temporary discomfort or inconsequential injury; Level 2 (Minor) corresponds to temporary injury not requiring professional medical intervention; Level 3 (Serious) corresponds to injury or impairment requiring professional medical intervention; Level 4 (Critical) corresponds to permanent impairment or irreversible injury; while Level 5 (Catastrophic) corresponds to life-threatening injury or death.

Both source studies used five-level scales that map directly to this canonical framework [7, 8]. The definitions provided to respondents in both studies were conceptually equivalent. The *scale_map* table in the database documents these direct one-to-one correspondences.

Direct count pooling

For the 19 tests present in both studies, pooled proportions are calculated using the direct count pooling method. This approach sums the raw category counts across studies before computing proportions, thereby naturally weighting each study by its effective sample size for that test. Formally, for each test t and severity category k (where k ranges from 1 to 5), the pooled count $C_k(t)$ is defined as the sum of category counts $c_{k,s,t}$ across all contributing studies s:

$$C_k(t) = \sum_s c_{k,s,t} \text{ for } k = 1, 2, 3, 4, 5 \quad (1)$$

The pooled total n(t) is the sum of all five pooled category counts:

$$\mathrm{n}(t) = \sum_{k=1}^{5} C_k(t) \quad (2)$$

The pooled proportion $p_k(t)$ is then calculated as $C_k(t)$ divided by n(t):

$$p_k(t) = C_k(t) \,/\, \mathrm{n}(t) \quad (3)$$

For the 177 tests present in only one study, the single-study counts and proportions are carried forward directly into the *pooled_analysis* sheet with a pooling_status flag of “SINGLE_STUDY” to distinguish them from truly pooled results.

Confidence intervals

Ninety-five percent confidence intervals (95 % CI) for all proportions are calculated using the Wilson score method for binomial proportions [11].

Derived metrics

For each test, the database provides several derived statistics calculated from the pooled (or single study) proportions. The mean SoH score is calculated as the sum of each category code (1 through 5) multiplied by its proportion. The median SoH is the 50th percentile of the ordinal distribution. The mode SoH is the most frequently chosen severity category.

Two additional risk metrics are provided. P(≥4), the critical/catastrophic risk, represents the combined proportion of respondents rating the test as critical (4) or catastrophic (5)., while P(=5), the catastrophic risk, represents the proportion of respondents rating the test at the maximum severity level.

Risk of bias assessment

Each source study was assessed for risk of bias across four domains relevant to SoH survey methodology. Table 4 summarizes the assessment.

The most notable risk factor across both studies is geographic representativeness. Çubukçu et al. collected data exclusively from Turkey, though from a diverse range of medical specialties including both clinicians and medical biochemistry specialists [7]. Peltier et al. achieved broader geographic coverage with 43 countries represented, but respondents were heavily weighted toward Europe (68.2 %), with four countries (France, Italy, Germany, and Spain) accounting for 46.4 % of all participants. Neither study included substantial representation from Asia (2.3 % in Peltier) and the Americas (1.5 % in Peltier). Despite these limitations, the 91 % agreement rate between medical biochemistry specialists and clinicians reported by Çubukçu et al. suggests that SoH assessments may be relatively stable across professional backgrounds, even if geographic variation remains underexplored [7].

Table 4: Risk of bias assessment summary.

Domain	CUB2024	PEL2025	Key considerations
Respondent representativeness	Moderate	Moderate	CUB2024: Turkey only, diverse specialties. PEL2025: 43 countries but 68 % European, lab professionals only.
Scale validity	Low (good)	Low (good)	Both use ISO-aligned 5-level scales with explicit definitions.
Scenario clarity	Low (good)	Low (good)	CUB2024 implied average unfavourable scenario framing. PEL2025 implied most unfavourable scenario framing.
Data granularity	Low (good)	Moderate	CUB2024: Full raw counts. PEL2025: Percentages with derived counts (rounding).

Regarding data granularity, CUB2024 benefits from full individual-level raw responses (514 respondents × 195 tests), enabling precise category counts for each test [7]. PEL2025 reported data as percentage distributions in the published table, from which category counts were derived by multiplication [8]. While Cubukcu et al. used scenario implying average impact of erroneous results [7], Peltier et al. used most unfavourable scenario [8].

Web application

The database is accompanied by an interactive web application (https://severityofharm.streamlit.app/) that provides a user interface for data exploration (Figure 1). The application consists of seven functional modules. The Dashboard module presents an overview of key metrics including total studies, unique tests, total respondents, and pooled tests, alongside aggregate SoH distribution charts. The Pooled Analysis module (Figure 2) provides a filterable table and visualizations of pooled results, including distribution bar charts and heat maps that can illustrate the tests with the different SoH distributions.

The Test Explorer module (Figure 3) enables detailed per-test investigation, showing the full SoH distribution with Wilson confidence intervals, individual study contributions, and all derived statistics for any selected test. The Study Comparison module (Supplementary Material 2) provides side-by-side comparison of study-level results for the 19 overlapping tests, facilitating assessment of inter-study consistency. The Risk of Bias module (Supplementary Material 3) presents a visual heatmap of risk-of-bias ratings across studies and domains. The Methodology module (Supplementary Material 4) documents the mathematical basis of the pooling method, confidence interval formulae, and derived metrics with rendered equations. Finally, the Data Download module allows export of individual tables in CSV format or the complete database as an Excel file.

The application was developed using Python programming language [12], with packages including Streamlit (version 1.30.0) [13], Pandas (version 2.0.0) [14], Plotly (version 5.18.0) [15], openpyxl (version 3.1.0) [16], and python-docx (version 1.0.0) [17]. The application reads that data directly from the soh_database_db.xlsx file and caches data for performance. The code script and the database is available at https://github.com/hikmetc/soh_database.

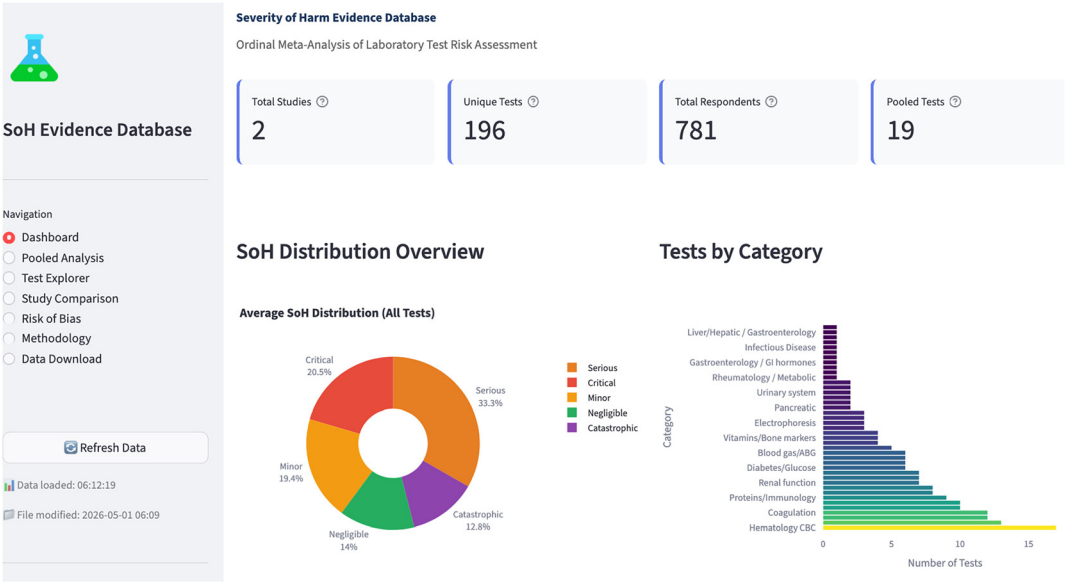

Figure 1: Dashboard of severity of harm score database.

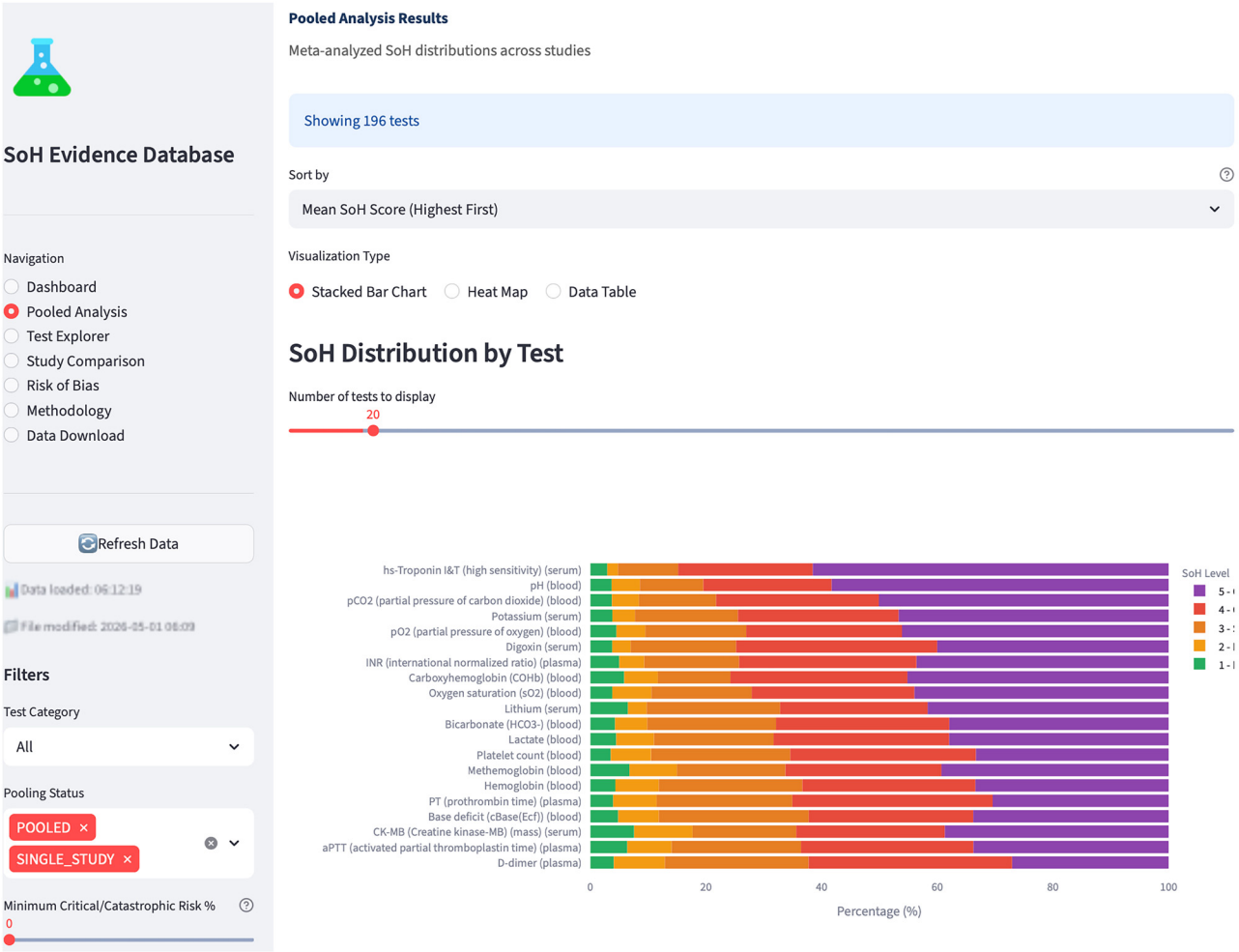

Figure 2: The pooled analysis module.

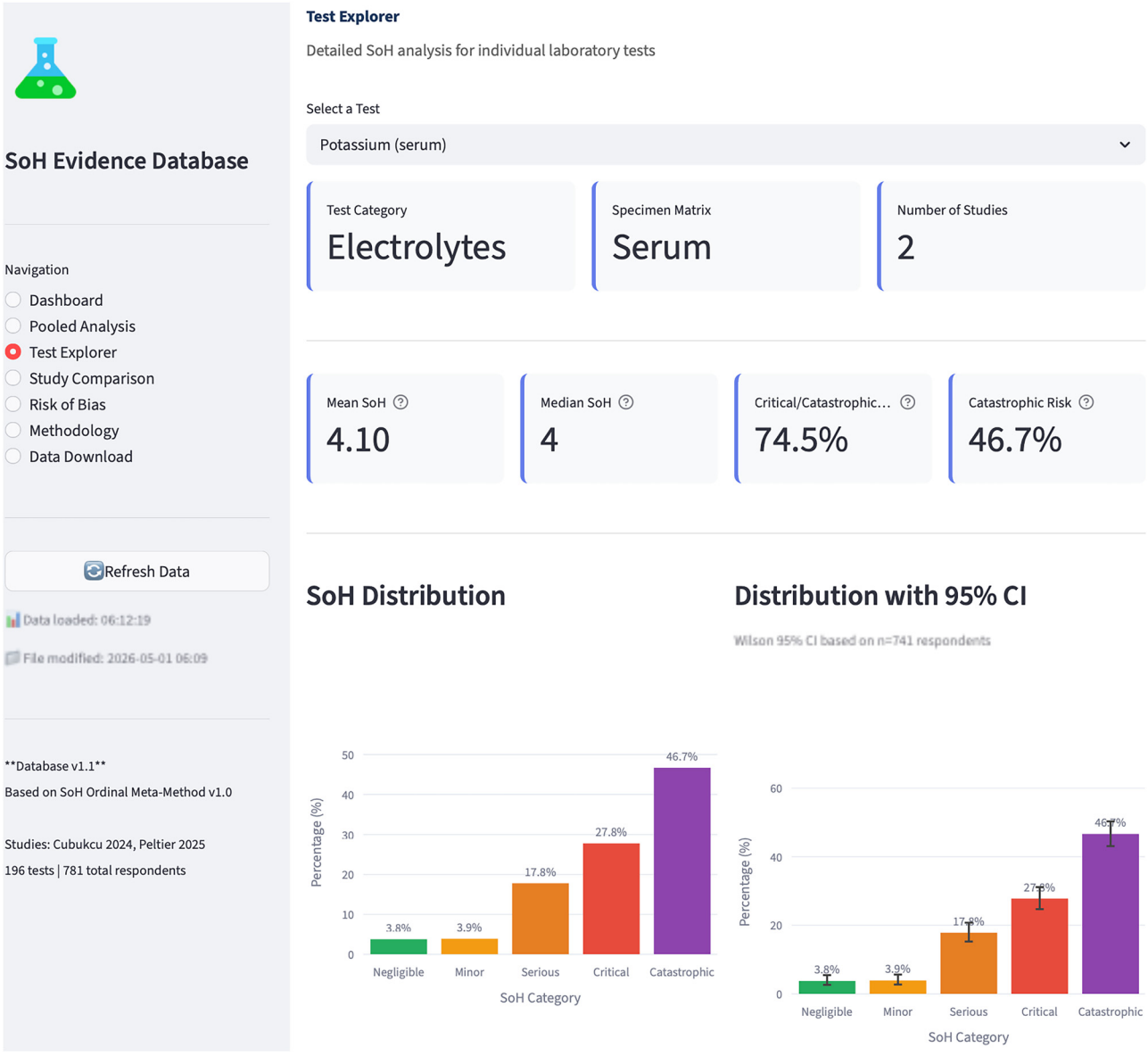

Figure 3: The test explorer module.

Results

Analysis of the pooled database reveals several consistent patterns across both source studies. Tests associated with the highest perceived severity fall into a small number of clinical categories. Cardiac biomarkers, most notably cardiac troponins, receive the highest ratings: in the pooled data, 84.8 % of respondents rated these biomarkers as critical or catastrophic, making them the test with the greatest perceived potential for patient harm from erroneous results. Blood gas parameters (pH, partial oxygen pressure (pO_2), partial carbon dioxide pressure (pCO_2), oxygen saturation) also received very high severity scores, with mode scores of 5 (catastrophic) for pH, pCO_2, and pO_2. Coagulation tests including activated partial thromboplastin time (APTT), PT/international normalized ratio (INR), and D-dimer consistently received high ratings across both studies.

Some electrolytes emerged as another high-severity group. Potassium was rated as critical or catastrophic by 74.5 % of pooled respondents, consistent with a case report of pseudohyperkalemia due to hemolysis, in which treatment for hyperkalemia led to cardiac arrest [18]. Therapeutic drug monitoring tests for narrow-therapeutic-index drugs similarly received high severity ratings: digoxin had a pooled P(≥4) of 74.8 %, reflecting the well-known risk of toxicity or therapeutic failure from erroneous drug level reporting [19]. What all tests

in the highest category have in common is that the release of the laboratory result is followed up by medical treatment shortly. That means that there is only a very short window of opportunity to correct an incorrect result and inform the requesting physician as intended by clause ISO15189 7.5.f [4].

Tests in the moderate-severity range included analytes such as creatinine (pooled P(≥4) of 51.0 %), glucose (58.9 %), and haemoglobin (63.4 %). These tests have significant clinical decision-making value; however, may allow for some clinical verification before irreversible harm occurs particularly if not performed as STAT tests. Nevertheless, the potential severity of harm may increase when these tests are performed in point-of-care or STAT settings, where the intrinsic purpose is immediate clinical decision-making once the result becomes available. In such contexts, it may be appropriate to assign the test to one higher severity category, reflecting the reduced opportunity to detect, verify, or correct an erroneous result before clinical action is taken.

Importantly, the cross-study concordance for the 19 overlapping tests was high (Pearson correlation coefficient=0.957) (Supplementary Material 2). Both studies classified cardiac troponins and potassium at the highest SoH level, and both identified iron as the analyte with the lowest perceived severity among those assessed in common. This concordance is notable given the differences in respondent populations (Turkish specialists vs. international laboratory professionals) and geographic scope (single country vs. 43 countries). Nevertheless, absolute severity assignments may still be influenced by clinical setting, test urgency, respondent expertise, and local healthcare practices. Therefore, these findings support the external consistency of the proposed severity framework, but they should not be interpreted as definitive evidence that severity judgements are universally fixed across all clinical contexts.

Additionally, to assess the degree of conformance between the two studies for the 19 laboratory tests rated by both, a linear-weighted Cohen's kappa (κ) analysis was performed. Each study was treated as an independent rater and each shared test as a single observation. Three approaches were evaluated for deriving a categorical SoH rating from each study's distribution: the modal category (the most frequently assigned rating), the median category (the lowest category at which the cumulative distribution reached or exceeded 50 %), and the mean SoH score rounded to the nearest integer. Results were interpreted according to the classification proposed by McHugh (2012): κ≤0.20=none, 0.21–0.39=minimal, 0.40–0.59=weak, 0.60–0.79=moderate, 0.80–0.90=strong, and >0.90=almost perfect agreement. For mean, median, and modal SoH scores the Kappa values were determined as 0.757 (moderate), 0.790 (moderate) and 0.550 (weak). The weak conformance observed for modal SoH scores (κ=0.550) likely reflects the sensitivity of the mode to distributional skewness and bimodality rather than genuine inter-study divergence in severity consensus, as the modal value is disproportionately influenced by the single most frequent response and fails to capture the overall shape of the rating distribution. In contrast, both the mean- and median-based approaches yielded moderate conformance (κ=0.757 and κ=0.790, respectively), with the three discordant tests (creatinine (serum), total protein (serum), and sodium (serum) each differing by only one severity level considering mean and median values. The moderate rather than strong kappa values reflect the conservative nature of the statistic, which penalises any categorical mismatch regardless of magnitude, and should be interpreted alongside the Pearson r=0.957 and the 84 % exact categorical agreement as collectively indicating robust cross-study concordance.

Discussion

The SoH Evidence Database addresses the lack of a consolidated, evidence-based resource for severity of harm assessment, which is a practical need in laboratory medicine. While individual laboratories have always been expected to perform their own risk analyses, the subjective nature of SoH determination has made this a challenging and often inconsistent exercise. By integrating data from two independent international surveys and providing harmonized, pooled estimates with CI, this database offers laboratories an evidence based, objective starting point for their risk management processes.

Several aspects of the database design deserve discussion. The direct count pooling method was chosen for its simplicity and transparency. Because both studies used identical five-level scales, and because the survey questions were similar, the data can be meaningfully combined by summing counts. This approach naturally weights each study in proportion to its sample size for each test, which is appropriate given the comparable methodological quality of both studies.

The database is designed for extensibility. The relational schema can accommodate additional studies as they become available, and the pooling methodology can be applied to any new study that uses a compatible five-level SoH scale. The *changelog* sheet ensures that all modifications are transparently documented. As the field matures and additional surveys are conducted in underrepresented regions and for additional analytes, the database can serve as the central repository for the growing body of SoH evidence.

The practical application of this database will address laboratory quality management. For ISO 15189:2022 compliance [4], the database provides published evidence of the SoH categories of the tests that laboratories can reference

when establishing their risk-based quality management. For CLSI EP23 risk-based quality control, the pooled SoH categories can be used directly to determine acceptable probabilities of patient harm and to design quality control strategies with predicted probabilities of harm. The risk management index, which compares the predicted probability of harm to an acceptable level, depends critically on the assigned severity category; the database provides an evidence-based foundation for this assignment [6]. The database also complements the tools, which supports risk-based quality control practices using severity of harm scores [20].

Some limitations should be acknowledged. The database currently draws from only two source studies, and geographic representation is skewed toward Europe and Turkey. In addition, direct cross-study comparison was possible for only 19 common analytes, which limits the strength and generalizability of the concordance analysis. Neither study included substantial participation from Asia, the Americas, or Oceania. Both studies relied on perceived opinion obtained by surveys rather than real cases on actual patient outcomes, and the perceived severity of harm may differ from observed harm in clinical practice. It is important to recognize that the specific intended clinical use of a measurand can affect the severity of potential harm (e.g., neonatal bilirubin monitoring vs. evaluating jaundice in adults). Indeed, descriptive studies of real laboratory failures indicate that most errors do not actually lead to patient harm, as errors are often detected before reaching the patient by laboratory personnel or by experienced clinical doctors; however, this does not diminish the importance of SoH estimation for risk management purposes, as emphasized by both CLSI EP23 [6], ISO 14971:2019 [9], and ISO 24971:2020 [10].

Conclusions

The Severity of Harm Evidence Database v1.1 represents the first structured, multi-study resource for ordinal meta-analysis of laboratory test risk assessment. By integrating data published literature, the database provides harmonized SoH distributions, pooled estimates, and derived risk metrics for 196 laboratory tests across 41 categories. The accompanying web application enables interactive exploration, study comparison, and data export.

The high concordance between the two source studies – despite differences in geography, and respondent populations, provides confidence in the robustness of the pooled estimates for the 19 overlapping tests. As additional SoH survey data are expected to become available from underrepresented regions and for additional analytes in the near future, the database can be extended to incorporate new studies, refine pooled estimates, and broaden the evidence base for patient risk management in laboratory medicine.

Acknowledgments: The present study is the product of a joint study of the Committee: Laboratory Error Database and the Committee: Accreditation and ISO/CEN Standards of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM).
Research ethics: Not applicable.
Informed consent: Not applicable.
Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Use of Large Language Models, AI and Machine Learning Tools: None declared.
Conflict of interest: The authors state no conflict of interest.
Research funding: None declared.
Data availability: Not applicable.

References

1. Rohr UP, Binder C, Dieterle T, Giusti F, Messina CG, Toerien E, et al. The value of in vitro diagnostic testing in medical practice: a status report. PLoS One 2016;11:e0149856.
2. Signori C, Ceriotti F, Sanna A, Plebani M, Messeri G, Ottomano C, et al. Process and risk analysis to reduce errors in clinical laboratories. Clin Chem Lab Med 2007;45:742–8.
3. Linko S, Boursier G, Bernabeu-Andreu FA, Dzneladze N, Vanstapel F, Brguljan PM, et al. EN ISO 15189 revision: EFLM committee accreditation and ISO/CEN standards (C: A/ISO) analysis and general remarks on the changes. Clin Chem Lab Med 2025;63:1084–98.
4. International Organization for Standardization. ISO 15189:2022. Medical laboratories – requirements for quality and competence. Geneva: International Organization for Standardization; 2022.
5. International Organization for Standardization. ISO 22367:2026 medical laboratories – application of risk management to medical laboratories. Geneva: International Organization for Standardization; 2026.
6. Clinical and Laboratory Standards Institute. CLSI document EP23-ED2: 2023 – laboratory quality control based on risk management, 2023. Wayne, Pennsylvania, US: Clinical and Laboratory Standards Institute.
7. Çubukçu HC, Cihan M, Alp HH, Bolat S, Zengi O, Uçar KT, et al. Measuring the impact: severity of harm from laboratory errors in 195 tests. Am J Clin Pathol 2025;163:453–63.
8. Peltier L, Van Aelst S, Peeters B, Raimbourg JB, Yundt-Pacheco J. Patient risk management in laboratory medicine: an international survey to assess the severity of harm associated with erroneous reported results. Clin Chem Lab Med 2025;63:1548–54.

9. International Organization for Standardization. ISO 14971:2019 medical devices – application of risk management to medical devices. Geneva: International Organization for Standardization; 2019.
10. International Organization for Standardization. ISO 24971:2020 medical devices – guidance on the application of ISO 14971. Geneva: International Organization for Standardization; 2020.
11. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. Am Statistician 1998;52: 119–26.
12. Van Rossum G, Drake FL. Python 3 reference manual. Scotts Valley, CA: CreateSpace; 2009.
13. Streamlit • A faster way to build and share data apps. Available at: https://streamlit.io/ [Accessed 12 Apr 2023].
14. Team TPD. Pandas-dev/pandas. Zenodo: Pandas; 2020.
15. Collaborative data science publisher: Plotly Technologies Inc. Available at: https://plot.ly [Accessed 21 Jun 2023 2015].
16. Gazoni E, Clark C. Openpyxl: A Phyton library to read/write Excel (Version 3.1.5) 2024.
17. Canny S. Python-docx (Version 1.2.0) 2025. https://pypi.org/project/python-docx/
18. Devera JL, Barnes DK, Lewis WR. AHRQ patient safety network (PSNet) collection A fatal twist in pseudohyperkalemia. WebM&M: case studies. Rockville (MD): Agency for Healthcare Research and Quality (US); 2019.
19. Kalra D, Tesfazghi M. Falsely elevated digoxin levels in patients on enzalutamide. Circ Heart Fail 2020;13:e007008.
20. Çubukçu HC. QC constellation: a cutting-edge solution for risk and patient-based quality control in clinical laboratories. Clin Chem Lab Med 2024;62:2185–97.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2026-0875).