IMDRF International Medical Device Regulators Forum

Draft

IMDRF/CEIVD WG/N91 DRAFT: 2026

Clinical Evidence for IVD Medical Devices – Definitions and Principles of Performance Evaluation

AUTHORING GROUP

Clinical Evidence for In Vitro Diagnostic Medical Devices

10 February 2026

Preface

Dr. Raymond Chua, IMDRF Chair

Contents

1. Introduction

This document supersedes earlier documents produced under the Global Harmonization Task Force (GHTF): GHTF/SG5/N6:2012 Clinical Evidence for IVD medical devices – Key Definitions and Concepts and GHTF/SG5/N7:2012 Clinical Evidence for medical devices – Scientific Validity Determination and Performance Evaluation.

It outlines the concepts, principles and terminology for clinical evidence related to in vitro diagnostic (IVD) medical devices, taking into account the differences between IVD medical devices and other medical devices.

It is anticipated that convergence of requirements for clinical evidence, including common data submissions, will lead to better understanding of an IVD medical device's scientific validity and performance by all stakeholders, more efficient use of resources of the clinical community, medical device regulators and industry, and increased transparency and confidence in the global regulatory model. Ultimately, there should be more efficient, predictable and timely access to IVD medical devices by patients and society worldwide.

When placing an IVD medical device on the market the manufacturer must have demonstrated through the use of appropriate conformity assessment procedures that the device complies with applicable essential principles of safety and performance of IVD medical devices as described in IMDRF/GRRP WG/N47 FINAL:2024 (Edition 2). Generally, from a clinical perspective, it is expected that the manufacturer has demonstrated the device achieves its intended performance during normal conditions of use in the intended environment (e.g. laboratories, physician's offices, healthcare centers, home environments) and in the intended use population. As IVD medical devices are used for the examination of specimens derived from the human body, some performance characteristics which are applicable to medical devices other than IVD may not be relevant. In turn, additional performance characteristics specific to IVD may be more appropriate.

2. Scope

The primary purpose of this document is to provide manufacturers with guidance on how to collect and document clinical evidence for an IVD medical device as part of the conformity assessment procedure prior to placing an IVD medical device on the market as well as to support its ongoing marketing. It is also intended to provide guidance to all those involved in the review and assessment of clinical evidence to support the marketing of IVD medical devices (Regulatory Authorities (RA), Conformity Assessment Bodies (CAB), other stakeholders).

This document provides the following guidance:

- the concepts and definitions of clinical evidence for IVD medical devices;
- general principles of clinical evidence;
- scientific validity and how to document it;
- analytical performance;
- clinical performance;
- how to demonstrate clinical performance (e.g. studies, literature);
- how to appraise and analyse the data; and
- the content of the clinical evidence report.

The guidance contained within this document applies to devices that meet the definition of an IVD medical device as outlined in ‘Definition of the Terms Medical Device and In Vitro Diagnostic (IVD) Medical Device’.

This document should not be used in isolation but read together with ‘Clinical evidence for IVD medical devices – Clinical Performance Studies for In Vitro Diagnostic Medical Devices’ as well as ‘Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices’ (references provided in the reference section).

3. References

IMDRF/GRRP WG/N47 FINAL:2024 (Edition 2) *Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices*

IMDRF/IVD WG/N64 *Principles of In Vitro Diagnostic (IVD) Medical Devices Classification*

GHTF/SG1/N044:2008 *Standards in Assessment of Medical Devices*

GHTF SG1/N046:2008 *Principles of CA for IVD Medical Devices*

GHTF/SG1/N071:2012 *Definition of the Terms Medical Device and In Vitro Diagnostic (IVD) Medical Device*

GHTF/SG5/N6:2012 *Clinical Evidence for IVD medical devices – Key Definitions and Concepts*

GHTF/SG5/N7:2012 *Clinical Evidence for medical devices – Scientific Validity Determination and Performance Evaluation*

GHTF/SG5/N8:2012 *Clinical evidence for IVD medical devices – Clinical Performance Studies for In Vitro Diagnostic Medical Devices*

ISO 18113-1:2022 *In vitro diagnostic medical devices — Information supplied by the manufacturer (labelling)*

4. Definitions

4.1 Analytical performance of an IVD medical device

Definition: The ability of an IVD medical device to detect or measure a particular analyte.

4.2 Clinical benefit of an IVD medical device

Definition: The impact of a device related to its function, such as that of screening, monitoring, diagnosis or aid to diagnosis of patients, or a positive impact on patient management or public health.

Explanation: The positive impact of a correctly performed result on patient management or public health. This includes how the IVD result contributes to diagnosis, screening, staging, monitoring, or prognosis of a disease or condition.

4.3 Clinical evidence for an IVD medical device

Definition: Clinical evidence for an IVD medical device is all the information that supports the scientific validity, analytical and clinical performance of the device, to allow an assessment of whether the device is safe and achieves its intended use as claimed by the manufacturer.

4.4 Clinical performance of an IVD medical device

Definition: The ability of an IVD medical device to yield results that are correlated with a particular clinical condition, physiological or pathological process/state in accordance with the intended use (clinical test purpose, intended use population and intended user).

4.5 Clinical performance study

Definition: A study undertaken to establish or confirm the clinical performance of an IVD medical device.

For further information, refer to GHTF/SG5/N8:2012 *Clinical evidence for IVD medical devices – Clinical Performance Studies for In Vitro Diagnostic Medical Devices*.

NOTE: This term is similar to ‘clinical investigation’ for medical devices other than IVD medical devices.

4.6 Companion diagnostics

Definition: Companion diagnostics (CDx) are IVD medical devices that are essential for the safe or effective use of a medicinal product/therapy. In some jurisdictions, the intended use of CDx may also include monitoring response to treatment with the therapeutic product for the purpose of adjusting treatment to achieve improved safety or effectiveness.

4.7 Diagnostic sensitivity (Clinical sensitivity)

Definition: The ability of an IVD medical device to identify the presence of a target marker associated with a particular disease or condition in the intended population.

NOTE 1: Also defined as percent positivity in specimens from subjects where the target disease or condition is known to be present.

NOTE 2: Diagnostic sensitivity is a number fraction, calculated as true positive values divided by the sum of true positive plus false negative values.

NOTE 3: The disease or condition is defined by criteria independent of the IVD medical device under consideration.

Source: Modified from ISO 18113-1:2022.

4.8 Diagnostic specificity (Clinical specificity)

Definition: The ability of an IVD medical device to recognize the absence of a target marker associated with a particular disease or condition in the intended population.

NOTE 1: Also defined as percent negativity in specimens from subjects where the target disease or condition is known to be absent.

NOTE 2: Diagnostic specificity is a number fraction, calculated as true negative values divided by the sum of true negative plus false positive values.

NOTE 3: The disease or condition is defined by criteria independent of the medical device under consideration.

Source: Modified from ISO 18113-1:2022.

4.9 Intended use/purpose

Definition: The objective intent regarding the use of a product, process or service as reflected in the specifications, instructions and information provided by the manufacturer.

NOTE: The intended use/purpose can include the indications for use.

Explanation: The intended use/purpose should be supported by the data supplied on the label, in the instructions for use, promotional or sales materials or statements, and should be specified by the manufacturer in the performance evaluation. The intended use/purpose for an IVD medical device should define the intended device function, such as diagnosis, aid to diagnosis, screening, monitoring, predisposition, prognosis, prediction, patient management and physiological/pathological status determination. Other relevant aspects include: the indications for use (e.g. specific disorder, condition or risk factor of interest), target marker, specimen type, data output type (e.g. qualitative or quantitative), technology/assay type/test principle, intended use population, intended use environment/settings (e.g. self-test, point of care) and the intended user.

This information will determine the type and depth of the clinical evidence as explained in later sections.

Source: Modified from GHTF/SG1/N77:2012.

4.10 In vitro diagnostic (IVD) medical device

Definition: In Vitro Diagnostic (IVD) medical device means a medical device, whether used alone or in combination, intended by the manufacturer for the *in vitro* examination of specimens derived from the human body solely or principally to provide information for diagnostic, monitoring or compatibility purposes.

Explanation: IVD medical devices include reagents, calibrators, control materials, specimen receptacles, software, and related instruments or apparatus or other articles and are used, for example, for the following test purposes: diagnosis, aid to diagnosis, screening, monitoring, predisposition, prognosis, prediction, determination of physiological status.

Source: Modified from GHTF/SG1/N071:2012.

4.11 Performance evaluation of an IVD medical device

Definition: Assessment and analysis of data to establish or verify the scientific validity, analytical and clinical performance of an IVD medical device.

Explanation: Performance evaluation for an IVD medical device is the investigation process by which data are assessed and analysed to demonstrate the scientific validity, analytical and clinical performance of the envisioned IVD medical device for the intended use/purpose as stated by the manufacturer. Data are typically generated from verification and validation studies (including, where appropriate, clinical performance studies using human specimens) or obtained from literature reviews that confirm the performance of the product. Performance evaluation should also include post-market data (e.g. adverse event reports, post-market surveillance reports, published literature) to support a life-cycle approach to generating clinical evidence for IVD medical devices. The performance evaluation supports the demonstration of the conformity to the relevant essential principles of safety and performance.

4.12 Performance of an IVD medical device

Definition: The ability of an IVD medical device to achieve its intended use/purpose as claimed by the manufacturer. The performance of an IVD medical device consists of the analytical and clinical performance supporting the intended use of the IVD medical device.

4.13 Predictive value

Definition: The probability that a person with a positive result has a given condition under investigation, or that a person with a negative result does not have a given condition.

NOTE 1: The predictive value is determined by the diagnostic sensitivity and diagnostic specificity of the IVD medical device, and by the prevalence of the disease or condition for which the IVD medical device is used in the intended use population. As such, predictive value is not an intrinsic property of the test.

NOTE 2: Prevalence means the proportion of persons with a particular disease or condition within a given population at a given time.

NOTE 3: The positive predictive value is the probability that following a positive result, that individual will truly have that specific disease or condition, it indicates how effectively an IVD medical device distinguishes true positive results from false positive results for a given disease or condition in a given population.

NOTE 4: The negative predictive value is the probability that following a negative result, that individual will truly not have that specific disease or condition, it indicates how effectively an IVD medical device distinguishes true negative results from false negative results for a given disease or condition in a given population

Source: Modified from ISO 18113-1:2022.

4.14 Recognized standards

Definition: Standards deemed to offer the presumption of conformity to specific Essential Principles of Safety and Performance[1].

Source: GHTF SG1/N044:2008.

4.15 Scientific validity of an analyte

Definition: The association of an analyte to a clinical condition/physiological state.

4.16 State of the art

[1] IMDRF/GRRP WG/N47 FINAL:2024 (Edition 2) Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices

Definition: Developed stage of technical capability at a given time in regards to products, processes and services, based on the relevant consolidated findings of science, technology and experience.

NOTE: The state of the art embodies what is currently and generally accepted as good practice in technology and medicine. The state of the art does not imply the most technologically advanced solution. The state of the art described here is sometimes referred to as the “generally acknowledged state of the art”.

Source: IMDRF/GRRP WG/N47 FINAL:2024 (Edition 2).

4.17 Technical documentation

Definition: The documented evidence, normally an output of the quality management system and performance evaluation, that demonstrates compliance of a device to the applicable essential principles of safety and performance.

Source: Modified from GHTF SG1/N046:2008.

5. General principles of clinical evidence

Clinical evidence for an IVD medical device is all the information on the scientific validity, analytical and clinical performance of the device, to allow an assessment of whether the device is safe and achieves its intended use/purpose as claimed by the manufacturer.

Clinical evidence is an important component of the technical documentation of an IVD medical device, which along with other design verification and validation documentation, device description, labelling, risk analysis and manufacturing information, is needed to allow a manufacturer to demonstrate conformity with the applicable essential principles of safety and performance[2]. Clinical evidence is used to support the labeling and marketing of the IVD medical device, including any claims made about the scientific validity and performance of the device.

To understand clinical evidence for IVD medical devices, it must be taken into account that IVD medical devices differ from other medical devices in that the risks and benefits they pose are related to impact on patient management and public health rather than direct interaction between the device and the patient. A significant percentage of all clinical decisions rely on information provided by IVD medical devices and these decisions can profoundly influence diagnosis and patient management.

Clinical evidence requirements are influenced by the risk class of the device, the degree of innovation reflected in the scientific/clinical recognition of the analyte/target marker, the assay technology, the novelty, the degree of variability of the intended use population and disease state, and the intended user(s) of the device. For the intended use/purpose of the IVD medical device, clinical evidence must:

- demonstrate that the IVD medical device achieves the intended clinical benefit as stated by the manufacturer
- address the relative risks to the patient and user, as applicable, associated with the use of the device.

The process of establishing the clinical evidence for an IVD medical device is termed performance evaluation, which consists of gathering and assessing information for three major pillars: scientific validity, analytical performance and clinical performance of the IVD medical device.

[2] IMDRF/GRRP WG/N47 FINAL:2024 (Edition 2) Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices

6. Performance evaluation of IVD medical devices

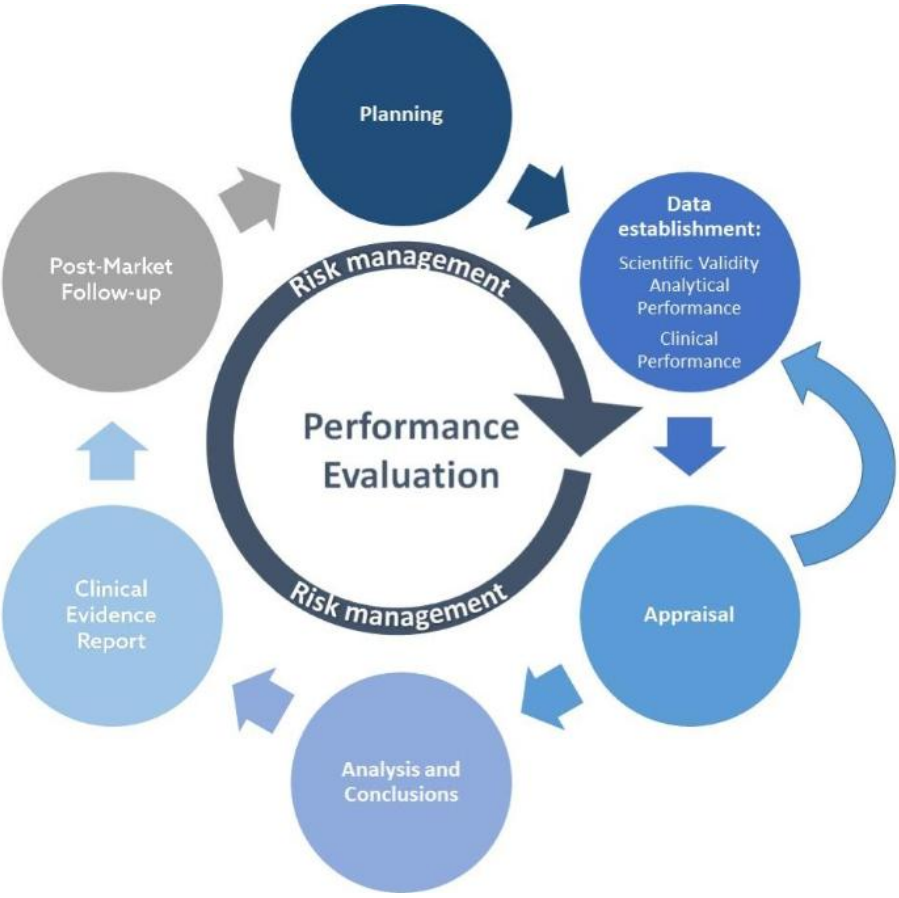

Figure 1: Lifecycle approach to performance evaluation of IVD medical devices.

Performance evaluation should be regarded as a continuous process where clinical evidence is demonstrated prior to placing the device on the market and updated as needed throughout the entire lifecycle of an IVD medical device, including after it is placed on the market (the "lifecycle approach", see Figure 1). The manufacturer should regularly review the scientific validity and performance of the device from post-marketing information (such as adverse event reports, results from post-market performance studies, published literature) to determine possible impact on the benefits and risks of the IVD medical device throughout the device lifecycle. The manufacturer should also take into account scientific developments and improvements in the state of the art.

For IVD medical devices that have a new intended use/purpose, the existing clinical evidence should be assessed as to its suitability to support the new intended use/purpose. If the new intended use/purpose is not supported by the existing evidence, then additions to the scientific validity, analytical performance and clinical performance should be demonstrated.

In what follows, it is important to distinguish between the terms “document”, “demonstrate” and “generate” in relation to clinical evidence or its components. “Documentation” refers to written consideration of the relevant concepts in the technical documentation of the device. “Demonstration” refers to laying out the available evidence, including already existing evidence such as published literature. “Generation” refers to production of new evidence through a study. Consideration of scientific validity, analytical performance and clinical performance must always be documented by the manufacturer. In general, this involves demonstration of scientific validity, analytical performance and clinical performance, with some exceptions as explained below. However, the manufacturer may not always need to generate new evidence – this is needed only when existing evidence is not sufficient.

7. Scientific validity determination

7.1. Is the scientific validity of the analyte established?

Scientific validity must be demonstrated for each IVD medical device.

Scientific validity can be demonstrated through the use of existing evidence, where available. Generation of new evidence for scientific validity is not necessary where the association of an analyte to a clinical condition, physiological or pathological process/state is well established, based on available information such as peer-reviewed literature, textbooks and published consensus expert opinions. The manufacturer is expected to demonstrate the association through consideration of appropriate published literature and document the scientific validity for the IVD medical device. If the scientific validity is well established and existing evidence considered sufficient, a brief rationale should be documented in the final clinical evidence report (e.g. a brief list of key references).

For a new analyte or new intended use/purpose, the availability of existing evidence may be limited, and prospective data may need to be generated to demonstrate scientific validity.

As scientific and medical knowledge further develops, the initially established scientific validity might change and/or expand.

There are certain situations where it may not be necessary to demonstrate scientific validity due to the nature of the device (e.g. instrumentation, calibrators, controls). For certain devices intended to be used together, for example a reagent intended to be used with calibrators and controls, or an instrument to be used with reagents, it may be more appropriate to establish the scientific validity in the context of this combination rather than for each device individually. As certain instruments may have an independent measuring function which does not use any additional reagents, scientific validity determination would also be required unless sufficient justification is provided. In all cases the scientific validity must be documented, including a justification for where scientific validity is not relevant.

7.2. Identification of scientific validity of the analyte

Potential sources for the identification of scientific validity information are:

- Information on IVD medical devices that measure the same analyte and with the same intended use/purpose that have marketing history (e.g. Instructions for Use)
- Literature review: this information might be found in peer-reviewed articles, regulatory guidance documents, conference proceedings, etc.
- Review of expert opinions: this information might be found in sources that include textbooks, clinical guidance documents, published consensus expert opinions
- Results from proof of concept/feasibility studies: these studies are usually smaller scale scientific studies to identify the fundamental association of the analyte with the clinical condition/physiological state

- Results from clinical performance studies.

Scientific validity can be based on one or more of these potential information sources.

7.3. Appraisal and analysis of scientific validity data and information

The purpose of scientific validity appraisal is to assess data and information based on its relevance, merits and limitations. Each piece of data or information is appraised to determine its relevance and quality, including the method used, for establishing the association between the analyte and the clinical condition, physiological or pathological process/state.

To be relevant, the source should reference both the target marker and the clinical condition, physiological or pathological process/state in question, and should show a direct association between the two. The data or information provided should be of sufficient quality and detail to enable a rational and objective assessment of the scientific validity and should reflect the state of the art.

The scientific validity analysis aims to collectively evaluate all of the appraised information, in terms of weight and significance. The outcome of the analysis should include references, justifications and conclusions supporting a valid association between the analyte and the clinical condition/physiological state.

The scientific validity must be documented as part of the clinical evidence report.

8. Analytical and clinical performance

Analytical and clinical performance must be documented for each IVD medical device. Generally, this will involve the demonstration of the performance using already existing evidence, and generation of new evidence where existing evidence is not sufficient.

For established devices/technologies, there may be published literature and experience with the use of the IVD medical device which may serve as sources for performance data, provided that the manufacturer shows that the data is relevant to the IVD medical device in question. For novel devices/technologies, existing evidence may be limited, therefore there may be a greater need to generate new evidence. This includes novel technologies, for which it should be demonstrated that the IVD medical device is technically capable of achieving the intended use/purpose, using appropriate performance parameters for the technology/device in question.

8.1. Analytical performance

Analytical performance studies are always required for each IVD medical device. The demonstration of analytical performance supports the intended use/purpose of the IVD medical device. Analytical performance is determined by the collection of testing results (analytical performance data) from analytical performance studies used to assess the ability of the IVD medical device to measure a particular analyte.

Analytical performance may include, but is not limited to:

- analytical sensitivity as measured by limit of blank, limit of detection, limit of quantitation and inclusivity
- analytical specificity that measures cross-reactivity, endogenous and exogenous interference
- accuracy that is derived from trueness and precision
- precision as measured by reproducibility and repeatability
- linearity
- cut-off validation
- measuring interval (range)
- carry-over
- traceability of calibrators and controls
- stability of the specimen and device stability using shelf-life, in use, transport measures.

It should be noted that analytical studies for trueness are usually based on one or more of the following:

- methods described in a recognized standard

- comparison with an international reference method (e.g. atomic absorption for calcium)
- comparison with a reference material of a higher order (e.g. international reference materials)

Where the methods described above are not readily available, it may be possible to perform a comparison with an already available IVD medical device (e.g. method comparison between immunoassays) or a comparison to a recognized method, (e.g. Sanger sequencing for genotyping assays).

For novel devices, it may not be possible to demonstrate trueness since recognized reference materials or reference method are not likely to be available. If there are no comparative methods then different approaches can be used with a justification (e.g. comparison to some other well-documented method). In the absence of such approaches, a clinical performance study comparing test performance to the current clinical standard practice would be needed.

In most circumstances, analytical performance testing using human specimens is required. If specimens are of limited availability or do not cover the desired range of the assay or presence of the target marker, the use of contrived samples may be acceptable with appropriate justification and demonstration of comparable characteristics to human specimens.

For the purpose of reviewing clinical evidence for the IVD medical device in question, an assessment needs to be made whether the analytical performance data in conjunction with the scientific validity is sufficient to declare conformity with relevant essential principles of safety and performance[3].

The analytical performance data must be summarised and documented as part of the clinical evidence report.

8.2. Clinical performance

The clinical performance of an IVD medical device may include, but is not limited to:

- diagnostic/clinical sensitivity and diagnostic/clinical specificity,
- negative and positive predictive value,
- expected normal and abnormal distributions in test populations.

The parameters depend on the intended use/purpose of the IVD medical device (diagnosis, screening, classification, therapy selection) and other relevant aspects such as intended use environment/settings (e.g. self-test, point of care) and the intended user (e.g. qualified healthcare professional, lay person).

Clinical performance must always be documented. Clinical performance must be demonstrated (using existing and/or newly generated evidence) unless it is justified by the manufacturer that this is not applicable, for example for certain low-risk IVD medical devices. Whether the IVD medical device in question is established or novel does not exempt the manufacturer from the requirement to demonstrate clinical performance. However, in case of a novel IVD medical device it is more likely that it is necessary to generate new evidence through a clinical performance study to demonstrate clinical performance.

[3] IMDRF/GRRP WG/N47 FINAL:2024 (Edition 2) Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices

The clinical performance data should be documented as part of the clinical evidence report.

8.2.1. Identification of clinical performance

Manufacturers are able to draw on one or more data sources to demonstrate clinical performance. Potential sources for the identification of clinical performance information are:

- Clinical Performance Studies: Clinical performance studies carried out by or on behalf of a manufacturer.
- Literature: Scientific peer-reviewed published studies of clinical performance studies for the IVD medical device.
- Real-world data/evidence: Published experience gained by routine diagnostic testing or the delivery of health care from a variety of sources other than clinical performance studies, pre- and post-market data.

Determination of clinical performance needs to be based on representative and appropriately varied populations. The use conditions and environment need to be similar to those specified in the IVD medical device's intended use/purpose. Data relevant to the clinical performance may be held by the manufacturer (e.g. manufacturer sponsored pre- and post-market study reports and adverse event reports for the IVD medical device in question) or in the scientific literature (e.g. published articles of clinical performance studies for the IVD medical device).

The manufacturer is responsible for identifying data relevant to the IVD medical device and determining the types and amount of data needed to establish clinical performance, giving consideration to the advantages and limitations of each data type. Literature should be assessed to see if it is fit-for-purpose to support demonstration of clinical evidence and regulatory decision making, including an evaluation of both the relevance and the reliability of data.

8.2.2. Potential sources of clinical performance data

8.2.2.1 Clinical performance studies

The guidance included within this section applies to clinical performance studies carried out by or on behalf of a manufacturer specifically for the purposes of conformity assessment in accordance with applicable regulations. Such studies are generally expected to be designed, conducted and reported in accordance and in compliance with local regulations and guidance.

Clinical study design is crucial to determining clinical performance and the claims a manufacturer intends to pursue, particularly in the context of conformity assessment. The study design should ensure that the generated data is robust and sufficient to support the intended claims.

The design of clinical performance studies must reflect the nature of the IVD medical device and its intended use/purpose. Further explanation of clinical performance study design is discussed in GHTF/SG5/N8:2012 *Clinical evidence for IVD medical devices – Clinical Performance Studies for In Vitro Diagnostic Medical Devices*.

While not intended to impose unnecessary burdens, the clinical performance studies must provide evidence to support the intended use/purpose of the IVD medical device while addressing the relative risks/benefits associated with the use of the IVD medical device.

8.2.2.2 Literature

Systematic literature review can be used to identify published clinical performance data that is not in the possession of the manufacturer that may assist the manufacturer in demonstrating acceptable clinical performance of an IVD medical device. Systematic reviews aim to provide comprehensive and unbiased summary synthesis of existing evidence on a single topic by using scientific methods to synthesise multiple individual studies in a single place.

Systematic reviews can be used to synthesise evidence about the association of an analyte with a clinical condition or physiological state to determine scientific validity, or to estimate the analytical and clinical performance of an IVD medical device at any stage of development. The data generated through literature review should relate directly to the IVD medical device in question.

Specifically, systematic reviews can be used to:

- provide more precise estimates than can be obtained from single studies of association between an analyte and a clinical condition/physiological state for scientific validity or of diagnostic or prognostic analytic or clinical performance
- undertake comparisons of analytes (scientific validity) or performance (analytic or clinical) in the absence of primary comparative evidence
- understand what factors moderate associations or cause variation in performance.

When conducting a systematic review of the literature, there are core methodological steps that need to be taken to minimise the risk of bias being introduced into the review process:

1. Definition of a precise review question
2. A comprehensive search for published and unpublished literature using bibliographic databases and grey literature and based on the pre-defined review question
3. Selection of evidence for inclusion using objective criteria derived from the pre-defined review question
4. Quality assessment of included evidence. This assessment should be performed at the individual study level and using tools developed for this purpose. A variety of quality assessment tools are available according to the type of study/evidence (e.g. randomised controlled trial, test accuracy study). A list of quality assessment tools is provided in Appendix A.
5. Comprehensive reporting of methods and findings (including reasons for exclusion of certain studies) to ensure transparency and reproducibility

A narrative synthesis of evidence is always performed as part of a systematic review. For systematic reviews with numerical outcome measures (e.g. association, accuracy or effectiveness), statistical synthesis (meta-analysis) is appropriate if included studies are sufficiently similar that derivation of a single, summary estimate to represent the results of all included individual studies is considered appropriate and meaningful.

The above key steps are common to all types of systematic review, but the methods for undertaking them differ according to the review type (the methodological type of primary study being included in the review). For example, review question components to be defined as part of step 1 will differ for a review of scientific validity compared to a review of performance. Sources of relevant literature and terms to be used for step 2 will need to be adapted depending on the type of literature to be included.

The merits and limitations of published and unpublished literature should be considered. Unpublished literature, or grey literature, may not be widely accessible or standardised in terms of format and methodology, however, could be more up to date than published literature and is not affected by publication bias. Likewise, commercial publication is not a guarantee of quality and should not be a pre-requisite for inclusion in a review. Quality assessment tools used in step 4 will differ for different evidence types. If meta-analysis is undertaken as part of synthesis, statistical methods to combine studies will differ according to the outcomes being measured. There are a range of resources available to support the conduct and reporting of systematic reviews. Resources relevant to systematic reviews of scientific validity, analytical and clinical performance are detailed in Appendix A.

Once the literature review has been executed, a summary should be prepared and included in the clinical evidence report. A copy of the methodology (e.g. databases, search terminology) should be included and any deviations noted. It is important that the literature review is documented in sufficient detail so that the methods can be appraised critically, the results can be verified, and the review reproduced if necessary.

It is recognised that in cases where manufacturers source clinical performance data reported in the scientific literature (i.e. studies of the IVD medical device in question undertaken by a third party), the documentation readily available to the manufacturer for inclusion in the clinical evidence report is likely to be limited to the published paper itself.

8.2.2.3 Real-world data/evidence

These types of performance data are generated in actual use conditions that are outside the conduct of clinical performance studies. Use of this evidence is part of the lifecycle approach to performance evaluation described above. However, use of real-world data/evidence alone may not be sufficient to demonstrate clinical performance or extending the intended use/purpose of an IVD medical device.

While much of the experience with routine diagnostic testing is found in literature, additional data may include:

- manufacturer-generated post-market surveillance data, studies and reports;
- adverse events data, vigilance reports (held by either the manufacturer or Regulatory Authorities) and clinically relevant evidence related to field safety corrective actions;
- data from proficiency testing schemes, external quality assessment (EQA), and ring-trials (inter-laboratory comparison);
- investigator-initiated studies;
- studies of device performance by third-party organizations published in peer-reviewed literature.

The value of real-world data/evidence is that it provides actual use experience obtained in larger, heterogeneous and more complex populations (e.g. with regard to interfering substances). The data is most useful for identifying less common but potentially serious device-related adverse events. It is also a particularly useful source of diagnostic testing data for low-risk devices that are based on long standing, well-characterized technology and, therefore, unlikely to be the subject of either reporting in the scientific literature or performance study.

The manufacturer should choose and justify a methodology for gathering and appraising real-world data/evidence. This methodology should be objective so that both favourable and unfavourable data are included. Real-world data/evidence may lead to either extension or limitation of the intended use/purpose of an IVD medical device.

It is important that any reports or collations of data contain sufficient information to be able to undertake a rational and objective assessment of the information and make a conclusion about its significance with respect to the performance of the IVD medical device in question. Reports of such experience that are not adequately supported by data, such as anecdotal reports or opinion, should not be used.

8.2.3. Appraisal and analysis of clinical performance data

The purpose of the clinical performance appraisal is to assess information based on its merits and limitations. Each piece of information is appraised to determine its relevance and quality to address questions about the IVD medical device, and its contribution to demonstrate the clinical performance of the IVD medical device (including any specific claims about performance).

To be relevant, the information source should be specific to the IVD medical device in question and reflect its intended use/purpose.

The information provided should be of sufficient quality and details to enable a rational and objective assessment of the clinical performance.

The clinical performance appraisal aims to collectively evaluate all of the documented information, in terms of weight and significance. For the purpose of clinical performance evaluation, clinical performance study data is typically weighted higher than literature data and real-world data/evidence.

When weighting the results, particular attention should be given to circumstances where there are repeated publications on the same group of patients by the same authors, in order to avoid over-weighting the experience.

The different data sets should be reviewed for consistency of results across multiple sites and as appropriate, the intended use populations.

If the different data sets report comparable performance characteristics, certainty about the clinical performance increases. If different results are observed across the data sets, it will be helpful to determine the reason for such differences. Regardless, all data sets should be included.

As a final step, the evaluator should consider the basis on which it can be demonstrated that the combined data show:

- the IVD medical device performs as intended by the manufacturer; and
- any risks associated with the use of the IVD medical device are acceptable when weighed against the benefits to the patient; and
- compliance with the relevant essential principles of safety and performance.

Such considerations should take into account the (non-exhaustive list):

- intended users and the intended use environment
- intended use population
- number and severity of adverse events, the adequacy of the estimation of associated risk for each identified hazard
- severity and natural history of the condition being diagnosed or treated

- availability of alternative diagnostic tests and current standard of care
- novelty and degree of innovation of the device
- state of the art.

The outcome of the analysis should be summarised in a report comprising references, justifications and conclusions regarding the clinical performance of the IVD medical device.

The instructions for use should be reviewed to ensure they are consistent with the clinical performance data and that all the hazards and other clinically relevant information have been identified appropriately.

9. Specific considerations for certain IVD medical devices

9.1. Software as an IVD

Software as an IVD medical device (SaIVD) exhibit unique characteristics compared to traditional or hardware IVD medical devices that warrant additional regulatory considerations. For example, software typically undergoes rapid development cycles, frequent algorithm changes, remote updates with version controls. Post-market changes for SaIVD are often driven by adaptive (e.g. to keep pace with changes in operating systems, network and platforms) and perfective needs (e.g. to add new features), rather than corrective actions, which are more common in traditional IVD medical devices. These continual and cumulative changes mean that the clinical evidence for SaIVD are subject to more frequent and rapid re-verification throughout the product lifecycle compared to traditional IVD medical devices.

The clinical evidence should demonstrate that the SaIVD meets the essential requirements specific to software devices as outlined in IMDRF/GRRP WG/N47. This also applies to electronic programmable systems, including software, that drive or influence or is an accessory to an IVD medical device.

The scientific validity of a SaIVD's output (e.g. risk score, diagnosis) must be established. For novel technologies such as machine learning or artificial intelligence (AI), the scientific validity of the technology itself may also need to be demonstrated. For example, in the case of an AI-based software intended to identify cancer cells in pathology slides as an alternative to a pathologist, the biological characteristics of cancer cells are well-established; however, the novel use of AI in a context traditionally reliant on human expert interpretation introduces a specific need to independently establish scientific validity.

The analytical performance of a SaIVD is the ability of the medical device to generate the intended technical output from the input data. Analytical validation is required for any SaIVD to support its intended use/purpose. In addition to the analytical performance characteristics described in Section 8.1, SaIVD require verification that the software algorithm produces accurate and reliable outputs across all anticipated input scenarios. This includes testing the SaIVD against a representative range of input data, ensuring that the software consistently interprets signals, measurements, or images according to its intended use/purpose.

For some SaIVD, particularly those that employ machine learning or AI algorithms to recognize complex or hidden patterns in datasets, it may be difficult to define or establish appropriate methods for validating certain analytical performance parameters. For example, a SaIVD may detect expression patterns of a particular analyte in tissue sections associated with a particular clinical condition in ways that cannot be readily recognized or verified even by trained medical experts. In such cases, both the exact nature of the analyte and its corresponding ground truth may be unclear. When it is difficult to demonstrate certain analytical performance characteristics, greater emphasis may be placed on clinical performance outcomes, such as predictive values, clinical specificity and sensitivity, rather than on analytical performance characteristics that are directly linked to the analyte.

When establishing the clinical evidence of SaIVD that incorporate AI models with interpretative functions, additional considerations should include:

1. Documentation and traceability of the entire data lifecycle – from collection to training, testing, and validation. This includes:
   - Identifying data sources and their provenance
   - Ensuring data representativeness and diversity to avoid bias (geographically and demographically diverse datasets)
   - Maintaining version control and audit trails for datasets used in model development
2. Continuous monitoring for performance drift due to the dynamic and adaptive nature of AI algorithms, particularly as the SaIVD is exposed to new clinical environments and patient populations. The manufacturer should:
   - Implement robust risk management plans that include AI-specific failure modes
   - Establish post-market surveillance mechanisms to detect and respond to anomalies or adverse events linked to AI outputs
3. Transparency about model performance and limitations. The manufacturer should:
   - Provide clinically interpretable outputs
   - Ensure that healthcare professionals can understand and validate the AI's recommendations (human – AI team performance)
   - Avoid over-reliance on AI by designing systems that inform and support human clinical judgement rather than replace it. While AI technology can significantly enhance the interpretative capabilities of IVD medical devices, the final clinical decision must always rest with the qualified healthcare professionals.

Sources:

- IMDRF/GRRP WG/N47 *Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices*
- IMDRF/SaMD WG/N41FINAL:2017 *Software as a Medical Device (SaMD): Clinical Evaluation*
- IMDRF/SaMD WG/N23 FINAL: 2015 *Software as a Medical Device (SaMD): Application of Quality Management System*
- IEC PAS 63621:2025 *Artificial Intelligence Enabled Medical Devices - Data Management.*

9.2. Companion diagnostics

The clinical evidence for companion diagnostics (CDx) differs from other types of IVD medical devices with regards to the association of its clinical benefits with the medicinal product/therapy. The intended use/purpose of a CDx is to identify a patient population who are most likely to benefit from a particular therapy, or patients who are at increased risk of serious side effects from a particular therapeutic product. In some jurisdictions, the intended use/purpose of CDx may also include ongoing monitoring of a patient's response to treatment with the therapeutic product, for the purpose of adjusting treatment to achieve improved safety or effectiveness. Therefore, the primary clinical performance characteristic of a CDx is the clinical outcome(s) provided by the therapeutic intervention in the population identified by the CDx. Appropriate identification of the intended population is critical because false-positive results may harm patients by resulting in administration of unnecessary therapy, while false-negative results may deprive patients of potentially beneficial treatment, with both scenarios potentially distorting estimates of therapeutic effects and response proportions. Therefore, therapeutic product and IVD medical device manufacturers should collaborate closely to understand how the analytical performance of the CDx affects subject selection in all validation studies, ensuring appropriate analytical validation studies are conducted and potential sources of bias are addressed before utilising the CDx within a clinical trial study.

Due to the direct association between a CDx and its corresponding medicinal product, analytical and clinical performance studies are generally required to demonstrate the clinical evidence for the device. Multiple study designs may be employed to generate this evidence. Often, the clinical performance study of a CDx is conducted in parallel with the clinical trial of the associated medicinal product, commonly referred to as a combined trial. Two primary clinical trial designs are commonly employed:

- Therapy stratification: this trial design enrolls both CDx-positive and CDx-negative subjects with randomization, providing the most informative approach for assessing treatment-by-marker interactions and distinguishing prognostic versus predictive value
- Therapy selection: this trial design enrolls only CDx-positive subjects, which cannot determine predictive value since no data exists for the negative population.

The clinical performance of a CDx should be established by the pivotal clinical trials that support regulatory approval of the corresponding medicine or biological indication. Alternatively, when a clinical trial assay has been used in the clinical development of the medicinal product and in the pivotal clinical trials, a clinical bridging study may be used to establish the clinical performance of a subsequent CDx following the clinical trial. The bridging study should establish clinical comparability between the clinical trial assay and the subsequent CDx, which can be achieved through provision of direct or indirect comparability data.

Direct clinical comparability data comes from a study in which clinical samples are tested using both the clinical trial assay and subsequent CDx. The samples may be:

- Samples from the original pivotal clinical trial; or
- Samples from patients who are clinically similar to those who participated in the pivotal clinical trial.

Indirect clinical comparability data may consist of:

- Robust evidence that the scientific validity, analytical and clinical performance of the subsequent CDx are highly comparable to the clinical trial assay; and

- Scientific and clinical justification that a direct clinical comparability study is not required to conclude that use of the proposed subsequent CDx will result in clinically comparable outcomes to those seen with the clinical trial assay in the pivotal clinical trial.

New intended uses of an approved CDx device, including use in new indications or with different medicinal products, require additional clinical performance data to support the expanded intended use.

Sources:

- In Vitro Companion Diagnostic Devices, Guidance for Industry and Food and Drug Administration Staff, issued August 2014
- MDCG 2020-16 Rev.3 - Guidance on Classification Rules for in vitro Diagnostic Medical Devices under Regulation (EU) 2017/746, revised July 2024
- MDCG 2022-2 Guidance on general principles of clinical evidence for In Vitro Diagnostic medical devices (IVDs), published January 2022:
- MedTech Europe Regulatory eBook Clinical Evidence Requirements under the EU In Vitro Diagnostics Regulation (IVDR), 4th edition, published August 2025
- Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product, Draft Guidance for Industry and Food and Drug Administration Staff, issued July 2016.
- Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU
- Understanding regulatory requirements for in vitro diagnostic (IVD) companion diagnostics (CDx), Guidance by Australian Therapeutic Goods Administration, published December 2019

10. Clinical evidence report

The clinical evidence report is a compilation of the scientific validity, analytical and clinical performance. If the manufacturer concludes there is insufficient clinical evidence to be able to declare conformity with the relevant essential principles of safety and performance, the manufacturer will need to generate additional data (e.g. conduct a clinical performance study, broaden the scope of literature searching) to address the deficiency or amend the intended use/purpose to match the clinical evidence. In this respect, development of clinical evidence, and in particular, clinical performance, should be an iterative process.

The clinical evidence report may reference or be a compilation of the information related to scientific validity, analytical performance and clinical performance. It is important that the report outline:

- the justification for the overall approach taken to gather the clinical evidence;
- the technology on which the medical device is based, the intended use/purpose of the device and any claims made about the device's clinical performance or safety;
- the nature and extent of the scientific validity and the performance data that has been evaluated;
- how the referenced information (recognized standards and/or clinical data) demonstrate the clinical performance and safety of the device in question; and
- the literature search methodology, if a literature review is the approach taken to gathering clinical evidence.

It should be noted that the level of detail in the report content will vary according to the risk class of the IVD medical device and its intended use/purpose.

Appendix A: Guidance for the conduct and reporting of systematic reviews

Best practice for the conduct and reporting of systematic reviews varies according to the review type. The review type is determined by the research question to be answered which in turn determines the type of research evidence that is to be synthesised. Guidance to support the conduct and reporting of systematic reviews relevant to scientific validity, analytic and clinical performance are included below.

1. Systematic reviews of scientific validity

A systematic review question concerned with aetiology should be constructed using the following components: Population; Exposure of interest (analyte) and Outcome (clinical condition/ physiological state).

Systematic reviews concerned with establishing evidence for an association will be synthesising observational epidemiological evidence that may include cohort, case control, cross-sectional studies and basic research studies.

Methods:

Dekkers OM, Vandenbroucke JP, Cevallos M, Renehan AG, Altman DG, Egger M. COSMOS-E: Guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. PLoS Med. 2019 Feb 21;16(2):e1002742. doi: 10.1371/journal.pmed.1002742. PMID: 30789892; PMCID: PMC6383865.

Resources on reporting:

- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. J Clin Epidemiol. 2021 Jun;134:178-189. doi: 10.1016/j.jclinepi.2021.03.001. Epub 2021 Mar 29. PMID: 33789819.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008 Apr;61(4):344-9. doi: 10.1016/j.jclinepi.2007.11.008. PMID: 18313558.

2. Systematic reviews of analytical or clinical performance

A systematic review question concerned with accuracy should include the following components: Population (including the setting) in which the test is intended to be used; Index test (IVD(s) being evaluated; Target condition (disease of condition to be detected by the test). Systematic reviews of test accuracy may include a range of study designs, the common study design feature being a comparison of index test results with a more accurate, reference standard test (the best available method for determining if the population have the target condition).

Methods:

- Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from https://training.cochrane.org/handbook-diagnostic-test-accuracy/current.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046.
- Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, Deeks JJ, Leeflang MMG; QUADAS-C Group†; Bossuyt PMM, Brazzelli MG, Dinnes J, Gurusamy KS, Jones HE, Lange S, Langendam MW, Macaskill P, McInnes MDF, Reitsma JB, Rutjes AWS, Sinclair A, de Vet HCW, Virgili G, Wade R, Westwood ME. QUADAS-C: A Tool for Assessing Risk of Bias in Comparative Diagnostic Accuracy Studies. Ann Intern Med. 2021 Nov;174(11):1592-1599. doi: 10.7326/M21-2234. Epub 2021 Oct 26. PMID: 34698503.

Reporting:

McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM; and the PRISMA-DTA Group; Clifford T, Cohen JF, Deeks JJ, Gatsonis C, Hooft L, Hunt HA, Hyde CJ, Korevaar DA, Leeflang MMG, Macaskill P, Reitsma JB, Rodin R, Rutjes AWS, Salameh JP, Stevens A, Takwoingi Y, Tonelli M, Weeks L, Whiting P, Willis BH. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

3. Additional resources

- Lattitudes (https://www.latitudes-network.org/) is a repository of validity assessment tools for different study designs including a range of systematic review types
- The equator network (https://www.equator-network.org/) is a repository of reporting guidelines for different study designs including a range of systematic review types

Prognostic accuracy:

- Debray T P A, Damen J A A G, Snell K I E, Ensor J, Hooft L, Reitsma J B et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017; 356 :i6460 doi:10.1136/bmj.i646
- Riley R D, Moons K G M, Snell K I E, Ensor J, Hooft L, Altman D G et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ 2019; 364 :k4597 doi:10.1136/bmj.k4597
- Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. (2014) Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med 11(10): e1001744 doi:10.1371/journal.pmed.1001744
- Snell KIE, Levis B, Damen JAA, Dhiman P, Debray TPA, Hooft L, Reitsma JB, Moons KGM, Collins GS, Riley RD. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). BMJ. 2023 May 3;381:e073538 doi: 10.1136/bmj-2022-073538. PMID: 37137496; PMCID: PMC10155050.

Appendix B: Possible format for the literature search report

1. **Device name/model**
2. **Scope of the literature search** [should be consistent with the scope of the scientific validity, analytical or clinical performance]
3. **Methods**
   (i) Date of search
   (ii) Name of person(s) undertaking the literature search
   (iii) Period covered by search
   (iv) Literature sources used to identify data

- scientific databases – bibliographic (e.g. PubMed, MEDLINE, EMBASE);
- specialized databases (e.g. MEDION)
- systematic review databases (e.g. Cochrane Collaboration);
- clinical trial registers (e.g. CENTRAL, NIH);
- adverse event report databases (e.g. MAUDE, IRIS)
- reference texts

[Include justification for choice of sources and describe any supplemental strategies (e.g. checking bibliography of articles retrieved, hand searching of literature) used to enhance the sensitivity of the search]

   (v) Database search details

- search terms (key words, indexing headings) and their relationships (Boolean logic)
- medium used (e.g. online, use of AI) [Attach copy of downloaded, unedited search strategy]

   (vi) Selection criteria used to choose articles

4. **Outputs**
   (i) Attach copy of literature citations retrieved from each database search
   (ii) Data selection process

[Attach flow chart and associated tables showing how all citations were assessed for suitability for inclusion in the clinical evidence (see Appendix A)]

NOTES:
EMBASE Excerpta Medica published by Elsevier
CENTRAL The Cochrane Central Register of Controlled Trials
IRIS The TGA’s medical device **I**ncident **R**eport **I**nvestigation **S**cheme
MAUDE US FDA’s **M**anufacturer **A**nd **U**ser Facility **D**evice **E**xperience database

Please visit our website for more details.

www.imdrf.org