Science & Justice 65 (2025) 101304

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

Review

A guide to ISO 21043 *Forensic Sciences* from the perspective of the forensic-data-science paradigm

Geoffrey Stewart Morrison [a,b,*], Simon Elliott [c,d], June Guiness [e], Lisa Sonden [f], Denise Syndercombe Court [d]

[a] Forensic Data Science Laboratory, Aston University, Birmingham, UK
[b] Forensic Evaluation Ltd, Birmingham, UK
[c] Elliott Forensic Consulting Ltd, Birmingham, UK
[d] Department of Analytical, Environmental and Forensic Sciences, King's College London, London, UK
[e] Office of the Forensic Science Regulator, Birmingham, UK
[f] Defence Science and Technology Laboratory, Wiltshire, UK

ARTICLE INFO

Keywords:
Forensic science
Interpretation
ISO standard
Likelihood ratio
Reporting
Vocabulary

ABSTRACT

ISO 21043 is a new international standard for forensic science. It provides requirements and recommendations designed to ensure the quality of the forensic process. It includes Parts on: 1 vocabulary; 2 recovery, transport, and storage of items; 3 analysis; 4 interpretation; and 5 reporting. We provide a guide to ISO 21043 from the perspective of the forensic-data-science paradigm, which involves the use of methods that are transparent and reproducible, are intrinsically resistant to cognitive bias, use the logically correct framework for interpretation of evidence (the likelihood-ratio framework), and are empirically calibrated and validated under casework conditions. The guide focuses primarily on vocabulary, interpretation, and reporting, and on providing guidance for forensic-service providers and examiners who want to implement methods that are both consistent with the forensic-data-science paradigm and conformant with ISO 21043.

1. Introduction

ISO 21043 *Forensic Sciences* is a new international standard for *forensic* science published by the International Organization for Standardization (ISO).[1] Many *forensic service providers* currently seek accreditation to ISO/IEC 17020 *Conformity assessment — Requirements for the operation of various types of bodies performing inspection* for crime-scene investigation or to ISO/IEC 17025 *General requirements for the competence of testing and calibration laboratories* for laboratory-based activities, but neither of these standards is specific to *forensic* science.[2] ILAC G19 *Modules in a Forensic Science Process* [2] provides guidance to *forensic service providers* implementing ISO/IEC 17020 or ISO/IEC 17025, but this is a supplement to those non-forensic-science-specific standards. In contrast, ISO 21043 is specific to *forensic* science and could therefore be considered a better fit in terms of providing requirements and recommendations specifically related to *forensic* activities. ISO 21043 provides requirements and recommendations designed to ensure the quality of the *forensic process*, and covers both investigative and evidential applications.

ISO 21043 currently consists of the following Parts:

- ISO 21043-1:2025 *Forensic Science – Part 1: Vocabulary*
- ISO 21043-2:2018 *Forensic Sciences – Part 2: Recognition, recording, collection, transport and storage of items*
- ISO 21043-3:2025 *Forensic Sciences – Part 3: Analysis*
- ISO 21043-4:2025 *Forensic Sciences – Part 4: Interpretation*
- ISO 21043-5:2025 *Forensic Sciences – Part 5: Reporting*

Each Part is published as a separate document. The Parts are available from the ISO website (https://www.iso.org/standards.html) or

* Corresponding author at: Forensic Data Science Laboratory, Aston University, Birmingham, UK.
E-mail address: geoff-morrison@forensic-evaluation.net (G.S. Morrison).

[1] In the present document, words that are defined terms in ISO 21043 are set in italics when they represent those defined terms. If a word that is a defined term is, instead, used with its everyday meaning, it is not set in italics. In the present document, italics are also used for the titles of standards and guidelines.

[2] In England & Wales, for certain branches of forensic science, the Forensic Science Regulator requires accreditation to ISO/IEC 17020 or ISO/IEC 17025 and to the Regulator's *Code of Practice* [1].

https://doi.org/10.1016/j.scijus.2025.101304
Received 21 May 2025; Received in revised form 25 June 2025; Accepted 27 June 2025
Available online 28 June 2025
1355-0306/© 2025 The Author(s). Published by Elsevier B.V. on behalf of The Chartered Society of Forensic Sciences. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

from national standards organizations' websites. Searching on the website for "ISO 21043" should return links to all available Parts of the standard. They will also be accessible via libraries that subscribe to catalogues of national and international standards. The content of Part 1 is available free of charge at https://www.iso.org/obp/ui/en/#iso:std:iso:21043:-1:ed-2:v1:en.

ISO 21043 can be used in a number of ways to meet differing end-user needs. A *forensic service provider* could use the standard as part of their quality-management system, conform to the requirements and recommendations of the standard (or to particular Parts of the standard) and monitor and document conformance. This could be done in order to meet a *forensic service provider's* internal goal of improving the quality of their *forensic* activities. Regulatory bodies, customers commissioning work, or courts receiving *reports* or *testimony* could also require *forensic service providers* to be conformant with ISO 21043 (or with particular Parts of ISO 21043). A *forensic service provider* could self determine and self declare whether they are conformant. In the future, third-party certification and accreditation options could become available, allowing *forensic service providers* to include ISO 21043 or Parts thereof in the scope of their certification or accreditation. Since ISO 21043 is new, it will probably take time for such options to become available, if they do become available. Certification bodies could develop programmes to certify forensic-service providers as conformant to ISO 21043 or Parts thereof. Accreditation bodies could develop programmes to assess *forensic service providers* with respect to ISO 21043 or Parts thereof as an extension to the scope of their accreditation under ISO/IEC 17020 or ISO/IEC 17025. ISO 21043 is not an accreditation standard, so *forensic service providers* cannot be accredited to ISO 21043 as a stand-alone standard.

ISO 21043 was developed by ISO Technical Committee 272 Forensic Sciences (ISO/TC 272). ISO standards are developed through a process that formally considers comments and proposals submitted by mirror committees' of participating national standards organizations.[3] Participating national mirror committees submit concrete proposals for changes to the existing draft of the document, i.e., they submit proposals to move text from one location in the document to another, to delete text, to add text, and to alter text. Each proposal is accompanied by a comment explaining the rationale for the change. Members of the ISO committee, which consist primarily of representatives of participating national mirror committees, then discuss each proposal and vote on whether to accept it or not. If the majority of the representatives of the participating national mirror committees who are present at an ISO committee meeting vote in favour of the proposal (one vote per participating national mirror committee), the proposal is accepted. Representatives of national standards organizations with observer status, and liaisons from other ISO committees and from organizations such as the International Laboratory Accreditation Cooperation (ILAC) can attend meetings but not vote. Toward the end of the development process, there is a public-comment stage, during which interested parties can submit comments and proposals via national mirror committees. With participation of national mirror committees from multiple countries around the world, different perspectives are represented, and the process of developing a standard usually requires compromise.

ISO standards are usually reviewed and (potentially) revised on a five-year cycle. Parts 1 and 2 of ISO 21043 were first published in 2018. A revised version of Part 1 was published in 2025 and a revised version of Part 2 is scheduled for publication in 2027. Parts 3, 4, and 5 were first published in 2025. Input from *examiners* and *forensic service providers* who use ISO 21043 will be welcome when the standard is next revised. Input can be provided through national mirror committees.

The authors of the present paper have contributed to the development of ISO 21043. They are all members of the British Standards Institution's (BSI's) Forensic Science Committee (a national mirror committee), and have all represented BSI at ISO/TC 272 meetings. The first author served as Chair of the BSI Committee 2019–2025, and was ISO Project Leader for the revision of Part 1. Any and all opinions expressed in the present commentary are, however, the authors' personal opinions and they should not be construed as representing the policies or positions of BSI or ISO.

The present paper comments on ISO 21043 from the perspective of the forensic-data-science paradigm [3]. Other publications offering other perspectives on ISO 21043 include [4–6]. In the forensic-data-science paradigm, *methods* for *analysis* and *interpretation* of *forensic* evidence are based on relevant data, quantitative measurements, and machine-learning algorithms / statistical models. The *methods*:

- are transparent and reproducible;
- are intrinsically resistant to *cognitive bias*;
- use the logically correct framework for interpretation of evidence (the *likelihood-ratio* framework); and
- are empirically calibrated and *validated* under casework conditions.[4]

The present paper focuses on providing guidance for *forensic service providers* and *examiners* who want to implement methods that are both consistent with the forensic-data-science paradigm and conformant with ISO 21043. We hope it will also be of value for *forensic service providers* and *examiners* interested in being conformant with ISO 21043 but who are not yet ready to adopt the forensic-data-science paradigm.

In the remainder of the present document, each Part of ISO 21043 is discussed in turn. Part 1 Vocabulary, Part 4 *Interpretation*, and Part 5 Reporting are discussed in greater depth than Parts 2 and 3.

2. Part 1: Vocabulary

Essential to understanding the meaning of the requirements and recommendations of a standard is understanding the definitions of the terms used in the standard. Part 1 of ISO 21043 provides those terms and definitions. Below, we discuss key terms and their definitions.

Fig. 1 illustrates the *forensic process*. The component processes of the *forensic process* are "*recognition*, recording, collection, transport and storage of *items*" (abbreviated in Fig. 1 to "*recovery* of *items*"), "*analysis*", "*interpretation*", and "reporting", corresponding to Part 2, Part 3, Part 4, and Part 5 of ISO 20143 respectively. *Recovery* is defined as "*recognizing*, recording, and collecting an *item*". *Recognition* is defined as "deciding whether *items* have potential *forensic* value". *Recovery* of *items*, *analysis*, and *interpretation* (all the component processes except reporting) are together termed *examination*.

Analysis is defined as "part of the *examination* consisting of detecting, measuring, or comparing properties of *items* in order to obtain *observations*". The input to *analysis* is *items*, and the output of *analysis* is *observations*. A note to the definition of *analysis* states that "*Analysis* can be instrumental, human-perception-based, or a combination of the two."

An *item* is defined as an "object, substance, material or mark that is collected, derived or sampled as part of the *forensic process*". This includes what readers could know as traces, specimens, samples, reference material, control material, questioned-source items, known-source items, and items sampled from the relevant population. It includes *items* recovered from *scenes* and from persons of interest, but also includes other *items* used in the *forensic process*. Part 1 does not include a term that would distinguish traces from other *items*. Although not explicit in the definition, in the context of ISO 21043 (including in the definition of *scene*, which is discussed below) it is apparent that an *item* does not have to be physical – an *item* can consist of digital information.

A note to the definition of *observation* states that "An *observation* can

[3] See the ISO/TC 272 website (https://www.iso.org/committee/4395817.html) for a list of its participating national standards organizations. Not all of these were active in the development of ISO 21043.

[4] See [3] for details.

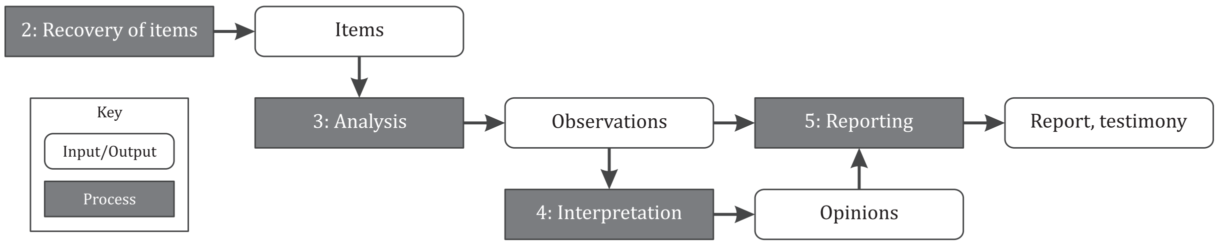

Fig. 1. The *forensic process*. The numbers in the process boxes represent the Parts of ISO 21043 to which the component processes relate.

be the result of human-perception-based *analysis*, instrumental *analysis*, or a combination of the two." Thus, *observation* covers both perceptual observations and quantitative measurements. Readers could be unused to a quantitative measurement being a meaning of *observation*.

Interpretation is defined as "part of the *examination* that uses professional judgement, logic, expertise, and relevant data and information and, if applicable, statistical models to infer the meaning of *observations* so as to provide *opinions* with respect to questions asked". The input to *interpretation* is *observations*, and the output of *interpretation* is *opinions*.

Opinion does not have its everyday meaning, but has the meaning of expert opinion. Given the definition of *interpretation*, an *opinion* can be directly an *examiner's* subjective judgement, or can be the output of a statistical model. Subjective decisions are required for deciding which statistical model to use, what data to use for training the model, etc., so professional judgement and expertise are always required. A *likelihood ratio* output by a statistical model would constitute an *opinion*. Readers could be unused to the output of a statistical model in general, or a *likelihood ratio* output by a statistical model in particular, being a meaning of *opinion*.

The input to *reporting* is *observations* or *opinions*, and the outputs of *reporting* are *reports* and, potentially, *testimony*. Some *reports* and some *testimony* present only *observations*, some only *opinions*, and some both.

Report and *testimony* form a pair of related terms. A *report* is defined as a "formal *document*, designated as a *report*, communicating the outcomes of the *examination*". The definition could appear to include tautology; however, the implication is that a document that communicates *observations* or *opinions* but that is not formal and is not designated as a *report* is not a *report*, even though in everyday language one would consider it a report. A *report* is, therefore, distinguished from, for example, preliminary communication of *observations* or *opinions*. *Testimony* is defined as a "statement which communicates information in a judicial setting".

To avoid confusion, "evidence" is not used for the input to *analysis*, or for the input to *interpretation*, or for the output of *reporting*. Neither "result" nor "finding" is used for the output of *analysis* nor for the output of *interpretation*. In the 2018 version of Part 2, "finding" was used with the meaning of *observation*, but "finding" is now deprecated. The revision of Part 2 will harmonize its vocabulary with that of the 2025 version of Part 1. "Conclusion" is not used for the output of *interpretation*. Whatever one's opinion may be about the merits of these (and other) vocabulary choices, once adopted, they remove potential confusion such as whether one is referring to the output of *analysis* or to the output of *interpretation*.

Also important to note with respect to terms and definitions are the following.

Likelihood ratio is defined as "expression of an *examiner's* assessment of the ratio of the *probabilities* of the *observations* if one of two competing *propositions* were true versus if the other *proposition* were true". The term *observation* is used where the common term in the *forensic* inference and statistics literature is "evidence", and the term *proposition* is used in preference to "hypothesis". "Hypothesis" is listed in Part 1 as an admitted term (an accepted synonym for the preferred term), but otherwise is not used in ISO 21043. Important notes to the definition of *likelihood ratio* state that, for continuously-valued data, probability density rather than *probability* is assessed, and that the term *likelihood ratio* is not intended to exclude Bayes factors.

Unfortunately, Part 1's definition for *probability* is less than helpful: *Probability* is defined as the "extent to which something is likely". In everyday language, "probable" and "likely" are synonyms, and providing a synonym does not in this case clarify meaning. In technical usage in statistics, "probability" and "likelihood" have different meanings, so defining one as being the other is incorrect and misleading. ISO 3534-1:1993 *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms* defined "probability" as "a real number in the scale 0 to 1 attached to a random event" and included a note: "It can be related to a long-run relative frequency of occurrence or to a degree of belief that an event will occur. For a high degree of belief, the probability is near 1." We think that a definition and note along these lines would have been more helpful. They cover both the frequentist concept and the subjectivist-Bayesian concept of probability. A note to the ISO 21043 definition of *probability* states that the "term includes qualitative and quantitative probabilities assigned subjectively and quantitative probabilities assigned through the use of statistical models and data."

Related to source-level *likelihood ratios*, the definition for *relevant population* is, unfortunately, incorrect and misleading. It is defined as the "set of *items* (or sources of *items*) which in the *examiner's* judgement best reflects a *proposition* and the circumstances of the case". Correctly, the *propositions* for a common-source *likelihood ratio*, for example, would be some version of "the questioned-source *item* and the known-source *item* come from the same source (a source selected at random from the relevant population)" and "the questioned-source *item* and the known-source *item* come from two different sources (each a source selected at random from the relevant population)". The relevant population does not, therefore, "reflect" a *proposition*, the relevant population is part of the specification of the *propositions*.

Fortunately, pending potential improvements when Part 1 is next revised, the problems with the definitions of *probability* and *relevant population* are unlikely to be impediments to use of the standard.

Scene and *facility* form a pair of related terms. *Scene* is defined as "place or object that either is subject to or requires *forensic examination*". A *scene* could be a non-stationary object such as a vehicle. A note states that "The *scene* can be a person, an animal or a digital location". Readers could be unused to objects, persons, animals, and non-physical locations being included in the meaning of *scene*. *Facility* is defined as "physical environment used to protect the *item* integrity, to conduct testing or to support any other aspect of the *forensic process*", and examples given are "Building, designated area, tent, (mobile) office, (mobile) laboratory, vehicle", e.g., *analysis* does not have to occur only in a permanent dedicated laboratory space in a building, it could occur at a *scene* or elsewhere.

Evaluative interpretation and *investigative interpretation*, and *proposition* and *explanation*, form contrasting pairs of terms. *Evaluative interpretation* is defined as "*interpretation* guided by a set of relevant *propositions* and aimed at generating *likelihood ratios*". *Proposition* is defined as "statement that is either true or false, the truth of which is uncertain". *Evaluative interpretation* therefore means using the *likelihood-ratio* framework. *Investigative interpretation* is defined as "*interpretation* guided by *observations* made and aimed at generating *explanations* or estimations". "Estimation" is not a defined term. Its meaning will be discussed in §5.5 below. *Explanation* is defined as "possible cause for *observations*, generated in an *investigative interpretation*". An *explanation* is, therefore, a hypothesis that is generated after considering the *observations* rather than a hypothesis that is specified before considering the *observations*. The distinction between *investigative interpretation* and *evaluative interpretation* appears to parallel what [7] referred to as "investigator" and "evaluator" modes of operation. Rather confusingly, a note to the definition of *investigative interpretation* states that it "can be used in an investigation or in a judicial setting", i.e., an *investigative interpretation* is not just for investigative purposes, it can also be used for evidential purposes.

Classification and *identification* form a pair of related terms. The definitions of these terms could differ from what readers are used to. *Classification* is defined as "assigning an *item* to one of a number of distinct predefined groups based on shared properties". This appears to be a usual definition for "classification", but a note to the definition states: "If classes are defined directly by *observations* that do not require *interpretation*, *classification* can be performed independent of the *examiner*." The meaning of this note will be discussed in §5.6 below. *Identification* is defined as "assignment to the most specific class attainable", with examples given of "*Identifying* a red substance as human blood, a white powder as cocaine, or a fibre as cotton." Whether the second class in each example is actually "the most specific class attainable" is debatable, but perhaps the examples are sufficient to elucidate the intent of the definition. A note to the definition of *identification* states "*Identification* is not defined as answering a *question of source*." The latter term is defined as "question of whether two or more *items* have the same origin or have different origins, or who or what the origin of an *item* is".[5] Thus, *identification* is defined as a subclass of *classification*, assigning an *item* to the most specific class attainable, and not as a *categorical opinion* that two *items* (e.g., a fingermark and a fingerprint) came from the same source. *Categorical opinion* is defined as an "*opinion* that, in the *examiner's* judgement, one *proposition* or *explanation* is true and it would be impossible to obtain the *observations* if any of the alternative *propositions* or *explanations* considered were true".

Examiner and *forensic service provider* form a pair of contrasting terms. *Examiner* is defined as an "individual who conducts any part of the *forensic examination*", but *examiner* is also used in Part 5 Reporting, so a more accurate definition would be "individual who conducts any part of the *forensic process*". *Examiners* include scene-of-crime officers, laboratory technicians, reporting officers, expert witnesses, etc. Neither "forensic scientist" nor "forensic practitioner" are used. In contrast, *forensic service provider* is defined as an "organization that performs all or part of the *forensic process*". A note to the definition of *forensic service provider* states that a *forensic service provider* can be an organization that consists of a single *examiner*. The distinction between *examiner* (an individual) and *forensic service provider* (an organization) is important because some requirements and recommendations apply to *examiners* and others to *forensic service providers*.

Method and *procedure* form a pair of contrasting terms. *Method* is defined as "systematic and structured approach to perform a task" and a note gives *analysis* and *interpretation* as examples of tasks. *Procedure* is defined as "pre-defined sequence of steps or actions that are followed to accomplish a specific goal" and examples given are procedures for *method validation*, for evaluating measurement uncertainty, and for *technical review*.

Verification and *validation* have the same meanings as in ISO/IEC 17025:2017, i.e., *verification* is "provision of objective evidence that a given process, *method* and/or equipment fulfils specified requirements", and *validation* is "*verification*, where the specified requirements are adequate for an intended use".

Unless preceded by "either", the meaning of "A or B" is inclusive, i.e., it is equivalent to "A and/or B".

As in all ISO standards, clauses containing "shall" specify requirements, clauses containing "should" specify recommendations, and clauses containing "may" specify permissions. These auxiliary verbs are not terms that are defined in Part 1, but a statement of their meaning appears in the Introduction of each of the other Parts. Part 1 does not contain requirements, recommendations, or permissions. Notes and annexes in the other Parts are informational and do not contain requirements, recommendations, or permissions.

3. Part 2: Recognition, recording, collection, transport and storage of items

Part 2 is primarily concerned with activities that occur at *scenes*. In addition to providing requirements and recommendations related to the *recovery* of *items*, it also provides requirements and recommendations for transport and storage of *items*. The latter continue to be relevant during the *analysis* stage of the *examination*. Part 2 also includes some requirements and recommendations related to the *forensic process* more broadly.

Part 2 is in the process of being revised, and its title and content could change. We therefore do not comment on it in any detail.

The scope for the 2018 version of Part 2 indicates that it applies to the *recovery* of physical *items* but not to the *recovery* of digital *items*. *Recovery* of digital *items* is covered in ISO/IEC 27037 *Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence*. We have no reason to believe that the revised version of Part 2 will differ with respect to this aspect of its scope.

4. Part 3: Analysis

4.1. Contents

The main sections and subsections of Part 3 *Analysis* are:

- §4 General requirements
  - §4.1 General
  - §4.2 *Validation* and *verification* of *methods*
  - §4.3 Equipment and consumables
  - §4.4 Calibration and performance checks
  - §4.5 Metrological traceability
  - §4.6 Reference *samples*, control *samples*, collections and databases
  - §4.7 Measurement uncertainty and significant figures
  - §4.8 *Performance monitoring*
- §5 Personnel
- §6 *Facilities*
- §7 Environmental conditions
- §8 Acceptance and rejection of requests and *items*
  - §8.1 Acceptance and rejection of the customer's request
  - §8.2 Recording of accepted or rejected *items*
- §9 *Analytical strategy*
  - §9.1 General requirements
  - §9.2 Assessment prior to *analysis*
  - §9.3 Selection of *methods*
  - §9.4 *Item* selection, sampling and preparation
- §10 Purpose of the *analysis*
  - §10.1 *Classification* or *identification*
  - §10.2 *Quantification*
  - §10.3 *Question of source*
  - §10.4 *Reconstruction*
- §11 Reliability of *observations*

[5] *Question of source* is grammatically inconsistent with its homologous terms, *classification/identification*, *quantification*, and *reconstruction*. "Source attribution" or "source comparison" would have been consistent.

4.2. Comments

The scopes of Part 3 *Analysis* and Part 4 *Interpretation* both state that they are applicable to all branches of *forensic* science. Once *observations* have been obtained from *items*, the same sorts of *interpretation methods* can be applied irrespective of the nature of the *items* and the branch of *forensic* science. In order to obtain *observations*, however, appropriate *analysis methods* will, to a much greater extent, vary depending on the nature of the *items* and the branch of *forensic* science. Requirements and recommendations essential for *analysis* of an *item* in one branch of *forensic* science could be irrelevant for *analysis* of a different type of *item* in another branch of *forensic* science. Statements of requirements and recommendations in Part 3 are, therefore, often hedged with phrases such as "if applicable". Many requirements and recommendations so hedged appear to be specific to (and therefore applicable to) chemical and biological *analysis methods* used to extract *observations* from physical *items* such as drugs, explosives, and biological *samples*. In contrast, few, if any, of the hedged requirements and recommendations appear to be applicable to signal-processing *methods* used to extract *observations* from digital *items* such as voice recordings and facial images. A challenge in applying Part 3 to a particular *analysis* task in a particular branch of *forensic* science will be to determine which requirements and recommendations are actually applicable and which are not, and to document and justify those determinations.

Requirements in Part 3 that are broadly applicable include requirements in §4.1:

- to document *analysis methods* and *procedures*,
- to record the "data, information and *methods* used during *analysis*", and
- to record any deviations from a *procedure*.

§4.1 also requires the *forensic service provider* to, "when applicable", have documented procedures for "mitigat[ing] the risk of loss, degradation, *contamination* or alteration of *items*". In additional to having, if applicable, procedures to reduce the impact of these adverse events if they occur (mitigation), we recommend that, if applicable, the *forensic service provider* have procedures to reduce the probability that these adverse events will occur.

§4.2 requires *methods* to be *validated* prior to implementation, and §9.3 states that "Any deviation from a *validated* or *verified* analytical *method* shall be technically justified and recorded. The expected impact of the deviation on the validity of the *method* shall be recorded." §4.2 also directs readers to ISO/IEC 17025:2017 §7.2 for additional guidance on *validation* and *verification*.

§4.7 requires "Estimation of measurement uncertainty [to] be included in the *validation* of quantitative *methods*." If an *analysis method* uses signal processing to automatically extract feature vectors (*observations*), such as coefficients of Fourier transforms, and those feature vectors are directly fed into a statistical model that implements an *interpretation method*, what would have to be validated would be the combination of the *analysis* and *interpretation methods* as a unified whole. Such feature vectors are usually high-dimensional and large numbers of feature vectors are often extracted. Such feature vectors are generally not human interpretable and are not examined by a human. It could be argued that measurement uncertainty per se is not a relevant concept for such feature vectors, but that the *analysis* and *interpretation methods* as a unified whole should take account of sources of variability (such as sampling variability) that could have a substantial impact on the *opinion* [8].

§4.4 states that "Equipment and software that can substantially impact the *observations* shall be checked against pre-established acceptance criteria … prior to use … [and] periodically". If "necessary to guarantee the reliability of *observations*", §4.5 requires measuring equipment to be calibrated.

§4.8 provides requirements and recommendations related to *performance monitoring*. These include that "The *forensic service provider* shall participate regularly in a relevant *proficiency testing* program or *interlaboratory comparison*, if available. … If an appropriate proficiency test or *interlaboratory comparison* does not exist, the *forensic service provider* should implement an *intralaboratory comparison*." §4.8 also states that "Performance monitoring can be declared or blind."

In chemical and biological *analysis* in particular, it is important to have requirements and recommendations related to the selection, sampling, and preparation of *items*, the selection of analytical *methods*, and the order in which *methods* are applied (e.g., applying non-destructive *methods* before applying destructive *methods*). Such requirements and recommendations are provided in §9.[6]

5. Part 4: Interpretation

5.1. Contents

The main sections and subsections of Part 4 *Interpretation* are:

- §4 General requirements
- §5 Questions and *propositions*
  - §5.1 Questions
    - §5.1.1 General
    - §5.1.2 Types of questions
    - §5.1.3 *Classification*, including *identification*
    - §5.1.4 *Quantification*
    - §5.1.5 *Question of source*
    - §5.1.6 *Reconstruction*
  - §5.2 Addressing the question with either *investigative* [*interpretation*] or *evaluative interpretation*
    - §5.2.1 General
    - §5.2.2 *Investigative interpretation*
    - §5.2.3 *Evaluative interpretation*
- §6 Types of *opinions*
  - §6.1 General
  - §6.2 *Likelihood ratios*
  - §6.3 *Categorical opinions*
  - §6.4 *Explanations*
  - §6.5 Estimations
- §7 Assigning *probabilities*
  - §7.1 General
  - §7.2 Statistical-model-based assignment of *probabilities*
  - §7.3 Human-based assignment of *probabilities*
- §8 Expression of the *opinion*
  - §8.1 General
  - §8.2 Quantitative expression of *opinion*
  - §8.3 Qualitative expression of *opinion*
  - §8.4 *Categorical opinions*

The remainder of this section is primarily organized in terms of types of *opinions*.

5.2. Likelihood ratios

In the context of *interpretation* related to the source of an *item*, although Part 4 does not require *examiners* to calculate *likelihood ratios* using relevant data, quantitative measurements, and statistical models, there is a pathway through Part 4 that includes applicable requirements and recommendations if this is what an *examiner* is doing. We hope that the existence of this pathway will encourage greater adoption of *methods* that calculate *likelihood ratios* using relevant data, quantitative

[6] §9 is titled *analytical strategy*. The correct term would have been "analysis strategy" – it is a strategy related to how to conduct *analyses*, not a strategy that is analytical.

measurements, and statistical models.

Requirements and recommendations related to implementation of the *likelihood-ratio* framework are distributed across multiple sections and subsections of Part 4. Below, we gather together key requirements and recommendations related to implementation of the *likelihood-ratio* framework, and list and comment on them in what we hope the reader finds to be an easily understandable order.

§5.2.3 requires *propositions* to be specified before the *analysis* is conducted. It states: "When *propositions* are used to address case-relevant questions, they shall be specified before the *analysis* is conducted." We find it peculiar that the first clause in this sentence states a condition, because if a *proposition* does not address a case-relevant question, then it is not a meaningful *proposition*. The title of §5.2.3 is "*Evaluative interpretation*", so the intent could have been: "If conducting *evaluative interpretation*, *propositions* shall be specified before the *analysis* is conducted."

§5.2.3 also lists the following requirements for *propositions*:

> "The *examiner* shall consider at least two competing *propositions* which shall be:
> a) relevant to answering the question posed in the case;
> b) mutually exclusive;
> c) explicit and specific;
> d) considered during *analysis*, *interpretation*, and *reporting*."

5.1.5.1a and §5.1.5.1b state that either specific-source *propositions* or common-source *propositions* can be used. For the distinction between specific-source and common-source *likelihood ratios*, see [9–11]. §5.1.5 does not explicitly mention *likelihood ratios* or *propositions*, but it clearly assumes the use of the *likelihood-ratio* framework.

§5.1.5 requires between-source variability, within-source variability, and measurement variability to be taken into account, and it requires "rarity"[7] with respect to the *relevant population* to be taken into account. The exact wording requires the rarity of "observed properties" to be taken into account or the rarity of "observed comparison scores" to be taken into account. Thus, either feature-based methods or score-based methods may be used. "Comparison score" is not defined, but we assume its references include "similarity score". Similarity scores measure the degree of similarity (or the degree of difference) between the *observations* made on the questioned-source and known-source *items* in a case, but they do not take account of the typicality of those *observations* compared to *observations* made on *items* sampled from the *relevant population*. Although Part 4 permits similarity-score-based *methods* to be used, such *methods* do not properly take account of typicality [12–14]. We therefore recommend that they not be used.

§7.1 requires data used for assigning *probabilities* to "be as representative as possible of the *relevant population* of *items* and as representative as possible of the conditions of the case". As written, conformity with this requirement is not assessable: The requirement does not state who shall do what, and it is impossible to know whether data are "as representative as possible of the *relevant population* … and … the conditions of the case". §7.2 states: "Data shall only be used for assigning *probabilities* if, in the *examiner's* judgement, they are sufficiently representative of the *relevant population* and the conditions of the case." Despite some grammatical ambiguity,[8] the latter is conceptually more appropriate, it requires the *examiner* to do something, and is in line with recommendations in [15]. What is missing, however, is a requirement for the *examiner* to record the result of their assessment of whether the data are sufficiently representative of the *relevant population* for the case and sufficiently reflective of the conditions of the case. Also missing (in Part 5) is a requirement for the *examiner* to report the result of that assessment. We recommend that *examiners* both record and report the results of their assessment of whether "data … used for assigning *probabilities* … are sufficiently representative of the *relevant population* and the conditions of the case."

§6.2 permits *likelihood-ratio* values to be assigned either quantitatively or qualitatively. A note states that "A *likelihood ratio* can be assigned quantitatively using statistical models, or quantitatively or qualitatively based on professional judgement." §7.1 states that "*Probabilities* are assigned using *observations*, existing data or statistical models"[9]; however, "in the absence of existing data or statistical models" it explicitly gives permission to assign "subjective *probabilities*". In this context we interpret "subjective *probabilities*" as meaning probabilities assigned using "professional judgement", rather than meaning a subjectivist Bayesian concept of probability as opposed to a frequentist concept. Uncalibrated unvalidated subjective assignment of *likelihood ratios* has been criticized in [16–19], and we recommend that this not be done. Rather than the permission to assign subjective *probabilities* "in the absence of existing data or statistical models" being exploited as an excuse not to develop and adopt *methods* based on relevant data, quantitative measurements, and statistical models, we hope that ISO 21043 will encourage development and adoption of such *methods*.

§7.2 states that "Statistical models … used to assign *probabilities* shall be suitable for their intended use", and §4 states that "The *interpretation methods* shall be suitable for the intended use and should be supported by relevant *validation* studies." We find this mixture of requirement and recommendation to be incompatible. The way to demonstrate that a *method* (including a *method* instantiated in a statistical model) is suitable for its intended use is via *method validation* – the definition of *validation* is "provision of objective evidence that a … *method* … fulfils specified requirements, where the specified requirements are adequate for an intended use". This mixture of requirement and recommendation would have been better written as:

- a requirement for someone or some organization (e.g., the *forensic service provider*) to conduct *validation* studies under casework conditions;
- a requirement for the *examiner* to use their professional judgement to assess whether the conditions of the case are sufficiently similar to those of a *validation* study that it is appropriate to use that *validation* study to demonstrate that the *method* is suitable for use in that case;
- a requirement for the *examiner* to record the outcome of that assessment; and
- a requirement (in Part 5) for the *examiner* to report the outcome of that assessment.

We recommend that *forensic service providers* and *examiners* follow these steps.

ISO 21043 does not provide requirements or recommendations for to how to validate *interpretation methods* that calculate *likelihood ratios*. For guidance on how to validate *interpretation methods* that calculate source-level *likelihood ratios*, we recommend the "Consensus on validation of forensic voice comparison" [15] – although its title and scope refer to forensic voice comparison, with minor wording changes it is applicable in other branches of *forensic* science.

§8.2 includes a recommendation that "If a numerical value (e.g., a

[7] Our preferred term would be "typicality". A higher value in the denominator of the *likelihood ratio* corresponds to greater typicality but to less rarity. We think it clearer to pair "higher" with "greater" rather than with "less".

[8] The phrasing of the sentence using "only" is ambiguous. We recommend not interpreting it as a permission not to use any data at all, but to interpret it as a requirement to use suitable data. The sentence could have been better phrased as: "Data used for assigning *probabilities* shall be data that, in the *examiner's* judgement, are sufficiently representative of the *relevant population* and sufficiently representative of the conditions of the case.".

[9] We have difficulty understanding this sentence and wonder whether, especially given the context of the following sentence, the intent was: "Probabilities are assigned using *observations*, existing data, and statistical models." The wording does not specify a requirement or a recommendation.

probability or a *likelihood ratio*) is calculated using quantitative *observations*, quantitative data, and a statistical model, the *opinion* should be expressed using that numerical value." Part 5 §5.5.2 and §5.5.3 have parallel recommendations for reporting. We recommend that users of ISO 21043 treat these recommendations as requirements. If a numerical *likelihood-ratio* value has been calculated, not reporting that value would be non-transparent. That transparency is an important principle for *interpretation* is apparent from multiple requirements and recommendations stated in Part 4, including the following from §7.1: "The *examiner* shall document how *probabilities* are assigned, including the basis for such assignments. The *examiner* shall be transparent about the sources of information used to assign *probabilities*. Any assumptions used to assign *probabilities* should be clearly recorded."

§8.3 requires qualitative *opinions* (including subjectively assigned *likelihood ratios*) to be "drawn from a predefined *opinion scale* on which each level has a verbal expression". Examples of *opinion scales* containing verbal expressions are provided in Part 4 Annex B. If a *likelihood-ratio* value is calculated using relevant data, quantitative measurements, and statistical models, Part 4 does not require or recommend the use of a verbal expression from an *opinion scale*, but it does not prohibit it either. We recommend reporting only the numerical *likelihood-ratio* value, and not using an *opinion scale* – providing a verbal expression in addition to a numerical *likelihood ratio* has not been found to improve understanding of the meaning of the *likelihood ratio* (see [20] for a review of the empirical literature). For criticisms of the use of *opinion scales*, see [21–24].

§6.2 includes a warning not to transpose the conditional (not to commit the prosecutor's fallacy). The warning in §6.2 is actually phrased as a requirement: "The *examiner* shall not confuse the *likelihood ratio* with the ratio of the *probabilities* of the *propositions*." As written, however, it is not possible to assess conformance with this requirement because it refers to the *examiner's* state of mind, not to their actions. Part 5 §5.5.3 includes an appropriately phrased requirement: "When reporting a *likelihood ratio* the *examiner* … shall not transpose the conditional."

Part 4 Annex A provides an explanation of the meaning of a *likelihood ratio*.

5.3. Categorical opinions

Part 4 permits examiners to state *categorial opinions*. §6.3 states "*Categorical opinions* shall only be formed when, in the *examiner's* judgement, it would be impossible to obtain the *observations* if any alternative *proposition* were true." This is the Part 1 definition of *categorical opinion* restated as a requirement. "impossible to obtain the *observations* if any alternative *proposition* were true" could seem to be a high bar, but the decision as to whether the bar is surpassed is dependant on the *examiner's* judgement, and thus is susceptible to *cognitive bias*. There has been a great deal of concern about *cognitive bias* in forensic science, e.g., [25–36]. *Categorial opinions* based on both human perception and subjective judgement would be especially susceptible to *cognitive bias*. §8.4 recommends that a second *examiner* "check" the *opinion* and examine the *items*. In our opinion, this does not adequately address concerns regarding susceptibility to *cognitive bias* – the check by the second *examiner* of the subjective *opinion* reached by the first *examiner* is not blind to that *opinion*.

Categorial opinions, are in general, incompatible with the *likelihood-ratio* framework. *Categorial opinions*, are definitely incompatible if the *observations* are quantitative, continuously valued, and have within-source variability. Perhaps the only exceptions to the incompatibility of *categorical opinions* with the *likelihood-ratio* framework would be exclusions when invariant discrete data (e.g., single-donor high-template DNA profiles) do not match exactly, and exclusions in trivial circumstances such as the calibre of a cartridge case being different from the calibre of a firearm. Apart from in such exceptional cases, we recommend that *categorial opinions* not be used.

5.4. Explanations

As explained in §2 above, an *explanation* is a hypothesis that is generated after considering the *observations*. This contrasts with a *proposition*, which is a hypothesis that §5.2.3 requires "be specified before the *analysis* is conducted." §5.2.2 states that "In *investigative interpretation*, the *examiner* will not have *propositions* at the outset. The *examiner* may generate possible *explanations* for the *observations*." And §6.4 requires that "When assigning *probabilities* of *explanations*, the *examiner* shall take into account not only the current *observations* but all available relevant *observations* and information in the case."

Annex C of Part 4 includes an example of the use of *explanations*:

> "After a building burned down, the police consider whether a criminal investigation should be started, and ask an *examiner* about the cause of the fire. After examining the *scene* of the fire, the *examiner* sees three possible *explanations* for the fire: (*Explanation A*) an electrical malfunction, (*Explanation B*) an overheated stove while cooking, and (*Explanation C*) addition of a source of ignition by a person. For the purpose of helping the police decide how to proceed, the *examiner* ranked the *explanations* for the fire by their *probability*. The *examiner* expressed the *investigative opinion* by stating that 'based on my *observations* and not knowing anything else about the case, in my *opinion*, *Explanation A* is the most probable, *Explanation C* is much less probable, and the probability of *Explanation B* is very low.'"

As previously mentioned, there has been a great deal of concern about *cognitive bias* in forensic science, e.g., [25–36]. [27] was particularly critical of the practice of generating post hoc *explanations* for *observations*, and likened it to shooting at the side of a barn then drawing targets around the bullet holes. In our opinion, use of *investigative interpretation* to generate *explanations* and to assign *probabilities* to those *explanations* would be highly susceptible to *cognitive bias*.

Recall that (as mentioned in §2 above) a note to the definition of *investigative interpretation* states that it "can be used in an investigation or in a judicial setting." Even if *investigative interpretation* which generates *explanations* and assigns *probabilities* to those *explanations* is necessary in investigative contexts, we recommend that, because of its high susceptibility to *cognitive bias*, it not be used for evidential purposes.

§5.2.2 states that "*explanations* may be turned into *propositions* for a later *evaluative interpretation*."[10] If this is done when transitioning from a investigative context to an evidential purpose, we recommend that steps be taken to reduce the potential for *cognitive bias*, e.g.:

- The evidential *analysis* and *interpretation* should be conducted from scratch by an *examiner* who has had no prior exposure to the case (hereinafter the "new *examiner*").
- The new *examiner* should not be provided with any task-irrelevant information about the case.
- The new *examiner* should not be provided with any information about what the first *examiner* did.
- The new *examiner* should only be provided with the *propositions* and the *items* of interest, or, if reanalysis of the *items* is not practical and the *observations* are instrument based rather than human-perception based, the new *examiner* should only be provided with the *propositions* and the *observations*.

5.5. Estimates

"Estimation" is not a term that is defined in Part 1. It therefore has its everyday meaning. "Estimation" is the process of making an approximate calculation or evaluation, and an "estimate" is the output of that process. Unfortunately, in Part 4, "estimation" is often used where the

[10] [7] discussed "the generation of explanations [as] an intermediate stage between collecting observations and formulating propositions."

correct word would be "estimate" (including in the header for §6.5 – all the homologous headers refer to outputs, not to processes). Hereinafter, we use the two nouns, "estimation" and "estimate", according to their different meanings.

§5.2.2 states that "Estimation of a value does not involve the use of *propositions*." §5.2.1 includes examples to help explain when it would be appropriate to give an estimate and when it would be appropriate to apply the *likelihood-ratio* framework (which does involve *propositions*). One pair of examples is:

> "If the question were at what speed a car was moving, the *examiner* would use *investigative interpretation* to give an [estimate] of the speed. If the question were whether the speed was above the speed limit or not, the *examiner* would use *evaluative interpretation* [i.e., the *likelihood-ratio* framework]."

In the latter example, the two *propositions* would be some version of "the car was travelling above the speed limit" and "the car was travelling at or below the speed limit".

§6.5 states that "The *examiner* should assess the uncertainty of the [estimate], e.g., in the form of a distribution, or a range of values." §8.2 states that "[Estimates] shall be expressed numerically and should be accompanied by an expression of the uncertainty of the [estimate]."

5.6. Loophole

The scope for Part 4 states that the requirements in Part 4 "are applicable to all *forensic* disciplines"; however, the scope also states that "*Interpretation* is not necessary, and the requirements of [Part 4] do not apply, if the *observations* resulting from the *analysis* directly answer the relevant question", and §5.1.3 states "If classes are defined directly by *observations* that do not require *interpretation*, the requirements of [Part 4] do not apply to that *classification*." Also, as mentioned in §2 above, a note to the definition of *classification* states that "If classes are defined directly by *observations* that do not require *interpretation*, *classification* can be performed independent of the *examiner*." The scope for Part 4 includes an example in a particular branch of *forensic* science:

> "In analytical chemistry, substances are often *identified* or *classified*. Provided that the applied analytical *methods* are not limited in *selectivity* or *sensitivity* for the given question, the *observations* can lead to a direct statement of the name of the substance (*identification*) or a type of material (*classification*). This is not considered *interpretation* for the purposes of [Part 4]."

This appears to provide what we consider a worrying loophole. Although the example is specific to *classification* or *identification* in analytical chemistry, it is not clear how many types of *opinions* in how many branches of *forensic* science could pass through the loophole. The phrase "not limited in *selectivity* or *sensitivity*" implies a zero error rate, so the broader text permits an *examiner* to claim that their *method* has a zero error rate and then present an *opinion* without following the ISO 21043 requirements for *interpretation*. Claims that *methods* have zero error rates were severely criticized by the President's Council of Advisors on Science and Technology (PCAST) [37]. A forensic-chemistry standard, ASTM E2329-14 *Standard Practice for the Identification of Seized Drugs*, also claimed that some *methods* (or combinations of *methods*) have zero error rates. ASTM E2329-14 was the first standard to begin development under the National Institute of Standards and Technology's (NIST's) Organization of Scientific Area Committees for Forensic Science (OSAC) and then to be included in the OSAC Registry of Standards. Immediately after ASTM E2329-14 was posted to OSAC's registry, however, NIST issued a statement raising concerns about the standard's claim of zero error rates [38,39]. In application, all *methods* have a non-zero error rate. The error rate for a *method* can be proven to be small, but it can never be proven to be zero. If one runs a series of tests and does not observe an error, that does not mean that the error rate is zero because the number of tests performed is finite. An error could occur on the next test. There are established statistical techniques for calculating error rates for a series of tests in which no errors are observed (or in which only a small numbers of errors are observed). Using these techniques, as the number of tests performed increases the estimated error rate decreases, but, since the number of tests performed is finite, the estimated error rate never reaches zero. The example given in the scope for Part 4 is, therefore, invalid: No *methods* have zero error rates, hence no *interpretation methods* can be exempt from the ISO 20143 requirements for *interpretation*. We recommend that users of ISO 21043 not attempt to exploit the apparent loophole.

6. Part 5: Reporting

6.1. Contents

The main sections and subsections of Part 5 Reporting are:

- §4 General requirements
- §5 Report contents
  - §5.1 General
  - §5.2 Administrative requirements
  - §5.3 *Item* integrity
  - §5.4 Measurement uncertainty and significant figures
  - §5.5 Reporting *opinions*
    - §5.5.1 General
    - §5.5.2 Reporting *probabilities*
    - §5.5.3 Reporting *likelihood ratios*
    - §5.5.4 Reporting estimations
- §6 Case file review
  - §6.1 General
  - §6.2 *Technical review*
  - §6.3 *Administrative review*
- §7 Issue and control of *reports*
  - §7.1 Issuing *reports*
  - §7.2 Supplementing, amending and withdrawing *reports*
  - §7.3 Controlling *reports*
  - §7.4 Communications of *observations* or *opinions* other than by way of a report
- §8 *Testimony*
  - §8.1 General
  - §8.2 *Testimony* training and evaluation

6.2. General

§4 states general requirements for reporting, including:

- that *examiners* "not report beyond their area of expertise" and "not report beyond what can be based on the available information";
- that known limitations of *methods* and *procedures* be reported; and
- that "any deviations from documented *methods* or *procedures* … [that] could have a substantial impact on the *observations* or *opinions*" be reported, and that "The reason for any such deviation … be explained".

§5.1–§5.4 provide requirements and recommendations related to what content to include in *reports*, and related to the formatting of *reports*.

§7 provides requirements and recommendations related to the issuing and control of *reports*, including supplementing, amending, and withdrawing *reports*, ensuring the traceability of any authorized alterations of *reports*, and reducing the risk of unauthorized alteration of *reports*.

§7.4 provides requirements and recommendations related to communicating *observations* and *opinions* verbally or in writing outside the context of a *report*, e.g., preliminary communication of *observations* and *opinions*, or communication of *observations* and *opinions* from mass *examinations* for investigative use. These requirements are laxer than those for *reports*, but §7.4 also includes a requirement to issue a *report* if the information communicated is intended for use by a court.

§8.1 states "*Observations* and *opinions* should be supported by either *reports*, case records data or documents, or all." We have difficulty understanding this sentence, and wonder whether the intended meaning was: "*Observations* and *opinions* should be supported by one or more of: *reports*, case records, data, or documents."

6.3. Reporting likelihood ratios (and other types of opinions)

§5.5 provides requirements and recommendations related to reporting *opinions*. These include, in §5.5.1, that *reports* "shall differentiate *opinions* from *observations*" (§5.5.1.1), that "*opinions* … shall not be presented as if they were statements of fact" (§5.5.1.4),[11] and (in §5.5.1.1):

"Reports that contain *opinions* shall include:

a) the questions to be answered;
b) the set of *propositions*, where applicable;
c) a description of assumptions that are relevant for forming the *opinion*;
d) the basis of the *opinion*."

and (in §5.5.1.2):

"Reports that contain *opinions* should include or reference:

a) a description of any data used;
b) a description of any statistical models used."

§5.5.1.4 also recommends that when reporting source-level *opinions*, the *examiner* should state that the *opinion* only relates to source level and not to activity level.

§5.5.2 and §5.5.3 permit *probabilities* and *likelihood ratios* to be expressed quantitatively or qualitatively. §5.5.1.2 recommends that, if an *opinion scale* is used (which Part 4 §8.3 requires for qualitative *opinions*), the whole *opinion scale* be included or referenced in the *report*.

§5.5.2 recommends that data used to assign *probabilities* be described or referenced in the *report* (this is a repetition of §5.5.1.2a). §5.5.2 also recommends that, if a *method* for assigning *probabilities* outputs a numerical value, the numerical value be reported, and §5.5.3 recommends that, if a *method* calculates a numerical *likelihood-ratio* value, the numerical *likelihood-ratio* value be reported. As stated in §5.2 above, we recommend that users of ISO 21043 treat these recommendations as requirements. Not reporting the numerical value would be non-transparent.

§5.5.4 requires estimates to be reported as numerical values and each estimate to be accompanied by an expression of the uncertainty of its value.

§5.5.3 requires that, when expressing a *likelihood ratio*, the *examiner* not transpose the conditional (not commit the prosecutor's fallacy).

§5.5.1.3 states that "*Opinions* shall not be presented in such a way as to overstate or understate their strength."[12] But goes on to state:

> "Where the strength of the *opinion* is reported as a *likelihood ratio*, a bound of the *likelihood ratio* closer to the neutral value of 1 may be reported, making clear that the *likelihood ratio* exceeds that bound. The *forensic service provider* shall have a documented rationale for choosing that bound."

An example is provided of reporting "In a DNA profile comparison … that 'the *likelihood ratio* is greater than 1 billion'", and it is stated that "This would not be understating the strength of the *opinion*." A note states that:

> "Lowering a *likelihood ratio* based on limitations of the quality[13] of the *item*, measurement or data, is not understating the strength of the *opinion*, but is following the requirements [in Part 4] to take limitations into account in the *interpretation*".

These contortions (stating a requirement then giving permission to violate the requirement but pretending that it does not violate the requirement) could have been avoided if the initial requirement had simply been that the *examiner* not <u>deliberately</u> overstate the strength of the *opinion*. Since an *examiner* cannot know whether they are understating or overstating the strength of the *opinion*, it is impossible to comply with the requirement not to understate or overstate its strength.

"Quality" of *items*, should be taken account of by calibrating the system that calculates the *likelihood ratios*. The system should be calibrated using data that reflect the conditions of the questioned-source and known-source *items* in the case. For a system calibrated and *validated* using data reflecting poorer conditions, the log-likelihood-ratio values that it outputs will be closer to the neutral value of 0 than for a system calibrated and *validated* using data reflecting better conditions. See [40] Figure 8 for examples of Tippett plots showing such relationships in *validation* results from a system calibrated using different amounts data from questioned-source *items* (different durations of speech in questioned-speaker voice recordings).

In some jurisdictions, it is standard practice in DNA-profile comparison to place an upper bound on the magnitude of the *likelihood ratio* that is reported [41]. The rationale for this is that the calculated value is based on *sample* data, which are subject to sampling variability, and statistical models, which approximate but do not necessarily match the structure of the population distribution. The value output by the model is therefore subject to uncertainty.

One cannot know whether a particular calculated *likelihood-ratio* value understates or overstates what the value would be if one had access to the "true" distribution of the population. A common concern, however, has been about the danger of overstating the strength of the *opinion*. This has led to the use of *methods* that reduce the probability of overstating the strength of the *opinion*, i.e., *methods* that reduce the probability of presenting a *likelihood-ratio* value that is further from the neutral value of 1 than what the "true" *likelihood-ratio* value might be. If one takes steps to reduce the probability of overstating the strength of the *opinion*, however, then one will necessarily increase the probability of understating the strength of the *opinion*.

Even if a *likelihood-ratio* value calculated from single-donor high-template DNA profiles is many orders of magnitude larger than 1 billion, in order to reduce the probability of overstating the strength of the *opinion*, *examiners* in some jurisdictions report that the *likelihood-ratio* value is on the order of 1 billion [41]. For the legal-decision maker in the context of a case, a value of 1 billion is still a very large number.

If a *likelihood-ratio* value is calculated, for the sake of transparency, we recommend that the calculated value be included in the *report*, even if the headline value is reported as "greater than 1 billion". As mentioned in §5.2 above, Part 4 §8.2 includes a recommendation that "If a numerical value (e.g., a *probability* or a *likelihood ratio*) is calculated using quantitative *observations*, quantitative data, and a statistical model, the *opinion* should be expressed using that numerical value."

Practice for DNA in some jurisdictions is to report a *likelihood-ratio* value of 1 billion without actually calculating a *likelihood-ratio* value, in the knowledge that if a *likelihood-ratio* value were calculated it would be greater than 1 billion: if single-donor high-template DNA profiles match at all the loci of a multiplex of 15 or more loci, and the different-source *proposition* is that the questioned-source donor is a person who is from a large *relevant population* and who is unrelated to the known-source

[11] These requirements are repeated in §8.1 in the context of *testimony*.

[12] This requirement is repeated in §8.1 in the context of *testimony*.

[13] The use of the word "quality" here does not accord with the Part 1 definition of "degree to which a set of inherent characteristics fulfils a requirement".

donor, then, irrespective of what the alleles in the profiles actually are, the calculated *likelihood-ratio* would be orders of magnitude greater than 1 billion [41].

The logic behind placing a bound on the values of *likelihood ratios* could also apply to *likelihood ratios* calculated in other branches of *forensic* science in which the values of calibrated *likelihood ratios* will be (much) closer to the neutral value of 1 than 1 billion, and any bound used would therefore also be (much) closer to the neutral value of 1.

Despite being permitted in §5.5.1, calculating and reporting such bounds is controversial, and has been debated at length, including in [17,42–55]. A key factor in the debate has been whether one subscribes to a subjectivist Bayesian concept of probability or to a frequentist or likelihoodist concept of probability. [8] noted that, irrespective of one's philosophical position, the practical outcome would be the same: one would report a Bayes factor or a likelihood ratio that would be closer to the neutral value of 1 than if one simply calculated a likelihood ratio based only on the sample data. [8] proposed the use of methods (either Bayesian or likelihoodist) that shrink the calculated Bayes-factor or likelihood-ratio value toward the neutral value of 1, e.g., by using uninformative priors or by regularizing the model. Such methods are not explicitly permitted in §5.5.1, but we think they fall within the spirit of the permission to report a bound closer to the neutral value of 1 and should therefore be acceptable.

7. Conclusion

We have provided guidance which we hope will encourage and assist *forensic service providers* and *examiners* to get maximum benefit out of ISO 21043 *Forensic sciences*. We encourage *forensic service providers* and *examiners* to use ISO 21043 and, based on their experience of using it, to provide feedback that will assist in making improvements when the standard is next revised.

Disclaimer

Unless explicitly otherwise attributed, all opinions expressed in the present paper are personal opinions of the authors which should not be construed as representing policies or positions of any organizations with which the authors are associated.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Forensic Science Regulator for England & Wales (2025). *Code of Practice (version 2)*. https://www.gov.uk/government/publications/forensic-science-regulator-code-of-practice.
[2] International Laboratory Accreditation Cooperation (2022). *ILAC G19: Modules in a Forensic Science Process*. https://ilac.org/?ddownload=124605.
[3] G.S. Morrison, Advancing a paradigm shift in evaluation of forensic evidence: The rise of forensic data science, Forensic Sci. Int.: Synergy 5 (2022) 100270, https://doi.org/10.1016/j.fsisyn.2022.100270.
[4] L. Wilson-Wilde, The international development of forensic science standards — a review, Forensic Sci. Int. 288 (2018) 1–9, https://doi.org/10.1016/j.forsciint.2018.04.009.
[5] D. Meuwly, Implications of the forthcoming forensic sciences standard ISO/IEC 21043 for forensic biometrics, in: *Proceedings of the International Workshop on Biometrics and Forensics (IWBF)*, 2024, https://doi.org/10.1109/IWBF62628.2024.10701603.
[6] C.E.H. Berger, Finally a really forensic worldwide standard: ISO 21043 Forensic sciences, Part 4, Interpretation. Forensic Science International: Synergy 10 (2025) 100589, https://doi.org/10.1016/j.fsisyn.2025.100589.
[7] I.W. Evett, G. Jackson, J.A. Lambert, More on the hierarchy of propositions: Exploring the distinction between explanations and propositions, Sci. Justice 40 (2000) 3–10, https://doi.org/10.1016/S1355-0306(00)71926-5.
[8] G.S. Morrison, N. Poh, Avoiding overstating the strength of forensic evidence: Shrunk likelihood ratios / Bayes factors, Sci. Justice 58 (2018) 200–218, https://doi.org/10.1016/j.scijus.2017.12.005.
[9] D.M. Ommen, C.P. Saunders, Building a unified statistical framework for the forensic identification of source problems, Law Probab. Risk 17 (2018) 179–197, https://doi.org/10.1093/lpr/mgy008.
[10] D.M. Ommen, C.P. Saunders, A problem in forensic science highlighting the differences between the Bayes factor and likelihood ratio, Stat. Sci. 36 (2021) 344–359, https://doi.org/10.1214/20-STS805.
[11] G.S. Morrison, Taking account of typicality in calculation of likelihood ratios, Law Probab. Risk 24 (2025) mgaf009, https://doi.org/10.1093/lpr/mgaf009.
[12] G.S. Morrison, E. Enzinger, Score based procedures for the calculation of forensic likelihood ratios – scores should take account of both similarity and typicality, Sci. Justice 58 (2018) 47–58, https://doi.org/10.1016/j.scijus.2017.06.005.
[13] C. Neumann, M. Ausdemore, Defence against the modern arts: The curse of statistics – Part II: 'Score-based likelihood ratios', Law Probab. Risk 19 (2020) 21–42, https://doi.org/10.1093/lpr/mgaa006.
[14] C. Neumann, J. Hendricks, M. Ausdemore, Statistical support for conclusions in fingerprint examinations, in: D. Banks, K. Kafadar, D.H. Kaye, M. Tackett (Eds.), *Handbook of Forensic Statistics*, CRC, Boca Raton, FL, 2020, pp. 277–324, https://doi.org/10.1201/9780367527709.
[15] G.S. Morrison, E. Enzinger, V. Hughes, M. Jessen, D. Meuwly, C. Neumann, S. Planting, W.C. Thompson, D. van der Vloed, R.J.F. Ypma, C. Zhang, A. Anonymous, B. Anonymous, Consensus on validation of forensic voice comparison, Sci. Justice 61 (2021) 229–309, https://doi.org/10.1016/j.scijus.2021.02.002.
[16] D.M. Risinger, Reservations about likelihood ratios (and some other aspects of forensic 'Bayesianism'), Law Probab. Risk 12 (2013) 63–73, https://doi.org/10.1093/lpr/mgs011.
[17] K.A. Martire, G. Edmond, D.J. Navarro, B.R. Newell, On the likelihood of "encapsulating all uncertainty", Sci. Justice 57 (1) (2017) 76–79, https://doi.org/10.1016/j.scijus.2016.10.004.
[18] K.A. Martire, B. Growns, D.J. Navarro, What do the experts know? Calibration, precision, and the wisdom of crowds among forensic handwriting experts, Psychon. Bull. Rev. 25 (2018) 2346–2355, https://doi.org/10.3758/s13423-018-1448-3.
[19] G.S. Morrison, K. Ballantyne, P.H. Geoghegan, A response to Marquis et al. (2017) What is the error margin of your signature analysis? Forensic Sci. Int. 287 (2018) e11–e12, https://doi.org/10.1016/j.forsciint.2018.03.009.
[20] G.S. Morrison, A.S. Bali, K.A. Martire, R.H. Grady, W.C. Thompson, What is the best way to present likelihood ratios? A review of past research and recommendations for future research, Manuscript submitted for publication, 2025. Preprint at https://forensic-data-science.net/communication/.
[21] C. Mullen, D. Spence, L. Moxey, A. Jamieson, Perception problems of the verbal scale, Sci. Justice 54 (2014) 154–158, https://doi.org/10.1016/j.scijus.2013.10.004.
[22] K.A. Martire, I. Watkins, Perception problems of the verbal scale: A reanalysis and application of a membership function approach, Sci. Justice 55 (2015) 264–273, https://doi.org/10.1016/j.scijus.2015.01.002.
[23] R. Marquis, A. Biedermann, L. Cadola, C. Champod, L. Gueissaz, G. Massonnet, W. D. Mazzella, F. Taroni, T. Hicks, Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings, Sci. Justice 56 (2016) 364–370, https://doi.org/10.1016/j.scijus.2016.05.007.
[24] G.S. Morrison, E. Enzinger, What should a forensic practitioner's likelihood ratio be? Sci. Justice 56 (2016) 374–379, https://doi.org/10.1016/j.scijus.2016.05.007.
[25] D.M. Risinger, M.J. Saks, W.C. Thompson, R. Rosenthal, The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion, Calif. Law Rev. 90 (2002) 1–56. https://doi.org/10.15779/Z38GM7K.
[26] National Research Council of the National Academies, *Strengthening Forensic Science in the United States: A Path Forward*, National Academies Press, Washington, DC, 2009, https://doi.org/10.17226/12589.
[27] W.C. Thompson, Painting the target around the matching profile: The Texas sharpshooter fallacy in forensic DNA interpretation, Law Probab. Risk 8 (2009) 257–276, https://doi.org/10.1093/lpr/mgp013.
[28] Expert Working Group on Human Factors in Latent Print Analysis (2012). *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.7842.
[29] S.M. Kassin, I.E. Dror, J. Kukucka, The forensic confirmation bias: Problems, perspectives, and proposed solutions, J. Appl. Res. Mem. Cogn. 2 (2013) 42–52, https://doi.org/10.1016/j.jarmac.2013.01.001.
[30] B. Found, Deciphering the human condition: The rise of cognitive forensics, Aust. J. Forensic Sci. 47 (2015) 386–401, https://doi.org/10.1080/00450618.2014.965204.
[31] R.D. Stoel, C.E.H. Berger, W. Kerkhoff, E.J.A.T. Mattijssen, E.I. Dror, Minimizing contextual bias in forensic casework, in: K.J. Strom, M.J. Hickman (Eds.), *Forensic Science and the Administration of Justice: Critical Issues and Directions*, Sage, Thousand Oaks CA, 2015, pp. 67–86, https://doi.org/10.4135/9781483368740.n5.
[32] I.E. Dror, Human expert performance in forensic decision making: Seven different sources of bias, Aust. J. Forensic Sci. 49 (2017) 541–547, https://doi.org/10.1080/00450618.2017.1281348.
[33] G. Edmond, A. Towler, B. Growns, G. Ribeiro, B. Found, D. White, K. Ballantyne, R. A. Searston, M.B. Thompson, J.M. Tangen, R.I. Kemp, K.A. Martire, Thinking forensics: Cognitive science for forensic practitioners, Sci. Justice 57 (2017) 144–154, https://doi.org/10.1016/j.scijus.2016.11.005.

[34] G.S. Cooper, V. Meterko, Cognitive bias research in forensic science: A systematic review, Forensic Sci. Int. 297 (2019) 35–46, https://doi.org/10.1016/j.forsciint.2019.01.016.
[35] Expert Working Group on Human Factors in Handwriting Examination (2020). *Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach*. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8282.
[36] B.A. Spellman, H. Eldridge, P. Bieber, Challenges to reasoning in forensic science decisions, Forensic Sci. Int.: Synergy (2022), https://doi.org/10.1016/j.fsisyn.2021.100200.
[37] President's Council of Advisors on Science and Technology (2016). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. https://obamawhitehouse.archives.gov/administration/eop/ostp/pcast/docsreports/.
[38] National Institute of Standards and Technology (2016). *Statement on ASTM Standard E2329-14*. https://www.nist.gov/news-events/news/2016/03/nist-statement-astm-standard-e2329-14.
[39] G.P. Jackson, Error terror in forensic science: When spectroscopy meets the courts, Spectroscopy 31 (11) (2016) 12–16. https://www.spectroscopyonline.com/view/error-terror-forensic-science-when-spectroscopy-meets-courts.
[40] P. Weber, E. Enzinger, B. Labrador, A. Lozano-Díez, D. Ramos, J. González-Rodríguez, G.S. Morrison, Validation of the alpha version of the E^3 Forensic Speech Science System (E^3FS^3) core software tools, Forensic Sci. Int.: Synergy 4 (2022) 100223, https://doi.org/10.1016/j.fsisyn.2022.100223.
[41] A.J. Hopwood, R. Puch-Solis, V.C. Tucker, J.M. Curran, J. Skerrett, S. Pope, G. Tully, Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts, Sci. Justice 52 (2012) 185–190, https://doi.org/10.1016/j.scijus.2012.05.005. [Corrigendum: Sci. Justice 53 (2013) 371. https://doi.org/10.1016/j.scijus.2013.03.003].
[42] F. Taroni, S. Bozza, A. Biedermann, C. Aitken, Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio, Law Probab. Risk 15 (2016) 1–16, https://doi.org/10.1093/lpr/mgv008.
[43] A. Nordgaard, Comment on 'Dismissal of the illusion of uncertainty on the assessment of a likelihood ratio' by Taroni F., Bozza S., Biederman A. and Aitken C. Law Probab. Risk 15 (2016) 17–22, https://doi.org/10.1093/lpr/mgv006.
[44] M.J. Sjerps, I. Alberink, A. Bolk, R.D. Stoel, P. Vergeer, J.H. van Zanten, Uncertainty and LR: To integrate or not to integrate, that's the question, Law Probab. Risk 15 (2016) 23–29, https://doi.org/10.1093/lpr/mgv005.
[45] G.S. Morrison, Special issue on measuring and reporting the precision of forensic likelihood ratios: Introduction to the debate, Sci. Justice 56 (2016) 371–373, https://doi.org/10.1016/j.scijus.2016.05.002.
[46] J.M. Curran, Admitting to uncertainty in the LR, Sci. Justice 56 (2016) 380–382, https://doi.org/10.1016/j.scijus.2016.05.005.
[47] D.M. Ommen, C.P. Saunders, C. Neumann, An argument against presenting interval quantifications as a surrogate for the value of evidence, Sci. Justice 56 (2016) 383–387, https://doi.org/10.1016/j.scijus.2016.07.001.
[48] C.E.H. Berger, K. Slooten, The LR does not exist, Sci. Justice 56 (2016) 388–391, https://doi.org/10.1016/j.scijus.2016.06.005.
[49] A. Biedermann, S. Bozza, F. Taroni, C. Aitken, Reframing the debate: A question of probability, not of likelihood ratio, Sci. Justice 56 (2016) 392–396, https://doi.org/10.1016/j.scijus.2016.05.008.
[50] A. van den Hout, I. Alberink, Posterior distributions for likelihood ratios in forensic science, Sci. Justice 56 (2016) 397–401, https://doi.org/10.1016/j.scijus.2016.06.011.
[51] D. Taylor, T. Hick, C. Champod, Using sensitivity analyses in Bayesian Networks to highlight the impact of data paucity and direct future analyses: A contribution to the debate on measuring and reporting the precision of likelihood ratios, Sci. Justice 56 (2016) 402–410, https://doi.org/10.1016/j.scijus.2016.06.010.
[52] A.P. Dawid, Forensic likelihood ratio: Statistical problems and pitfalls, Sci. Justice 57 (2017) 73–75, https://doi.org/10.1016/j.scijus.2016.09.002.
[53] A. Biedermann, S. Bozza, F. Taroni, C. Aitken, The consequences of understanding expert probability reporting as a decision, Sci. Justice 57 (2017) 80–85, https://doi.org/10.1016/j.scijus.2016.10.005.
[54] K. Slooten, C.E.H. Berger, Response paper to "the likelihood of encapsulating all uncertainty": The relevance of additional information for the LR, Sci. Justice 57 (2017) 468–471, https://doi.org/10.1016/j.scijus.2017.05.007.
[55] G.S. Morrison, What should a forensic practitioner's likelihood ratio be? II, Sci. Justice 57 (2017) 472–476, https://doi.org/10.1016/j.scijus.2017.08.004.