**Article:**
(2025) Development of a minimum checklist to assess the quality of evidence produced using registry data for the evaluation of medical device safety and performance. BMJ Surgery, Interventions, and Health Technologies. e000364. ISSN: 2631-4940

https://doi.org/10.1136/bmjsit-2024-000364

BMJ Surgery, Interventions, & Health Technologies

# Development of a minimum checklist to assess the quality of evidence produced using registry data for the evaluation of medical device safety and performance

Lotje A. Hoogervorst ,[1] Rob G.H.H. Nelissen ,[1] Tom Melvin,[2] Paul Piscoi,[3] Chris Wilkinson,[4,5] Anne Lubbeke,[6,7] Chris P. Gale,[8,9,10] David Epstein,[11] Soren Overgaard,[12,13] Phil Walmsley,[14,15] Piotr Szymanski,[16,17] Maziar Mohaddes,[18,19,20] Donal B O'Connor,[21,22] Robert E. Geertsma ,[23] Joëlle M. Hoebert,[23] Alan G. Fraser ,[24] Perla J. Marang-van de Mheen,[25] CORE-MD study group

**To cite:** Hoogervorst LA, Nelissen RG.H.H, Melvin T, *et al* . Development of a minimum checklist to assess the quality of evidence produced using registry data for the evaluation of medical device safety and performance. *BMJ Surg Interv Health Technologies* 2025;**7**:e000364. doi:10.1136/bmjsit-2024-000364

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjsit-2024-000364).

Received 18 November 2024
Accepted 07 October 2025



For numbered affiliations see end of article.

**Correspondence to**
Dr Lotje A. Hoogervorst;
l.a.hoogervorst@lumc.nl

## ABSTRACT

**Objectives** Medical device registries in Europe report limited information about their structure and methodological characteristics. This hinders their utility for evaluation of medical device safety and performance under the Medical Device Regulation. This study aimed to define a minimum checklist of items necessary for regulators to assess the quality of evidence produced using registry data for the evaluation of medical device safety and performance.

**Design** A three-round Delphi panel.

**Setting** A task within the Coordinating Research and Evidence for Medical Devices project.

**Participants** 101 experts in the medical device community (healthcare professionals, methodologists, registry experts, regulators, and assessors from notified bodies) were invited.

**Interventions** Based on a literature review and expert advice, 27 items relating to the quality of registry data and the analysis of medical device safety and performance were selected. In round 1, participants selected which items were required for a minimum checklist. They could also propose new items. Items selected by ≥70% of participants indicated consensus. Remaining items were discussed in round 2, resulting in a final checklist that was ranked by participants for importance (round 3).

**Main outcome measures** Consensus of items to be included in the minimum checklist.

**Results** 51 experts participated in round 1, achieving consensus on 18 (67%) items and suggesting 12 items. After discussion in round 2, 5 additional items were selected, resulting in a final set of 15 data quality items and 8 data analysis items. The most important items were 'completeness of procedures' (data quality) and 'definition of outcome analyzed'" (quality of analysis).

**Conclusions** Reporting all items from the minimum checklist will facilitate judgment of the utility of registry data to evaluate medical devices during post-market surveillance.

### WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ European Medical device registries report limited information about their structure and methodological characteristics, which impedes their utility for evaluation of medical device safety and performance.

### WHAT THIS STUDY ADDS

⇒ Registries should publicly report on 15 data quality and 8 data analysis items, and specifically on the items 'completeness of procedures' and 'definition of outcome analyzed'" as these items were considered as the most important.

### HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Reporting all items from the minimum checklist will allow better judgment of the quality of evidence produced using registry data to evaluate medical devices during postmarket surveillance.

## INTRODUCTION

Postmarket surveillance is one of the crucial elements for assuring the safety and performance of medical devices in patients. The European (EU) Union Medical Device Regulation (MDR) requires manufacturers to plan and conduct postmarket surveillance of their medical devices (see Article 83 of (EU) 2017/745),[1 2] including the collection of real-world outcomes for patients receiving a specific medical device in clinical practice. For postmarket surveillance, different data sources can be used including medical device registries.[3] Manufacturers set up a postmarket surveillance system for their device(s) and notified bodies assess whether manufacturers plan and conduct it in a correct manner. Regulators also have responsibilities

to monitor the safety of medical devices placed on the market.

In the USA, the coordinated registry networks (CRNs) have been developed to produce all the necessary evidence for regulators and other stakeholders by combining data from multiple sources.[4 5] While there are examples of countries outside the USA where this concept can also be applied, such as the UK and Australia where registry data have been linked to, for example, hospital data, regulations in EU countries make this difficult. All EU databases, including medical device registries, need to comply with the General Data Protection Regulation (GDPR), implemented to protect individuals' data and privacy, which includes regulations regarding data sharing and privacy.[6 7] As a result, it is not always allowed to share registry data across (EU) countries or to link registry data to another data source within EU countries. In addition, the available registry data may be inadequate for decision-making due to significant heterogeneity across datasets in available registry-based studies or annual reports.[8] As a result, although tools and regulatory guidelines exist to assess the quality of registry data,[9 10] it is complex to judge the regulatory utility of these registry data.

Consensus among regulators internationally, with input from experts at the International Medical Device Regulators Forum (IMDRF), produced guidance documents on usability and methodological principles for using registry data.[11 12] Furthermore, the Food and Drug Administration (FDA) indicated relevance and reliability of data as key indicators when using real-world data.[13] However, these documents do not include more specific and detailed guidance on which items should be considered by regulators, notified bodies, and manufacturers when assessing the quality of evidence produced by using registry data. The previously developed maturity framework to assess the maturity of CRNs and registries[4] does include several items related to data quality—focusing on relevance, coverage, data completeness, and data verification—but not several other variables found in a previous systematic review[8] such as reporting how patient consent is managed and who can access and use the data, as well as items related to analysis of data regarding performance or safety of the device. Agreeing on a minimum checklist of items that medical device registries should publicly report would therefore assist manufacturers in their selection of data to be used for postmarket surveillance under the MDR, and it would allow EU regulators to determine whether the registry data may be reliable for the evaluation of medical device safety and performance during market surveillance.

As part of the Coordinating Research and Evidence for Medical Devices (CORE-MD) project (a summary of the deliverables and corresponding findings is available in an online report),[14–16] the aim of this study was to support the assessment of the quality of the evidence produced using registry data when evaluating medical device safety and performance during postmarket surveillance. This was achieved by reaching consensus on a minimum checklist of items that are essential to judge: (1) the quality of reporting of registry data and (2) the quality of methods reported to be used for analysis of medical device safety and performance.

## METHODS

### Patient and public involvement

No patients or members of the public were directly involved in the design of this study.

### Study design

A three-round Delphi method, consisting of two online surveys and one online consensus meeting (figure 1), was used to achieve consensus among EU experts in the evaluation of medical device safety and performance. The Delphi method is a validated method that can be used to transform individual opinions into group consensus.[17]

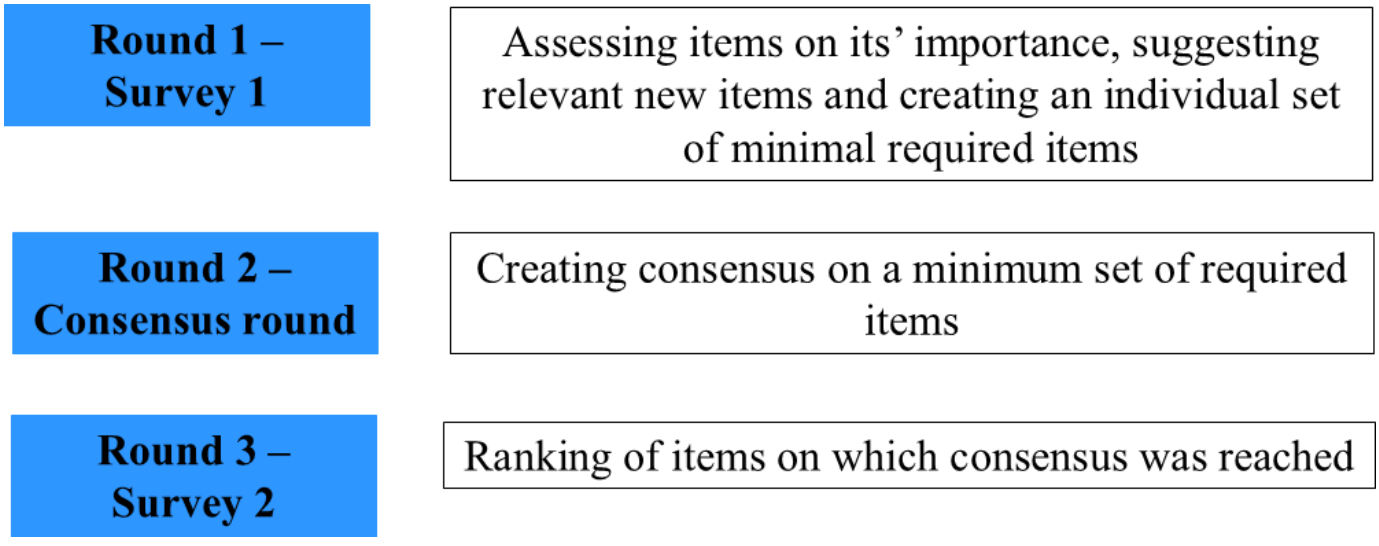


**Figure 1** Flow chart showing the consensus process.

**Table 1** Initial items evaluated, concerning the quality of registry data (seventeen items), and the quality of analysis of medical device safety and performance (10 items)

| Items concerning quality of registry data | Items concerning the quality of analysis of medical device safety and performance |
|---|---|
| (1) Goal of registry<br>(initial motivation/goal to set up the registry) | (1) Methods for handling missing data described<br>(eg, missing procedures will be sent every 3 months to each hospital department and request for data entry/missing data is considered as missing completely at random) |
| (2) Design<br>(eg, regional/national/multicountry) | (2) Time period during which devices were implanted |
| (3) Starting year<br>(year of first patient/procedure included) | (3) Minimum number of patients and/or procedures at risk |
| (4) Mandatory<br>(mandatory for surgeons/hospitals to submit data to the registry: yes/no) | (4) Minimum number of hospitals in which the device is used |
| (5) Patients' consent<br>(patients' consent required before entering their data into the registry: required/not required) | (5) Minimum number of surgeons using the device |
| (6) Funding<br>(eg, public/private/both) | (6) Minimum follow-up duration |
| (7) Data access<br>(who can access the data and see results? (eg, public access/only to members)) | (7) Statistical approach used to analyze performance<br>(eg, assessing superiority/non-inferiority in a relative benchmark/using an absolute benchmark defined by objective performance criteria) |
| (8) Privacy regulation for patients' identifiable information<br>(privacy regulation reported as implemented: yes/no? And if yes: how?) | (8) Adequate analysis to adjust for confounding (by indication)<br>(eg, propensity scores) |
| (9) Data capture and collection method<br>(eg, electronic/manual/barcodes-industry/reported by operator) | (9) Definition of outcome analyzed |
| (10) Method of access to registry for users/members<br>(eg, dashboard/real-time/secure server) | (10) Definition of outlier performance |
| (11) Level of information provided<br>(data is reported at hospital-/medical device-/surgeon-level) | |
| (12) Data linkage with other sources<br>(eg, registry data is linked to hospital statistics/manufacturer vigilance data/national competent authority on medical devices) | |
| (13) Quality assurance system defined/quality check of data described<br>(eg, data verification) | |
| (14) Reporting missing data for all patients' characteristics in registry (%)<br>(eg, Body Mass Index/American Society of Anesthesiologists classification/gender (%)) | |
| (15) Completeness of procedures<br>(number of procedures captured in registries relative to total number of procedures performed, as %) | |
| (16) Coverage (hospitals)<br>(number of participating hospitals relative to the total number of eligible hospitals, as %) | |
| (17) Collecting unique device identifier | |

During the entire Delphi process, equal weight was given to the opinions of each group of stakeholders.

In round 1, participants were asked to select items from an initial set of 27 items identified through literature review and expert advice.[8] Of the 27 items, 17 related to the quality of registry data, and ten concerned the quality of analysis of medical device safety and performance (table 1). The set of initial items was listed in an online survey and participants were asked to indicate using a 3-point Likert scale whether each item was: (1) not important, (2) somewhat important, or (3) very important. All items rated as 'somewhat important' and 'very important' were fed

into the second step, as the starting point for participants to create their own minimum checklist. For each item, participants were asked if the item was 'required' or 'not required' in the minimum checklist. In the third step of round 1, participants could suggest new items that they considered necessary. The first author (LAH) extracted all newly suggested items and harmonized similar items with different wording between participants.

As input for the online consensus meeting (round 2), LAH calculated for each item, the percentage of experts who had included it in their minimum checklist (all items ranked as 'required' as well as those newly suggested items); those selected by at least 70% of all participants were defined as indicating consensus.[18] By email, each participant then received a report detailing which items had reached consensus, together with their individual checklist (all items ranked as 'required' as well as their newly suggested items) and with information on how often the remaining items (ie, items not reaching consensus) appeared in the checklists across all participants. During the online consensus meeting, LAH first presented—for information purposes solely—the items on which consensus was reached.

All remaining items that did not receive consensus (ie, selected by <70% of all participants) but which were included at least once in an individual checklist as well as newly suggested items were then discussed. The discussion was chaired by PMvdM. After initial discussion on a specific item, a poll was created with the following question: 'Is this item needed in addition to those items already selected in the minimum dataset?' with two possible answers: (1) 'yes, it is required' and (2) 'no, it is not required'. As before, consensus was defined as ≥70% of participants voting for the item be included in the checklist.[18] If <70% of the participants considered that the item was required, the item was discussed until consensus was reached to either include or exclude the item from the checklist. Participants also had the option to rephrase items on which no consensus was reached, followed by a poll of the rephrased question. This resulted in a final minimum checklist across all participants.

In round 3 (survey 2), participants were asked to rank the items on which consensus had been achieved. Having an average rank for each item may subsequently guide regulators, notified bodies, manufacturers, and clinicians on how much weight they should place on an item, as in practice a registry may score poorly on one item but higher on another. A total of 100 points had to be allocated across all items related to the quality of registry data, and another total of 100 points across all items concerning the quality of analysis of medical device safety and performance. More points reflected greater importance. This method was used as it forces participants to choose between the items rather than merely rating all items as very important, since there is evidence that other rating scales (such as Visual Analog Scores) have limited capacity to differentiate between items.[19]

### Survey development

The two online surveys were developed by LAH using Sawtooth (Sun Valley, Idaho, USA) and survey links were distributed via email. Both surveys were first piloted by seven PhD students to ensure clear comprehensibility and reliability of the questions. The students provided comments that resulted in several (small) adjustments, and both adjusted surveys were tested again by the same group of PhD students.

### Expert panel recruitment

A total of 101 EU experts, divided into 4 groups of stakeholders, were invited to participate in our Delphi panel: (1) 30 regulators and notified body representatives, (2) 28 healthcare professionals particularly from the orthopedic and cardiovascular fields, as together they represent the majority of high-risk medical devices,[15] (3) 24 experts involved in (national) registries, and (4) 19 methodological experts (eg, on analysis of medical device safety and performance). These stakeholders were invited because they were involved in the evaluation of medical device safety and performance (ie, regulators, notified bodies, and clinicians in expert panels) or due to their knowledge and expertise regarding the quality of the evidence provided by using registry data. The aim was to include at least ten participants per stakeholder group to ensure sufficient sample size and distribution across groups. Experts had 2 weeks to complete each survey. If experts did not complete the survey within this timeframe, LAH sent a reminder to those who had not yet responded to give them another opportunity to complete the survey within 2 weeks. If they did not respond to the first survey after 4 weeks, they were considered non-respondents and excluded from further participation. If participants completed the first survey but did not participate in the consensus round (round 2), their input in the first survey was still used in the consensus round to calculate the percentage consensus. These participants were also invited to participate in round 3 (the second survey).

### Data analysis

Descriptive statistics were used to report the response rates in all 3 rounds; the response rate for round 1 was calculated as the percentage of participants filling in the first survey relative to all invited experts. Response rates for rounds 2 and 3 were calculated as the percentage of those participating in round 1. For each of the 27 items, the percentage of participants voting 'required' was calculated in round 1. For round 3 (survey 2), the total sum of points and the mean number of points assigned to each item were calculated. For each item, we calculated their relative weight (ie, importance) by dividing the mean number of points assigned to that item by the number of expected points if all items had equal weight (ie, 100 divided by the total number of items to be ranked).

For each participant filling in the online surveys, the time taken to complete the survey was extracted.

Consequently, the median time to complete the online surveys was calculated, together with the corresponding IQR.

Analyses were performed using Microsoft Excel (Redmond, USA).

## RESULTS

Of the 101 experts invited for the Delphi panel, 51 experts (50%) from 14 countries completed round 1 (survey 1). Of these 51 experts, 30 (59%) participated in the consensus meeting (round 2). And 38 of the 51 experts (75%) completed round 3 (survey) (online supplemental table 1). The median time to complete the first survey was 8 min (IQR: 6–19 min) and for the second survey 7 min (IQR: 5–11 min).

### Round 1: selecting an individual minimum checklist

Consensus was achieved on 10 of the 17 (59%) data quality items and 8 of the 10 (80%) items concerning the quality of analysis of medical device safety and performance (online supplemental figure 1A). The three data quality items most frequently selected in individual minimum checklists were: (1) the completeness of procedures (96%); (2) the level of information provided (ie,hospital, medical device, or surgeon level) (92%), and (3) the quality assurance system defined/quality check of data (90%). For items concerning the quality of analysis of medical device safety and performance, the top three were: (1) the definition of outcome analyzed (98%); (2) the time period during which devices were implanted (94%), and (3) the approach to analyze performance (eg, assessing superiority/non-inferiority in a relative benchmark or using an absolute benchmark) (92%) (online supplemental figure 1B). A total of 11 new data quality items and 1 quality of analysis item were suggested (online supplemental table 2).

### Round 2: creating consensus on a minimum checklist

During the online consensus meeting, the remaining 7 data quality items were discussed (online supplemental figure 1A). During the discussion, 2 items (items number 7 and 10 from table 1) were combined into 1 item 'reporting on procedures how to apply for data, who can access and use the data' which resulted in consensus (100% of participants voted for inclusion, online supplemental figure 2A). In addition, item number 5 from table 1 on patients' consent was rephrased for better interpretation into 'reporting how patient consent is managed and for which purposes', which then resulted in consensus (86% of participants voted for inclusion in the minimum checklist, online supplemental figure 2A).

Of the 11 newly suggested data quality items, only 3 items were discussed because none of the participants felt that any of the other 8 items added sufficiently to the minimum checklist. The 3 items that were discussed were: (1) 'clearly defined patient inclusion/exclusion criteria'; (2) 'important confounders/risk factors/exposures, with potential impact on outcome have been identified and recorded', and (3) 'reporting how validation of the standard is achieved'. Only the first item on patient selection reached consensus (76% of participants voted for inclusion, online supplemental figure 2B). In total, participants voted on 9 data-quality items, of which 5 items were included in the minimum checklist (online supplemental figure 2B).

For items concerning the quality of analysis of medical device safety and performance, 2 remaining items (online supplemental figure 1B) and 1 newly suggested item were discussed, but none of these was included in the minimum checklist (online supplemental figure 2B).

Combining the findings of Delphi rounds 1 and 2, table 2 shows the minimum checklist on which consensus was achieved, which includes fifteen items concerning quality of registry data and eight items concerning the analysis.

### Round 3: ranking items included in the minimum checklist

Given that 15 data quality items were selected, the number of expected points assigned if all items were equally important was 6.67. Of all data quality items, the item 'completeness of procedures' was deemed most important for reporting, with a total sum of 421 points assigned across participants (mean per participant 11.1 with SD=10.3), resulting in a relative weight of 1.66 (online supplemental figure 3A). The item 'reporting missing data for all patients' characteristics in registry (%)' was the second most important, with a total of 334 points (mean per participant 8.8 (SD=4.4) relative weight 1.32). The item with the lowest number of points assigned was: 'privacy regulation for patients' identifiable information' with 146 (mean per participant 3.8 (SD=3,0) relative weight 0.58). When analyzing the outcomes of each specific stakeholder group, the item 'completeness of procedures' was considered most important for reporting by both healthcare professionals and experts involved in (national) registries. In contrast, 'quality assurance system defined/quality check of data' and 'reporting missing data for all patients' characteristics in registry' were deemed most important by regulators and notified body representatives. Methodological experts gave the highest priority to 'coverage (hospitals)'.

As 8 data analysis items were selected, the number of expected points assigned to each item, if all items were considered equally important, was 12.5. Most points were assigned to 'definition of outcome analyzed'" with a total of 580 (mean per participant 15.3 points (SD=6,1) and relative weight 1.23) followed by 'minimum number of patients and/or procedures at risk' (534 points; mean per participant 14.1 (SD=7,2) and relative weight 1.13) (online supplemental figure 3B). The lowest number of points was assigned to the item 'definition of outlier performance' with 420 (mean per participant 11.1 points (SD=5.3) with a relative weight of 0.88). When analyzing the outcomes of each specific stakeholder group, the item 'definition of outcome analyzed'" was considered most

Table 2 Items included in the minimum required checklist

| Items concerning quality of registry data | Items concerning the quality of analysis of medical device safety and performance |
|---|---|
| (1) Design<br>(eg, regional/national/multicountry) | (1) Methods for handling missing data described<br>(eg, missing procedures will be sent every 3 months to each hospital department and request for data entry/missing data is considered as missing completely at random) |
| (2) Mandatory<br>(mandatory for surgeons/hospitals to submit data to the registry: yes/no) | (2) Time period during which devices were implanted |
| (3) Reporting how patient consent is managed and for which purposes | (3) Minimum number of patients and/or procedures at risk |
| (4) Funding<br>(eg, public/private/both) | (4) Minimum follow-up duration |
| (5) Reporting on procedures how to apply for data, who can access and use the data | (5) Approach to analyze performance<br>(eg, assessing superiority/non-inferiority in a relative benchmark/using an absolute benchmark defined by objective performance criteria) |
| (6) Privacy regulation for patients' identifiable information<br>(privacy regulation reported as implemented: yes/no? And if yes: how?) | (6) Adequate analysis to adjust for confounding (by indication)<br>(eg, propensity scores) |
| (7) Data capture and collection method<br>(eg, electronic/manual/barcodes-industry/reported by operator) | (7) Definition of outcome analyzed |
| (8) Level of information provided<br>(data is reported at hospital-/medical device-/surgeon-level) | (8) Definition of outlier performance |
| (9) Data linkage with other sources<br>(eg, registry data is linked to hospital statistics/manufacturer vigilance data/ national competent authority on medical devices) | |
| (10) Quality assurance system defined/quality check of data described<br>(eg, data verification) | |
| (11) Reporting missing data for all patients' characteristics in registry (%)<br>(eg, Body Mass Index/American Society of Anesthesiologists classification/ gender (%)) | |
| (12) Completeness of procedures<br>(number of procedures captured in registries relative to total number of procedures performed, as %) | |
| (13) Coverage (hospitals)<br>(number of participating hospitals relative to the total number of eligible hospitals, as %) | |
| (14) Collecting unique device identifier | |
| (15) Reporting on patient inclusion/exclusion criteria<br>(ie, patient selection) | |

important for reporting by both healthcare professionals and experts involved in (national) registries. Contrarily, 'minimum number of patients and/or procedures at risk' was deemed most important by regulators and notified body representatives. Methodological experts gave the highest priority to 'approach to analyze performance'.

## DISCUSSION

This Delphi study, using a large panel of EU experts involved in the evaluation of medical devices, reached consensus on a minimum checklist of 15 items concerning quality of registry data and 8 items concerning the quality of analysis of medical device safety and performance. Of all items included in the checklist, 'completeness of procedures' and 'definition of outcome analyzed' were deemed most important for data quality and quality of analysis, respectively. Public reporting by registries of this minimum checklist of 23 items will facilitate regulators, notified bodies, and manufacturers in judging the utility of registry data when evaluating medical devices during postmarket surveillance.

This is the first study to create a minimum required checklist consisting of items on structural and methodological characteristics of EU medical device registries that are important to judge the quality of data as well as analysis of these data. Previous EU initiatives have focused on achieving common definitions and outcomes across registries, to increase uniformity of data collected.[9 20–23] The IMDRF has produced guidance documents on assessing the usability of registry data and addressed methodological principles for performing clinical evaluation and signal detection using registry data[11 12] and other reports emphasised the importance of data completeness and accuracy,[24–26] to which our minimum checklist adds more items that are relevant. Compared with the FDA guidance[13] and the previously mentioned maturity assessment of registries,[5] several items are similar, such as common data capture, data verification procedures, and data completeness. Our minimum checklist includes additional items such as reporting on how patient consent is managed and items related to data analysis. Achieving consensus on items needed to judge the quality of evidence produced using registry data, for the evaluation of medical device safety and performance, is an important first step. Our minimum checklist, however, does not define what qualifies as sufficient quality data, especially when high scores on certain items are paired with lower scores on other items of the minimum checklist. The ranking provided in the current study may guide regulators, notified bodies, and manufacturers on which quality items are recommended to be assigned more weight.

As mentioned previously, due to the GDPR, it is not always permitted to share registry data across (EU) countries or to link registry data to other data sources within EU countries. However, within countries, there are some examples where data linkage is possible. For instance, the German Arthroplasty Register links data with health insurers, and the Irish National Orthopaedic Register links with the National database on discharges from acute public hospitals. Data linkage with other sources is often used for data verification purposes, such as assessing the completeness of procedures compared with electronic health records. However, when data linkage is not possible, verification can also be performed by comparing (aggregated) numbers. Therefore, although data linkage offers many advantages, verification can also be achieved through alternative methods. Hence, both were included as separate items. This may also explain why the item 'quality assurance system defined', which includes data verification procedures, was given higher priority in the ranking than 'data linkage with other sources'.

### Decision framework to assess the safety and performance of medical devices

The National Institute for Health and Care Excellence (NICE) framework in the UK is not exclusively designed for regulatory decision-making nor does it solely concentrate on medical devices.[24] Instead, it encompasses a broader spectrum of real-world data sources, including medical device registries, to support those developing evidence to inform NICE guidance. The framework highlights that real-world data should be 'of good provenance, relevant and of sufficient quality to answer the research question', and that evidence should be generated in a transparent way while using 'analytical methods that minimize risk of bias and characterize uncertainty'. Under data provenance, they consider knowledge about the purpose and methods of data collection to be important, as well as data coverage and governance. Relevance focuses on generalizable and robust results, where completeness and accuracy are key factors considered for data quality.

The aforementioned FDA guidance document states that the 2 key factors for assessing real-world data are 'relevance' and 'reliability'.[13] Under the key factor 'relevance', it is listed that: (1) 'real-world data should contain sufficient detail to capture the use of medical devices, exposure, and the outcomes of interest in an appropriate population'; (2) 'the use of a specific medical device in a real-world population should be representative as captured within the data source, and is generalizable to the relevant population being evaluated', and (3) 'available data elements should be able to address the question at hand when valid and appropriate methods are used. 'Reliability' covers various aspects of data collection (eg, common definitions and a relevant time window) but also data quality such as adherence to verification procedures. The previously mentioned framework to assess the maturity of CRNs and registries incorporates these principles, for example, in their data quality domain.[4 5] 3 of the 7 domains show overlap with the items included in our minimal checklist: 'device identification', 'data quality', and 'governance and sustainability'. However, the domain descriptions are generally broad rather than indicating which specific items are considered within each domain. On the other hand, they indicate a description of maturity levels showing how a registry may advance from, for example, a pilot registry including several sites to a national registry with greater than 80% coverage and greater than 80% data completeness.

Where previous frameworks give rather general descriptions with some examples, there may be other factors to consider and contextual factors may determine the acceptability of the evidence (eg, high-quality evidence may be more challenging to generate for rare diseases and devices). In addition, these frameworks do not specify a minimum checklist of items within each domain to allow regulators and manufacturers to assess the safety and performance of medical devices. We, therefore, mapped the items on which consensus was achieved in the current Delphi study to the more generic principles and domains found in previous frameworks. This resulted in a decision framework that may assist EU regulators when assessing the safety and performance of medical devices for market surveillance as well as manufacturers when using registry data for postmarket surveillance (figure 2).

The framework uses relevance and reliability as the guiding principles, consistent with previous FDA guidance.

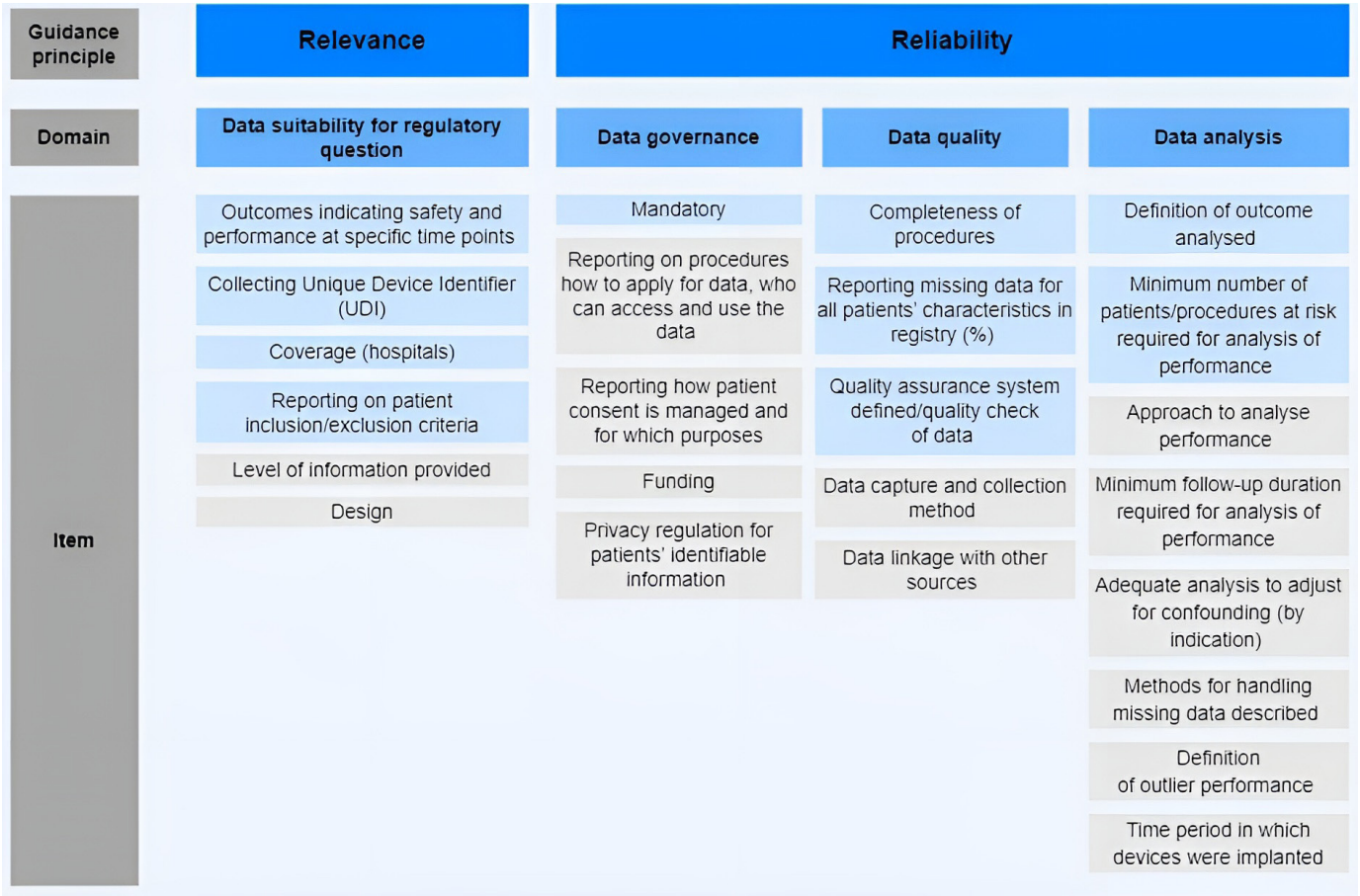


**Figure 2** Decision framework to assess safety and performance of medical devices (the items listed in light gray scoring lower than expected and the items listed in light blue higher than expected, based on their relative weight).

Within these principles, we distinguished 4 domains: data suitability for regulatory question (6 items), data governance (5 items), data quality (5 items), and data analysis (8 items). The outcome of interest at specific time points was added because of the large heterogeneity found in our previous systematic review between outcomes and time points captured by registries, and because of the lack of clarity regarding which of these outcomes could be included to calculate the benefit-risk ratio for the intended purpose of a particular medical device, that is,suitability for regulatory question.[8] If all these factors are explored and found to indicate good-quality data and analysis, particularly for the items deemed most important (indicated in blue), then the evidence produced by the registry can be considered trustworthy.

### Strengths and study limitations

Our study comprised a large representation of EU experts involved in the evaluation of medical devices and the management of national registries. It included good representation across multiple groups of stakeholders. Our results are, therefore, likely to reflect the opinion of other EU experts in the field of regulatory evaluation of medical devices. Nonetheless, some study limitations should be noted. First, we only included experts proposed from the professional network of the CORE-MD research group, which consisted solely of EU experts. Hence, the recommendations drawn from our study may not be generalizable to non-EU countries. A broader inclusion of non-EU experts may increase the external validity of the minimum checklist. On the other hand, we also showed overlap with the previously published maturity framework which was developed in the US, suggesting generalizability.

Second, there might be selection bias as only 51% of the invited experts participated in round 1, with fewer participants in the last 2 rounds. These response rates are lower than the Delphi panel guidelines.[18] We believe that the response rates did not relate to the length of the surveys, as they were relatively short (median times to complete the surveys were less than 8 min). Despite the relatively low response rates, our Delphi panel is still in line with sample size recommendation for a Delphi panel, namely: as small as three members or as large as eighty, whereby a sample of approximately fifteen participants is recommended.[18 27 28] Importantly, there was a balanced participation by all stakeholder groups in all rounds.

Third, no manufacturers were invited to participate in our Delphi, as they were not included in the CORE-MD network, to avoid any commercial influence. Moreover, manufacturers are not involved in the evaluation of

medical device safety and performance, but rather supply the data to be evaluated, which could pose conflicts if they would prioritize certain items based, for example, on the ease of collecting data rather than on their utility as a source of good-quality data.

Lastly, the time to respond in the Delphi rounds 1 and 3 (surveys 1 and 2, respectively) was limited, namely 4 weeks. However, as three-quarters (39 out of 52) of the respondents in round 1 (survey 1) also completed the second survey, the effect of this time limit seems to be negligible.

### Perspective and future research

The items listed in our proposed checklist are relatively easy to report publicly, as most EU medical device registries will include these items already. The practical implementation of the minimum required checklist has not been tested, so both its usefulness and effectiveness are currently unknown, indicating that further research is needed to evaluate the experience with the proposed minimum checklist. A first step toward implementation is the recently developed International Society of Arthroplasty Registries template, including items covering general descriptive information about registries, information related to governance, outcomes, data quality, data access, and registry production.[29] Further research can determine the thresholds to be used to indicate sufficient quality evidence for each item as well as for combinations, given that registries could score 'sufficient' on one item, but 'insufficient' on another, and to test these thresholds using empirical data. For instance, a registry may have good-quality data, but due to poor quality of analysis, may still be considered to produce lower-quality evidence. In addition, future research might focus on further specifying criteria for data access and governance, as these remained relatively broad in the current framework and lacked detail on what would be considered good quality.

We hope that the proposed checklist for the minimum number of items required to judge the quality of evidence produced by registries, for the evaluation of safety and performance of medical devices, will be implemented by studies reporting on registry data as well as by registries in their annual reports. That may not only benefit regulators, notified bodies, and clinicians, but also help to improve the comparability of data and interoperability between registries. Researchers can also refer to the checklist when they use registry data for scientific analyses, for example, reporting on specific items such as the completeness of reporting of procedures, which would indicate the quality of their data. Combining data from medical device registries—either through linkage of data or in a federated network analysis—is crucial to detect any safety and performance concerns related to medical devices as early as possible, in order to minimize harm to patients. That will be achieved only if the evidence produced using registry data is shown to be of sufficient quality.

## CONCLUSIONS

Registries reporting publicly on the proposed 15 items regarding the quality of registry data and the 8 items concerning the quality of analysis will allow regulators, notified bodies, manufacturers, and also clinicians to judge the evidence produced using registry data for the evaluation of medical devices during postmarket surveillance.

**Author affiliations**
[1]Department of Orthopaedics, Leiden Universitair Medisch Centrum, Leiden, The Netherlands
[2]Trinity College Dublin School of Medicine, Dublin, Ireland
[3]Directorate General for Health (DG SANTE), European Commission, Brussels, Belgium
[4]Academic Cardiovascular Unit, James Cook University Hospital, Middlesbrough, UK
[5]Hull York Medical School, University of York, York, UK
[6]Division of Orthopaedic Surgery and Traumatology, Geneva University Hospitals, Geneva, Switzerland
[7]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[8]Department of Cardiology, Leeds Teaching Hospitals NHS Trust, Leeds, UK
[9]Leeds Institute for Data Analytics, University of Leeds, Leeds, UK
[10]Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, UK
[11]Department of Applied Economics, University of Granada, Granada, Spain
[12]Department of Orthopaedic Surgery and Traumatology, Copenhagen University Hospital, Copenhagen, Denmark
[13]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[14]School of Medicine, University of St Andrews, St Andrews, UK
[15]Department of Orthopaedic Surgery, Victoria Hospital Kirkcaldy, Kircaldy, UK
[16]Centre for Postgraduate Medical Education, Warsaw, Poland
[17]Clinical Cardiology Center, National Medicines Institute, Warsaw, Poland
[18]Department of Orthopaedics, University of Gothenburg Institute of Clinical Sciences, Gothenburg, Sweden
[19]Department of Orthopaedics, Sahlgrenska University Hospital, Gothenburg, Sweden
[20]Department of Orthopaedics, Hässleholms Hosptial, Skane, Sweden
[21]Health Products Regulatory Authority, Dublin, Ireland
[22]Department of Surgery, Trinity College Dublin School of Medicine, Dublin, Ireland
[23]Centre for Health Protection, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[24]Department of Cardiology, University Hospital of Wales, Cardiff, UK
[25]Safety & Security Science and Centre for Safety in Healthcare, Delft University of Technology, Delft, The Netherlands

**Collaborators** CORE-MD study group: The CORE-MD study group for this task consists of all participants in the Delphi study: Alan Fraser, Alma Becic Pedersen, Anne Lubbeke, Bart Pijls, Chris Gale, Chris Wilkinson, Christina Dimopoulou, Christoph Ziskoven, Claudia Wild, David Epstein, David Erlinge, Donal O'Connor, Elin Karlberg, Emil Hagstrom, Enrico Caiani, Enrique Gomez Barrena, Erman Melikyan, Filippo Boniforti, Fintan Bergin, Frank Rademakers, Gavin Quigley, Gearoid McGauran, Gearoid O'Connor, Heather Prentice, Jantine van Baal, Juan Carlos Rejón Parrilla, Leo Hovestadt, Liza van Steenbergen, Marina Torre, Maziar Mohaddes, Michael Dunbar, Nunung Nurrahmah, Orjan Friberg, Paola Laricchiuta, Paul Piscoi, Per Kjærsgaard-Andersen, Peter Vasko, Philip Walmsley, Piotr Szymanski, Ria Mahon, Richard Holborow, Rob Nelissen, Robert Geertsma, Roelf Postema, Sergio Buccheri, Soren Overgaard, Stefan James, Stephan Windecker, Stephen Dean, Timothy Wilton, and Tom Melvin.

**Contributors** LAH (guarantor; conceptualization, data curation, formal analysis, methodology, project administration, writing–original draft). RN (conceptualization, funding acquisition, writing–review, and editing). TM (writing–review and editing). PP (writing–review and editing). CW (writing–review and editing). AL (writing–review and editing). CPG (writing–review and editing). DE (writing–review and editing). SO (writing–review and editing). PW (writing–review and editing). PS (writing–review and editing). MM (writing–review and editing). DBO'C (writing–review and editing). REG (writing–review and editing). JMH (writing–review

and editing). AF (funding acquisition, writing–review and editing). PJM-vdM (conceptualization, funding acquisition, data curation, methodology, project administration, writing–review and editing).

**Funding** This work was supported by the European Union Horizon 2020 Research and Innovation Program (grant number 965246) and was part of the Coordinating Research and Evidence for Medical Devices (CORE-MD) project.

**Disclaimer** CORE-MD is supported by a grant from the European Commission. The sole responsibility for the content of this manuscript lies with the authors. It is not intended to convey the views of their employers, their affiliated institutions or the European Commission.

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Ethics approval** This study was approved by the local ethics committee of the Leiden University Medical Center (W.23.009). Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information. Not applicable.



**ORCID iDs**
Lotje A. Hoogervorst https://orcid.org/0000-0001-5343-6965
Rob G.H.H. Nelissen https://orcid.org/0000-0003-1228-4162
Robert E. Geertsma https://orcid.org/0000-0003-0979-9077
Alan G. Fraser https://orcid.org/0000-0001-7083-6995

## REFERENCES

1 Medical Device Coordination Group (MDCG). Regulation (EU) 2017/745: clinical evidence needed for medical devices previously CE marked under directives 93/42/EEC or 90/385/EEC. A guide for manufacturers and notified bodies. Available: https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2020_6_guidance_sufficient_clinical_evidence_en_0.pdf [Accessed 26 Oct 2024].
2 Fleetcroft C, McCulloch P, Campbell B. IDEAL as a guide to designing clinical device studies consistent with the new European Medical Device Regulation. *BMJ Surg Interv Health Technol* 2021;3:e000066.
3 Massimo P. ANNEX III Technical documentation on post-market surveillance. *BMJ Surg Interv Health Technol* 2024;4:e000123.
4 Sedrakyan A, Marinac-Dabic D, Campbell B, *et al*. Advancing the Real-World Evidence for Medical Devices through Coordinated Registry Networks. *BMJ Surg Interv Health Technol* 2022;4:e000123.
5 Sedrakyan A, Aryal S. Maturity framework and select approaches for developing Coordinated Registry Networks (CRNs): Medical Device Epidemiology Network (MDEpiNet) supplement. *BMJ Surg Interv Health Technologies* 2022;4:e000148.
6 European Union. Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation). Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj [Accessed 03 Feb 2025].
7 European Data Protection Board. Guidelines 02/2024 on Article 48 GDPR. Available: https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2024/guidelines-022024-article-48-gdpr_en [Accessed 03 Feb 2025].
8 Hoogervorst LA, Geurkink TH, Lübbeke A, *et al*. Quality and Utility of European Cardiovascular and Orthopaedic Registries for the Regulatory Evaluation of Medical Device Safety and Performance Across the Implant Lifecycle: A Systematic Review. *Int J Health Policy Manag* 2023;12:7648.
9 Guilhaume C. A tool to assess the registries quality: The Registry Evaluation and Quality Standards Tool (REQueST). *Eur J Public Health* 2021;31.
10 European Medicines Agency. Guideline on registry-based studies. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf-0 [Accessed 26 Oct 2024].
11 International Medical Device Regulators Forum. Tools for assessing the usability of registries in support of regulatory decision-making. Available: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-180327-usability-tools-n46.pdf [Accessed 26 Oct 2024].
12 International Medical Device Regulators Forum. Methodological principles in the use of international medical device registry data. Available: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170316-methodological-principles.pdf [Accessed 26 Oct 2024].
13 Food and Drug Administration (FDA). Use of real-world evidence to support regulatory decision-making for medical devices. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices [Accessed 26 Oct 2024].
14 Marang-van de Mheen P. Improving the quality of post-market surveillance. Available: https://www.core-md.eu/wp-content/uploads/2024/06/7-Perla-Marang-van-de-Mheen.pdf [Accessed 03 Feb 2025].
15 Fraser AG, Nelissen RGHH, Kjærsgaard-Andersen P, *et al*. Improved clinical investigation and evaluation of high-risk medical devices: the rationale and objectives of CORE-MD (Coordinating Research and Evidence for Medical Devices). *Eur Heart J Qual Care Clin Outcomes* 2022;8:249–58.
16 CORE-MD. Decision framework to assess the performance of high-risk medical devices. Available: https://www.core-md.eu/wp-content/uploads/2024/11/CORE-MD_D3.1_LUMC_v1.1_final_clean.pdf [Accessed 30 Jul 2025].
17 Vogel C, Zwolinsky S, Griffiths C, *et al*. A Delphi study to build consensus on the definition and use of big data in obesity research. *Int J Obes (Lond)* 2019;43:2573–86.
18 Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;32:1008–15.
19 De Meyer D, Kottner J, Beele H, *et al*. Delphi procedure in core outcome set development: rating scale and consensus criteria determined outcome selection. *J Clin Epidemiol* 2019;111:23–31.
20 Hoeijmakers F, Beck N, Wouters MWJM, *et al*. National quality registries: how to improve the quality of data? *J Thorac Dis* 2018;10:S3490–9.
21 European Society of Cardiology. BigData@Heart. Available: https://www.escardio.org/Research/Big-Data-Heart [Accessed 26 Oct 2024].
22 European Federation of National Associations of Orthopaedic and Traumatology (NORE). Minimal datasets. Available: http://nore.efort.org/minimal-datasets [Accessed 26 Oct 2024].
23 European Network of Cancer Registries (ENCR). ENCR recommendations 2022 updated recommendations for a standard. Available: https://policycommons.net/artifacts/3358432/encr-recommendations-2022-updated-recommendations-for-a-standard/4157085/ [Accessed 26 Oct 2024].
24 National Institute for Health and Care Excellence (NICE). NICE real-world evidence framework. Available: https://www.nice.org.uk/corporate/ecd9/chapter/overview [Accessed 26 Oct 2024].
25 International Society of Arthroplasty Registries (ISAR). International prosthesis benchmarking working group guidance document hip and knee arthroplasty devices. Available: https://drive.google.com/file/d/0BwKvdROo5Eg-MjZYc2VHQUZGYzNJMlRaenZEVUN3cTdMYlBj/view?resourcekey=0-OtTMX1RmF7E-HAZgxXiNZg [Accessed 26 Oct 2024].
26 Tcheng JE, Fleurence R, Sedrakyan A. Electronic health data quality maturity model for medical device evaluations. *BMJ Surg Interv Health Technol* 2020;2:e000043.
27 McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm* 2016;38:655–62.
28 Mullen PM. Delphi: myths and reality. *J Health Organ Manag* 2003;17:37–52.
29 Lübbeke A, Hoogervorst LA, Marang-van de Mheen PJ, *et al*. Arthroplasty registries at a glance: an initiative of the International Society of Arthroplasty Registries (ISAR) to facilitate access, understanding, and reporting of registry data from an international perspective. *Acta Orthop* 2025;96:116–26.