



Universiteit
Leiden

The Netherlands

Safety and performance of high-risk medical devices: the role of real-world data

Hoogervorst, L.A.

Citation

Hoogervorst, L. A. (2025, November 18). *Safety and performance of high-risk medical devices: the role of real-world data*. Retrieved from <https://hdl.handle.net/1887/4283146>

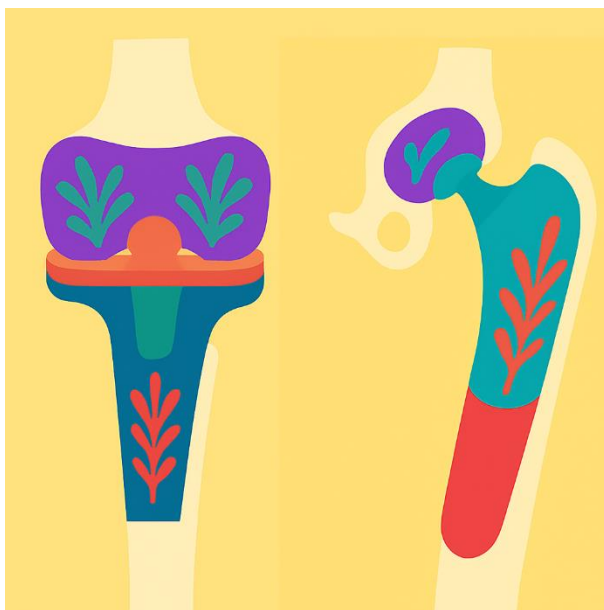
Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4283146>

Note: To cite this publication please use the final published version (if applicable).

**Chapter 6 – Validating Orthopaedic Data
Evaluation Panel (ODEP) ratings across 9
orthopaedic registries - total hip implants with an
ODEP rating perform better than those without
an ODEP rating**



Lotje A. Hoogervorst^{1,2}, Maartje M. van Tilburg¹, Anne Lübbecke^{3,4,5}, Tim Wilton^{6,7,8}, Rob G.H.H. Nelissen¹, Perla J. Marang-van de Mheen⁹

¹ Department of Orthopaedics, Leiden University Medical Center, Leiden, the Netherlands

² Department of Biomedical Data Sciences & Medical Decision Making, Leiden University Medical

³ Division of Orthopaedic Surgery and Traumatology, Geneva University Hospitals, Geneva Switzerland

⁴ University of Geneva, Geneva, Switzerland

⁵ Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

⁶ Department of Orthopaedics, Derby Teaching Hospitals NHS Foundation Trust, Derby, United Kingdom

⁷ Orthopaedic Data Evaluation Panel, United Kingdom

⁸ Beyond Compliance Steering Committee, Halesowen, United Kingdom

⁹ Safety & Security Science, Centre for Safety in Healthcare, Delft University of Technology, Delft, The Netherlands

J Bone Joint Surg Am. 2024 Sep 4;106(17):1583-1593

Abstract

Background: Orthopaedic Data Evaluation Panel (ODEP) ratings of total hip (TH) and total knee (TK) implants are informative for assessing implant performance. However, the validity of ODEP ratings across multiple registries is unknown. Therefore, we aimed to assess, across multiple registries, whether TH and TK implants with a higher ODEP rating (i.e. an A* rating) have lower cumulative revision risks (CRRs) than those with a lower ODEP rating (i.e. an A rating) and the extent to which A* and A-rated implants would be A*-rated on the basis of the pooled registries' CRR.

Methods: Implant-specific CRRs at 3-, 5-, and 10-years that were reported by registries were matched to ODEP ratings on the basis of the implant name. A meta-analysis with random-effects models was utilised for pooling the CRRs. ODEP benchmark criteria were utilised to classify these pooled CRRs.

Results: A total of 313 TH cups (54%), 356 TH stems (58%), 218 TH cup-stem combinations (34%), and 68 TK implants (13%) with unique brand names reported by registries were matched to an ODEP rating. Given the low percentage that matched, TK implants were not further analysed. ODEP-matched TH implants had lower CRRs than TH implants without an ODEP rating at all follow-up time points, although the difference for TH stems was not significant at 5-years. No overall differences in CRRs were found between A* and A-rated TH implants, with the exception of TH cup-stem combinations, which demonstrated a significantly lower CRR for A*A*-rated cup-stem combinations at the 3-year time point. 39% of A*-rated cups and 42% of A*-rated stems would receive an A* rating on the basis of the pooled registries' CRR at 3-years; however, 24% of A-rated cups and 31% of A-rated stems would also receive an A* rating, with similar findings demonstrated at longer follow-up.

Conclusion: At all follow-up time points, ODEP-matched TH implants had lower CRRs than TH implants without an ODEP rating. Given that

the performance of TH implants varied across countries, registries should first validate ODEP ratings with use of country-specific revision data to better guide implant selection in their country. Data source transparency and the use of revision data from multiple registries would strengthen the ODEP benchmarks.

Background

In the United States, medical devices are regulated by the Food and Drug Administration.¹ In the European Union, medical devices are regulated according to the Medical Device Regulation (MDR), which aims to provide “a robust, transparent, predictable and sustainable regulatory framework for medical devices which ensures a high level of safety and health whilst supporting innovation.”^{2,3} To ensure patient safety, the MDR requires manufacturers to monitor the performance of their implants, including total hip (TH) and total knee (TK) implants, with use of benchmarking – that is, “a systematic process of determining whether an implant meets specified performance levels.”^{4,5} Several methods for benchmarking TH and TK implants are utilised. These methods include comparing the performance of an implant to that of the best-performing implant, comparing it to the average performance of comparable implants, or comparing it to absolute thresholds determined by objective performance criteria (OPC).⁶⁻¹⁵

An example of the use of OPC to promote the evidence based selection of implants is the Orthopaedic Data Evaluation Panel (ODEP) rating, which is assigned to implants that show evidence of meeting survivorship criteria.¹⁰ ODEP ratings are available for TH components (cups and stems), TK implants (tibiofemoral combinations), unicondylar knee implants, shoulder components (glenoids and stems), reverse shoulder implants, total elbow implants, and spine implants (cervical discs). Implants are benchmarked by ODEP on the basis of revision data from observational studies (e.g. single-centre studies, manufacturers’ in-house sources, and registry data). Thus, not all ODEP ratings are based on registry data. The submitted data are supplied by manufacturers with use of standardised ODEP submission forms.¹⁶ Not all implants on the market are submitted to ODEP since data submission is voluntary, but surgeons and hospitals are encouraged to use ODEP-rated implants. As different data sources can be utilised by manufacturers to submit their application for an ODEP rating, the

data may not be representative of daily clinical practice. Therefore, before submission, manufacturers have to declare that the submitted clinical data are “representative of the results of all studies conducted in relation to it.”¹⁷ The ODEP rating includes a number (representing the years of evidence) and a letter (representing the strength of the evidence). The latter denotes the performance of implants at specific time points (i.e. 3-, 5-, 7-, 10-, 13-, and 15-years) based on the OPC, which include the minimum number of centres and surgeons, size of the cohort, number of patients at risk, and the maximum revision rate. Implants can be rated as A* (highest rating), A (lower rating), or B (a rating that is assigned either to implants that are extremely important but have limited usage or to new implants that are introduced in a limited manner), starting from 3-years of evidence. Implants that do not meet the ODEP benchmark criteria (Table 1) are not rated. Although originally developed for use in the United Kingdom (U.K.), the ODEP rating is increasingly utilised internationally for the quality assessment of implants.¹⁸⁻²⁰ In the Dutch Arthroplasty Register, 100% of all TH cups and TH stems and 92% of all TK implants utilised in 2019 were assigned an ODEP rating. In the U.K., comparable numbers were reported in 2018.^{19,21} Although ODEP ratings are increasingly utilised, to our knowledge, an external validation of ODEP ratings across multiple registries has never been undertaken.

Therefore, we aimed to assess, across multiple registries, whether TH and TK implants with a higher ODEP rating (i.e. an A* rating) have lower cumulative revision risks (CRRs) than those with a lower ODEP rating (i.e. an A rating) and the extent to which A* and A-rated implants would receive the A* rating on the basis of the pooled registries’ CRR. Since the maximum revision rate for an A* rating is lower than that for an A rating, we hypothesised that A*-rated implants would have lower CRRs across the registries than A-rated implants. Furthermore, we expected that the majority – but not all – of the A*-rated implants would also be A*-rated on the basis of the pooled registries’ CRR, as revision risks are also influenced by

variables such as surgeon factors that potentially affect implant performance.

Methods

The ODEP rating

The data submitted to ODEP is evaluated by a voluntary, independent panel of orthopaedic experts. To prevent camouflage (i.e. when the performance of a specific implant-design variant is concealed as a result of different variants existing under the same implant name)²², the panel reviews implants at the product-code level (Table 1¹⁰).²² After being assigned an ODEP rating, manufacturers have to resubmit new evidence at every ODEP milestone to prevent their implant ratings from being lapsed, which some manufacturers may not do.¹⁰ ODEP usually provides a grace period of 1-year before lapsing an ODEP rating. Implants that do not meet the benchmark criteria do not receive an ODEP rating.

Matching registry data to ODEP ratings

European registries were identified in a previous systematic review and were supplemented with non-European registries listed by the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR).^{23,24} Registries were included if they reported implant-specific CRRs with standard errors (SEs) and/or 95% confidence intervals (CIs) to allow the pooling of data and if they were “active” (i.e. “published at least one annual report and/or peer-reviewed paper containing registries’ data, during or later than 2018”²⁴). The CRR was defined as the number of patients who needed to undergo a revision up to a certain time point as a proportion of the total number of patients who were at risk after a primary procedure.

For TH components (cups or stems), TH cup-stem combinations, and TK implants (tibiofemoral combinations), the following registry data were extracted: name, manufacturer, type of fixation, number of implants, and the CRR with the SE and/or 95%

CI. If only the 95% CI was provided, then the SE was calculated by subtracting the values of the upper and lower bounds of the 95% CI and dividing the result by 3.92.²⁵

The implants in the registry data were identified, on the basis of the implant name, as having received an ODEP rating or not (Figures 1 and 2). ODEP-matched implants with a B rating were excluded because such a rating is assigned for implants with limited usage.

Statistical analysis

Before comparing higher and lower-rated implants with respect to their CRRs, we assessed whether ODEP-rated implants represented a selected group of implants. Therefore, with use of independent t tests, we evaluated whether ODEP-matched implants differed from unmatched implants with and without multiple ODEP ratings (red boxes; Figures 1 and 2) in terms of the CRR.

For ODEP-matched implants, random-effects models were utilised to calculate the pooled registries' CRR at 3-, 5-, and 10-years for A*- and A-rated implants. These models included the DerSimonian-Laird estimator to consider the extent of heterogeneity among the implant designs.²⁶ The ODEP rating (A* or A) was included as a factor to test for group differences. This analysis was performed separately for TH components and TK implants. Similar random-effects models were utilised to compare A*A*-rated and AA-rated TH cup-stem combinations. The I^2 was utilised to estimate the extent of heterogeneity in the pooled registries' CRR, which was defined as low (25%), moderate (50%), or high (75%).^{27,28} To explore possible reasons for the observed heterogeneities, TH components, TK implants, and TH cup-stem combinations were stratified by fixation type and the analyses were repeated. Additionally, another analysis was performed with TH cup-stem combinations stratified by whether the individual components were from the same manufacturer or different manufacturers.

To answer the second research question, random-effects models were utilised to calculate the pooled CRR with 95% CI at 3-, 5-, and 10-years for each TH component across all registries in which the component was reported. The pooled CRR was then compared with ODEP benchmark criteria (Table 1) to assess whether the TH component met the criteria for an A* rating. We then calculated the percentage of A*-rated TH components that would receive an A* rating on the basis of the pooled registries' CRR and performed a similar calculation for A-rated TH components. Considering that the performance of an implant may differ across registries, we also examined the median number (and range) of registries in which each TH component would be assigned an A* rating as well as examined how many TH components would consistently get an A* rating in all registries in which the component was reported.

The metafor package in R (R Foundation for Statistical Computing; version 4.1.2) was utilised for meta-analyses. The level of significance was set at $p\text{-value}<0.05$.

Results

Nine registries were included (Figure 3). The latest annual reports of eight registries^{19,29-35}, consisting of data up to December 2019, and the up-to-date data (as of March 2021) from the website of one registry³⁶ were utilised. The mean percentage completeness of patient or procedure-level data in the included registries was 87.3% (range: 40%²⁹ to 99%¹⁹).

Nine registries reported on a total of 583 TH cups with unique brand names (2,615,890 implants) and 618 TH stems (2,567,442 implants), and eight registries reported on a total of 634 TH cup-stem combinations (2,266,864 implants) and 508 TK implants (2,940,899 implants) (Supplementary Tables 1 to 4 – online available). A total of 313 (54%) of the unique TH cups, 356 (58%) of the unique TH stems, 218 (34%) of the unique TH cup-stem combinations, and 68 (13%) of the unique TK implants that were

reported by registries were matched to an ODEP rating. The percentage of implants with a matched ODEP rating varied widely between registries, ranging from 35% to 69% of cups, 46% to 80% of stems, 22% to 55% of TH cup-stem combinations, and 6% to 20% of TK implants. For implants that were unmatched as a result of multiple ODEP ratings, the median number of possible ODEP ratings was two (range: two to six) for cups, two (range: two to eight) for stems, and four (range: two to 48) for TK implants (data not shown). Since only 13% of TK implants were matched, they were not further analysed. The failure to match most of the TK implants was primarily due to the fact that the granularity with which ODEP ratings are applied to a TK implant is much more detailed than most registry reports of a TK implant.

ODEP-matched versus ODEP-unmatched TH implants

ODEP-matched cups had significantly lower 3-, 5-, and 10-year CRRs than unmatched cups without an ODEP rating, and ODEP matched stems had significantly lower 5- and 10-year CRRs than unmatched stems without an ODEP rating. However, ODEP matched cups and stems had comparable CRRs to unmatched cups and stems with multiple ODEP ratings (Table 2). ODEP matched TH cup-stem combinations had significantly lower CRRs than unmatched TH cup-stem combinations at all follow-up points (Table 3).

A-rated versus A-rated TH implants*

No overall differences in CRRs were found between A*-rated and A-rated TH implants, with the exception of the CRRs for TH cup-stem combinations, which were significantly lower for A*A*-rated cup-stem combinations at the 3-year time point (Tables 4 and 5). Moderate to high (range: 64% to 95%) heterogeneity was found, reflecting variation in CRRs between implants (Tables 4 and 5). To explore this heterogeneity, the analyses were repeated with the implants stratified by fixation type, which again showed no significant differences in the CRRs at 3-, 5-, and 10-years for all analysed groups and demonstrated moderate to high heterogeneity

(data not shown). Among TH cup-stem combinations that consisted of components from the same manufacturer, A*A*-rated implants had significantly lower 3- and 5-year CRRs than AA-rated implants. Among TH cup-stem combinations with components from different manufacturers, no significant differences were found (data not shown).

ODEP ratings based on pooled registries' CRR

Among all ODEP-matched A*-rated cups and stems, 39% of cups and 42% of stems would receive an A* rating on the basis of the pooled registries' CRR at 3-years, 44% of cups and 35% of stems would receive such a rating at 5-years, and 30% of cups and 5% of stems would receive such a rating at 10-years (Table 6; Supplementary Figures 1 and 2 for implant-level results – online available). Analysing A*-rated cups and stems that were reported by ± 2 registries resulted in similar percentages at 3- and 5-years but lower percentages at 10-years than in the previous analysis (Table 6). Cups and stems qualifying for an A* rating on the basis of the pooled registries' CRR would receive an A* rating in a median of one registry at all follow-up points (range: zero to four registries [cups] and zero to six registries [stems]; Supplementary Table 5 and Figures 1 and 2 – online available). Three cups and three stems would consistently get an A* rating in all registries at 3-years; four cups and two stems, at 5-years; and three cups and zero stems, at 10-years (Supplementary Tables 5 and 6 – online available).

Among all ODEP-matched A-rated cups and stems, 24% of cups and 31% of stems would receive an A* rating on the basis of the pooled registries' CRR at 3-years, 24% of cups and 32% of stems would receive such a rating at 5-years, and 22% of cups and 23% of stems would receive such a rating at 10-years (Table 6; Supplementary Figures 3 and 4 – online available). When analysing A-rated cups and stems that were reported by $\pm$ two registries, these percentages were as follows: 27% of cups and 30% of stems at 3-years, 18% of cups and 25% of stems at 5-years, and 33% of cups and

40% of stems at 10-years (Table 6). Cups qualifying for an A* rating on the basis of the pooled registries' CRR would receive an A* rating in a median of zero registries at all follow-up points (range: zero to five registries; Supplementary Table 7 – online available). Stems qualifying for an A* rating on the basis of the pooled registries' CRR would receive an A* rating in a median of one registry (range: zero to two) at 3-years, one registry (range: zero to two) at 5-years, and zero registries (range: zero to one) at 10-years (Supplementary Table 8 – online available). Zero cups and one stem would consistently receive an A* rating in all registries at 3-years; one cup and two stems, at 5-years; and zero cups or stems, at 10-years (Supplementary Tables 7 and 8 – online available).

Discussion

This multiregistry study showed that ODEP-matched TH implants had lower CRRs than unmatched TH implants without an ODEP rating. Among matched TH implants, CRRs did not differ between implants with a higher ODEP rating and those with a lower ODEP rating. TK implants were not analysed because only 13% of the TK implants reported by registries were matched to an ODEP rating. Only 39% of A*-rated cups and 42% of A*-rated stems would be assigned the A* rating on the basis of the pooled registries' CRR at 3-years. However, 24% of A-rated cups and 31% of A-rated stems would be assigned the A* rating at 3-years, with similar or lower percentages at longer follow-up. The assigned ODEP ratings varied across registries, implying that assigned ODEP ratings do not necessarily apply to the performance of TH implants in other countries. Therefore, registries should first validate ODEP ratings with use of country-specific data to better guide implant selection in their country.

In principle, OPC such as those utilised by ODEP can help stakeholders to monitor implant performance; to stimulate the continuous evaluation of implants, which may result in a higher ODEP rating and prevent losing an ODEP rating when no data are

provided 2-years (for 3-, 5-, and 13-year ODEP ratings) or 3-years (for 7-, 10-, and 15-year ODEP ratings) after an ODEP rating has been assigned; and to use ODEP ratings to guide implant selection. ODEP aims to “promote evidence-based selection of implants so that patients receive the very best and safest implants.”³⁷ The present study showed that ODEP-matched TH implants had better performance than unmatched TH implants without an ODEP rating, suggesting that ODEP achieves this aim by encouraging surgeons and hospitals to use ODEP-rated implants.

Some prior studies benchmarked against a predefined benchmark created by a quality institute, whereas others utilised relative benchmarks, such as the performance of the best-performing implant at that time or the average performance of similar implants.^{6,11-15} Using a relative benchmark means that the judgment of whether the performance of an implant is an outlier depends on the performance of the comparator. The performance of an implant can change over time, and so too can the performance of the comparator. Therefore, even if an implant continues to have the same performance over time, that implant could become an outlier if the comparator improves. This method differs from one using absolute benchmarks such as ODEP ratings, where the OPC is predefined on the basis of what is considered to be an acceptable level of implant performance, thereby making interpretations and assessments of implant performance more straightforward.¹⁰ However, absolute benchmarks may need to be updated over time (e.g. the ODEP rating originally had a 10-year benchmark threshold of <10%¹⁰), so it has to be considered whether the OPC are still acceptable.

A prerequisite for the assignment of ODEP ratings is that manufacturers must declare that the voluntarily submitted data—which may be based on various data sources—are representative of the performance of these implants in daily clinical practice.¹⁰ The present study tested the external validity of ODEP ratings across

multiple registries and showed that approximately 40% of A*-rated cups and stems would also receive the A* rating on the basis of the pooled registries' CRR; however, we found that approximately 25% of A-rated cups and stems would receive the A* rating as well and that the A* and A ratings were inconsistent across registries. This inconsistency may be a result of differences between registries with respect to case mix; revision indications; smaller 95% CIs due to the pooling of data, which resulted in meeting the OPC; or camouflage.²² Another explanation, particularly for implants that have been utilised for decades, and in recognition of the fact that the performance of implants has improved over time, may be that the CRR applies to patients who underwent the primary operation in a different period. For some registries, the 10-year CRR of implants may include patients who underwent the operation in the previous century, whereas for newer registries, it would include patients who underwent the operation more recently. This potential discrepancy highlights the importance of including patients from the same period when combining data across multiple registries. Nonetheless, if well-established implants continue to be utilised to the same extent, the impact of patients who underwent the operation long ago on the reported revision estimates will likely be small. This inconsistency also underscores the importance of transparent reporting of the types of submitted data sources that serve as the basis for ODEP ratings, which would allow for validation of whether the data are indeed representative, as claimed by manufacturers.

Some study limitations should be noted. First, there could have been selection bias because some implants could not be matched as a result of multiple ODEP ratings and were thus excluded. However, ODEP-matched TH cups and stems had similar CRRs to unmatched TH cups and stems with multiple ODEP ratings, making selection bias unlikely. The matching problem was due to insufficient details on implants reported by registries, resulting in a large number of compatible cup stem combinations within one implant name ("camouflage").²² To solve this matching problem, which was

most prominent among TK implants, registries should register the product codes of implants, which is already done by a few registries.³⁸ Second, some registries may not have included all patients or revisions, which may have influenced the CRRs. For example, the Swedish Hip Arthroplasty Register likely underestimates revisions because it excludes revisions due to infection, and thus the actual implant-level CRRs are higher than reported.³³ These underestimated CRRs may have resulted in the assignment of an A* rating to TH implants that were commonly reported in this registry, whereas such implants might have received an A rating when including all revisions. Similarly, the American Joint Replacement Registry only includes patients with osteoarthritis who are $\pm$ 65 years old, which may again have resulted in underestimated CRRs, as the literature has generally shown a lower CRR among older patients.^{29,39} Third, registries were excluded from the analysis primarily because they did not report CRRs with SEs or 95% CIs, making data comparison and pooling impossible. This highlights the importance of international agreement across registries with regard to definitions, the amount of detail (e.g. the reporting of product codes), and methodologies to enable data pooling.²⁴ Fourth, although we evaluated the performance of A*- and A-rated TH cup-stem combinations to give insight into possible performance differences, ODEP has never rated TH cup-stem combinations, only hip components (i.e. cups and stems) separately. Rating TH components separately is aligned with clinical practice, in which clinicians mix and match cups and stems from different manufacturers, often with excellent results.⁴⁰ However, the practice of not rating TH cup-stem combinations and instead rating TH components separately may be a potential reason for some of the differences between the ODEP ratings based on the pooled registries' CRR and the ODEP ratings assigned by ODEP. Lastly, we only analysed 3-, 5-, and 10-year CRRs because, besides the 1-year CRR, these years were the most frequently reported time points, with each registry contributing at least two time points. 1-year CRRs were not analysed because they

are not utilised for ODEP ratings, whereas the 3-year follow-up is the first time point utilised by ODEP.

Conclusion

In conclusion, clinicians should be encouraged to use implants with an ODEP rating, as these implants have better CRRs than unrated implants. A minority of A*-rated cups and stems would be eligible for an A* rating on the basis of the pooled registries' CRR, with the assigned ODEP ratings varying across registries, indicating that implant performance varies across countries. Therefore, registries should first validate ODEP ratings to better guide implant selection in their country, and they should preferably do so at the product-code level to prevent camouflage. The ODEP benchmarks could be strengthened by making data submission, including transparency regarding the data source, mandatory; removing the grace period of 1-year for ODEP ratings; and using revision data from at least two regional, national, or multicountry registries with >95% implant-level completeness.^{24,41}

Figures and Tables

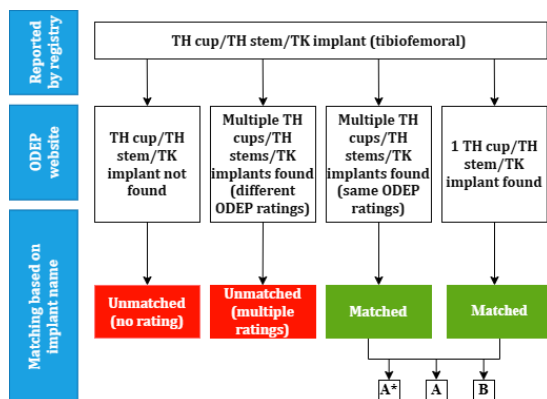


Figure 1. Diagram showing the process for matching registry-reported TH cups, TH stems, and TK implants to the ODEP rating for that implant

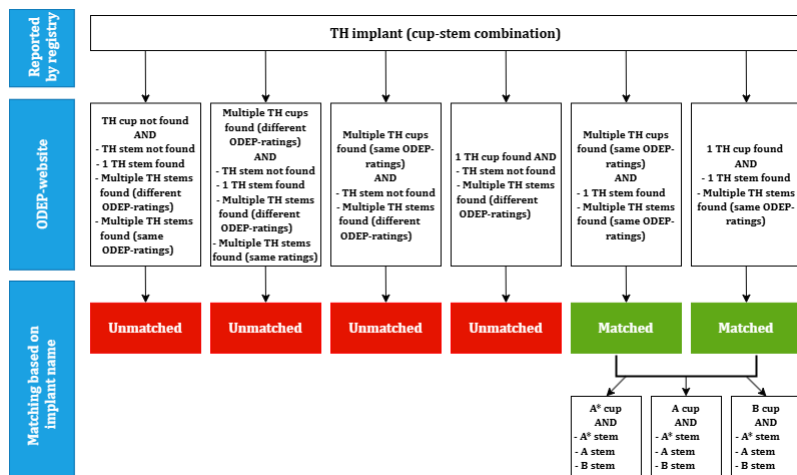


Figure 2. Flowchart showing the process for matching registry-reported TH implants (cup-stem combinations) to the ODEP rating for that implant

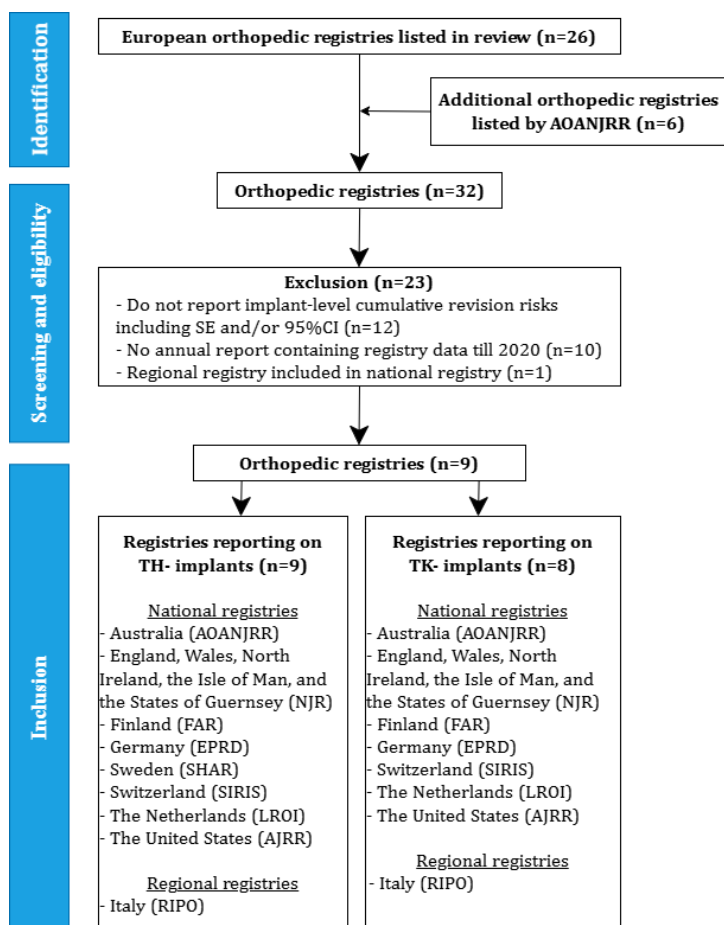


Figure 3. Flowchart showing the selection process for registries.
 AOANJRR = Australian Orthopaedic Association National Joint Replacement Registry; NJR = National Joint Registry; FAR = Finnish Arthroplasty Register; EPRD = The German Arthroplasty Registry; SHAR = Swedish Hip Arthroplasty Register; SIRIS = Swiss National Hip & Knee Joint Registry; LROI = Dutch Arthroplasty Register; AJRR = American Joint Replacement Registry; RIPO = Register of Orthopaedic Prosthetic Implantology

TH Implant: Criteria for A* Ratings	3A*	5A*	7A*	10A*	13A*	15A*
Minimum no. of centers outside development center(s)	3	3	3	3	3	3
Minimum no. of surgeons outside development center(s)	3	3	3	3	3	3
Minimum total cohort	150	250	350	500	500	500
Minimum at risk at benchmark time	150	225	300	400	400	400
Maximum revision rate [§]	3.0%	3.5%	4.0%	5.0%	6.5%	8.0%
TH Implant: Criteria for A Ratings	3A	5A	7A	10A	13A	15A
Minimum no. of centers and surgeons	3	3	3	3	3	3
Minimum total cohort	150	250	350	500	500	500
Minimum at risk at benchmark time	72	66	60	51	42	40
Maximum revision rate [§]	5.0%	5.5%	6.0%	7.0%	8.5%	10.0%
TH Implant: Criteria for B Ratings	3B	5B	7B	10B	13B	15B
Minimum no. of centers and surgeons	1	1	1	1	1	1
Minimum total cohort	100	100	100	100	100	100
Minimum at risk at benchmark time	40	40	40	40	40	40
Maximum value of 95% lower confidence limit for revision rate	3.0%	3.5%	4.0%	5.0%	6.5%	8.0%
TK Implant: Criteria for A* Ratings	3A*	5A*	7A*	10A*	13A*	15A*
Minimum no. of centers outside development center(s)	3	3	3	3	3	3
Minimum no. of surgeons outside development center(s)	3	3	3	3	3	3
Minimum total cohort	150	250	350	500	500	500
Minimum at risk at benchmark time	150	225	300	400	400	400
Maximum revision rate [§]	3.5%	4.0%	4.5%	5.0%	6.0%	6.5%
TK Implant: Criteria for A Ratings	3A	5A	7A	10A	13A	15A
Minimum no. of centers and surgeons	3	3	3	3	3	3
Minimum total cohort	150	250	350	500	500	500
Minimum at risk at benchmark time	66	60	55	51	51	45
Maximum revision rate [§]	5.5%	6.0%	6.5%	7.0%	8.0%	8.5%
TK Implant: Criteria for B Ratings	3B	5B	7B	10B	13B	15B
Minimum no. of centers and surgeons	1	1	1	1	1	1
Minimum total cohort	100	100	100	100	100	100
Minimum at risk at benchmark time	66	60	55	51	45	42
Maximum value of 95% lower confidence limit for revision rate	3.5%	4.0%	4.5%	5.0%	6.0%	6.5%

Table 1. ODEP benchmark criteria for TH and TK implants (reproduced including modification, with permission, from ODEP). For TH and TK implants, the criteria for a pre-entry A* rating is the launch of the product under Beyond Compliance, and the criteria for a pre-entry A rating is the supplying of the product details to ODEP. [§]The upper 95% CI bound for the Kaplan-Meier revision rate (1 minus survival) must be lower than the specified level.

	Matched components		Unmatched components (multiple ODEP-ratings)		Matched versus unmatched components (multiple ODEP-ratings)	Unmatched components (no ODEP-rating)		Matched versus unmatched components (no ODEP-rating)
	Revision risk	N	Revision risk	N	Mean difference (95% CI)	Revision risk	N	Mean difference (95% CI)
Cups – 3 year	2.6%	1,270,520	2.5%	645,191	0.1% (-0.25;0.39)	3.2%	379,345	-0.6% (0.32;0.94)*
Cups – 5 year	3.1%	1,406,957	3.2%	631,813	-0.1% (-0.49;0.30)	5.1%	370,942	-2.0% (1.37;2.58)*
Cups – 10 year	5.6%	944,820	5.4%	506,671	0.2% (-0.79;1.11)	11.8%	196,116	-6.3% (4.43;8.09)*
Stems – 3 year	2.7%	1,423,161	2.7%	165,456	0.0% (-0.47;0.46)	2.9%	692,944	-0.2% (-0.09;0.46)
Stems – 5 year	3.4%	1,418,673	3.4%	162,655	0.0% (-0.82;0.82)	4.2%	675,774	-0.7% (0.16;1.30)†
Stems – 10 year	6.7%	1,004,520	5.7%	112,264	1.0% (-1.73;3.80)	8.8%	606,571	-2.0% (0.33;3.74)†

* = p-value<0.001; † = p-value=0.013; ‡ = p-value=0.019.

Table 2. Cumulative revision risks of ODEP-matched versus ODEP-unmatched components

	Matched TH cup-stem implants		Unmatched TH cup-stem implants		Matched versus unmatched implants
	Revision risk	N	Revision risk	N	Mean difference (95% CI)
3 year	2.6%	799,382	2.9%	1,405,493	-0.3% (-0.58;-0.08)*
5 year	3.0%	793,761	4.0%	1,365,984	-1.0% (-1.52;-0.47)†
10 year	5.2%	503,730	8.6%	1,006,928	-3.4% (-5.08;-1.66)†
* = p-value=0.010; † = p-value<0.001					

Table 3. Cumulative revision risks of ODEP-matched versus ODEP-unmatched TH cup-stem combinations

	A* components			A components			A* versus A components	
	Revision risk	N	Registries included (n)	Revision risk	N	Registries included (n)	Mean Difference (95% CI)	I ²
Cups								
3 year	2.3%	1,058,495	7	2.6%	153,979	5	-0.2% (-1.19;0.71)	78%
5 year	2.7%	1,302,734	9	2.9%	180,830	7	-0.3% (-1.34;0.78)	86%
10 year	4.3%	1,030,923	6	5.9%	137,499	5	-1.5% (-3.55;0.49)	90%
Stems								
3 year	2.3%	1,098,938	7	2.3%	288,025	7	0.1% (-0.60;0.74)	67%
5 year	3.0%	1,109,707	8	3.0%	311,695	8	0.0% (-0.76;0.81)	70%
10 year	5.4%	1,001,275	6	6.7%	170,134	5	-1.3% (-4.63;2.08)	95%

Table 4. Cumulative revision risks of A*-rated versus A-rated TH components

	A*A* cup-stem implants			AA cup-stem implants			A* cup + A stem implants		
	Revision risk	N	Registries included (n)	Revision risk	N	Registries included (n)	Revision risk	N	Registries included (n)
3 year	2.1%	448,940	7	3.2%	16,066	4	2.5%	191,696	7
5 year	2.7%	452,788	8	3.7%	17,121	5	3.0%	211,212	8
10 year	5.1%	351,180	5	6.0%	14,891	4	4.7%	116,519	4

Table 5. Cumulative revision risks of higher versus lower-rated TH cup-stem combinations

	A cup + A* stem implants			A*A* versus AA implants	
	Revision risk	N	Registries included (n)	Mean Difference (95% CI)	I ²
3 year	2.2%	86,761	5	-1.1% (-0.08;-2.11)*	66%
5 year	2.6%	87,954	6	-1.1 (-2.15;0.04)	64%
10 year	4.6%	83,244	5	-0.9% (-3.45;1.61)	79%
* = p-value=0.035					

Table 5 – continued. Cumulative revision risks of higher versus lower-rated TH cup-stem combinations

	Unique components		Unique components utilised in ≥2 registries	
	Total	No. (%) that met the A* benchmark	Total	No. (%) that met the A* benchmark
A* cups				
3 year	33	13 (39%)	23	9 (39%)
5 year	36	16 (44%)	25	11 (44%)
10 year	30	9 (30%)	18	4 (22%)
A* stems				
3 year	33	14 (42%)	25	12 (48%)
5 year	31	11 (35%)	24	8 (33%)
10 year	20	1 (5%)	14	-
A cups				
3 year	17	4 (24%)	11	3 (27%)
5 year	17	4 (24%)	11	2 (18%)
10 year	9	2 (22%)	3	1 (33%)
A stems				
3 year	29	9 (31%)	10	3 (30%)
5 year	28	9 (32%)	12	3 (25%)
10 year	13	3 (23%)	5	2 (40%)

Table 6. A*- and A-Rated TH components that met the OPC for an A* rating on the basis of pooled registries' CRR

References

1. Jarow JP, Baxley JH. Medical devices: US medical device regulation. *Urol Oncol*. 2015 Mar;33(3):128-32
2. Melvin T. The European Medical Device Regulation-What Biomedical Engineers Need to Know. *IEEE J Transl Eng Health Med*. 2022 Jul 28;10:4800105
3. Fraser AG, Nelissen RGHH, Kjærsgaard-Andersen P, et al; CORE-MD Investigators. 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 May 5;8(3):249-58
4. Sayers A, Crowther MJ, Judge A, et al. Determining the sample size required to establish whether a medical device is non-inferior to an external benchmark. *BMJ Open*. 2017 Aug 28;7(8):e015397
5. 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 (last accessed 2 April, 2024)
6. de Steiger RN, Miller LN, Davidson DC, et al. Joint registry approach for identification of outlier prostheses. *Acta Orthop*. 2013 Aug;84(4): 348-52
7. Australian Government Department of Health. Prostheses List Guide to listing and setting benefits for prostheses. 2017 Feb. Available: <https://www.health.gov.au/sites/default/files/documents/2020/06/prostheses-listguide.pdf> (last accessed 2 April, 2024)
8. Poolman RW, Verhaar JA, Schreurs BW, et al. Finding the right hip implant for patient and surgeon: the Dutch strategy—empowering patients. *Hip Int*. 2015 Mar-Apr;25(2):131-7
9. Chubb HA, Cornish ER, Hallstrom BR, et al. Early Benchmarking Total Hip Arthroplasty Implants Using Data from the Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI). *Orthop Res Rev*. 2021 Nov 24;13:215-28
10. Orthopaedic Data Evaluation Panel (ODEP). Available: <https://www.odep.org.uk/> (last accessed 25 Augustus, 2023)
11. Keurentjes JC, Pijls BG, Van Tol FR, et al. Which implant should we use for primary total hip replacement? A systematic review and meta-analysis. *J Bone Joint Surg Am*. 2014 Dec 17;96(Suppl 1):79-97
12. Deere KC, Whitehouse MR, Porter M, et al. Assessing the noninferiority of prosthesis constructs used in hip replacement using data from the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man: a benchmarking study. *BMJ Open*. 2019 Apr 29;9(4):e026685
13. Deere KC, Whitehouse MR, Porter M, et al. Assessing the noninferiority of prosthesis constructs used in total and unicondylar knee replacements using data from the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man: a benchmarking study. *BMJ Open*. 2019 Apr 29;9(4): e026736

14. Wyatt M, Frampton C, Whitehouse M, et al. Benchmarking total hip replacement constructs using noninferiority analysis: the New Zealand joint registry study. *BMC Musculoskelet Disord.* 2021 Aug 21;22(1):719
15. Nederlandse Orthopaedische Vereniging (NOV). Procedure Uitkomstanalyse. Available: <https://www.orthopeden.org/kwaliteitsbeleid/procedureuitkomstanalyse/> (last accessed 4 January, 2023)
16. Nederlandse Orthopaedische Vereniging (NOV). Procedure Uitkomstanalyse. Available: <https://www.orthopeden.org/kwaliteitsbeleid/procedureuitkomstanalyse/> (last accessed 4 January, 2023)
17. Orthopaedic Data Evaluation Panel (ODEP). ODEP Hip Submission Form. Available: https://www.odep.org.uk/wp-content/uploads/2024/02/ODEP-HipSubmissionForm_2021_v7.98.xlsx (last accessed 2 May, 2024)
18. Orthopaedic Data Evaluation Panel (ODEP). Impact of ODEP. Available: <https://www.odep.org.uk/about/impact-of-odep/> (last accessed 25 Augustus, 2023)
19. Landelijke Registratie Orthopedische Implantaten (LROI). Online LROI annual report 2020. Available: <https://www.lroi-report.nl/app/uploads/2021/03/PDF-LROI-annual-report2020.pdf> (last accessed 2 April, 2024)
20. Gundersen T, Furnes O, Gjertsen JE, et al. Norwegian National Advisory Unit on Arthroplasty and Hip Fractures: Annual Report 2021. Available: <https://www.helse-bergen.no/nasjonalt-kvalitets-og-kompetansenettverkfor-leddproteser-og-hoftebrudd/arsrapporter/#annual-reports>
21. Phillips JRA, Tucker K. Implant brand portfolios, the potential for camouflage of data, and the role of the Orthopaedic Data Evaluation Panel in total knee arthroplasty. *Bone Joint J.* 2021 Oct;103-B(10):1555-60
22. Orthopaedic Data Evaluation Panel (ODEP). ODEP Rating System. Available: <https://www.odep.org.uk/about/rating-system/> (last accessed 25 Augustus, 2023)
23. Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR). Other Registries Worldwide. Available: <https://aoanjrr.sahmri.com/registries> (last accessed 31 January, 2023)
24. Hoogervorst LA, Geurkink TH, Lübbecke 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
25. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol.* 2005 Apr 20;5:13
26. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986 Sep;7(3):177-88
27. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ.* 2003 Sep 6;327(7414):557-60
28. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002 Jun 15;21(11):1539-58
29. American Joint Replacement Registry (AJRR). Annual Report 2020 The

Seventh Annual Report of the AJRR on Hip and Knee Arthroplasty. Available: <https://connect.registryapps.net/2020-ajrr-annual-report> (last accessed 2 April, 2024)

30. Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR). Australian Orthopaedic Association National Joint Replacement Registry Hip, Knee & Shoulder Arthroplasty 2020. Available: <https://benjeffcote.com.au/wp-content/uploads/Hip-Knee-Shoulder-ArthroplastyAnnual-Report.pdf> (last accessed 2 April, 2024)

31. German Arthroplasty Registry (EPRD). The German Arthroplasty Annual Report 2020. Available: https://www.eprd.de/fileadmin/user_upload/Dateien/Publikationen/Berichte/AnnualReport2020-Web_2021-05-11_F.pdf (last accessed 2 April, 2024)

32. Regional Register of Orthopaedic Prosthetic Implantology. Overall Data Hip, Knee and Shoulder Arthroplasty In Emilia-Romagna Region (Italy) 2000-2019. Available: https://ripo.cineca.it/authzssl/pdf/Annual%20report%202019%20RIPO%20register_v1.pdf (last accessed 2 April, 2024)

33. Swedish Hip Arthroplasty Register (SHAR). Swedish Hip Arthroplasty Register Annual report 2019. Available: https://registercentrum.blob.core.windows.net/shpr/r/VGR_Annualreport_SHAR_2019_EN_Digital-pages_FINAL-ryxaMBUWZ_.pdf (last accessed 2 April, 2024)

34. Swiss National Hip & Knee Joint Registry (SIRIS). Swiss National Hip & Knee Joint Registry Report 2020. Available: https://www.anq.ch/wp-content/uploads/2021/01/ANQakut_SIRIS_HipsKnee_Annual-Report_2020.pdf (last accessed 2 April, 2024)

35. National Joint Registry (NJR). 17th Annual Report 2020. Available: <https://reports.njrcentre.org.uk/portals/0/pdfdownloads/njr%2017th%20annual%20report%202020.pdf> (last accessed 2 April, 2024)

36. Finnish Arthroplasty Register (FAR). Finnish Arthroplasty Register. Available: <https://www.thl.fi/endo/report/#index> (last accessed 9 September, 2022)

37. Orthopaedic Data Evaluation Panel (ODEP). What is ODEP? Available: <https://www.odep.org.uk/> (last accessed 24 Augustus, 2023)

38. Denissen GAW, van Steenberghe LN, Lollinga WT, et al. Generic implant classification enables comparison across implant designs: the Dutch Arthroplasty Register implant library. EFORT Open Rev. 2019 Jun 3;4(6):344-50

39. Prokopetz JJ, Losina E, Bliss RL, et al. Risk factors for revision of primary total hip arthroplasty: a systematic review. BMC Musculoskelet Disord. 2012 Dec 15;13:251

40. Tucker K, Günther KP, Kjaersgaard-Andersen P, et al. EFORT recommendations for off-label use, mix & match and mismatch in hip and knee arthroplasty. EFORT Open Rev. 2021 Nov 19;6(11):982-1005

41. Nelissen RGHH, Kjaersgaard-Andersen P, Tucker K. Quality Assessment of Orthopaedic implants. White paper EFORT. 2019 Feb 19. Available: https://nore.efort.org/documents/7050849/7792706/EFORT_Implant_1Qualit_y_EFORT_Position_paper.pdf/d46de1ad-ff60-9b41-b8bb-efa76048aae4 (last

accessed 2 April, 2024)

42. Orthopaedic Data Evaluation Panel (ODEP). ODEP Methodology for Standard (and Patient Matched) Total Hip Replacements THRs. 2021 Feb 22. Available: <https://www.odep.org.uk/methodology/methodology-for-hips/> (last accessed 7 May, 2024)

43. Orthopaedic Data Evaluation Panel (ODEP). ODEP Methodology for Total Knee Replacements. Available: <https://www.odep.org.uk/methodology/methodology-for-tnr/> (last accessed 14 October, 2020)