


Early Benchmarking Total Hip Arthroplasty Implants Using Data from the Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI)

Heather A Chubb¹
Eric R Cornish²
Brian R Hallstrom¹
Richard E Hughes¹ 

¹Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, MI, USA; ²Department of Orthopedic Surgery, MidMichigan Health, Alpena, MI, USA

Background: Benchmarking arthroplasty implant revision risk is an informative way to address implant performance. National benchmarking efforts exist in the United Kingdom, Netherlands, and Australia. Recently, the International Prosthesis Benchmarking Working Group, including representatives from industry, academia, and national registries, produced a guideline describing arthroplasty benchmarking methodology. The proposal was applied to data from the Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) to assess its feasibility for benchmarking implants in the United States.

Methods: Primary elective total hip arthroplasty procedures performed for osteoarthritis between 2/15/2012 and 12/31/2018 and their associated revisions were identified in the MARCQI registry. The guidelines recommend that all prostheses combinations receive an early benchmark if they have at least 250 procedures at risk and the revision rate does not exceed the pre-determined standard of 2% at 2 years and 3% at 5 years.

Results: A total of 72,949 primary cases met the inclusion criteria. Of these, 1369 had revisions. Twenty-nine and six stem/cup combinations satisfied the minimum case requirement at 2 and 5 years, respectively. Three implant combinations would not receive a benchmark at 2 years: Secur-Fit/Trident, Anthology/Reflection 3, Taperloc 133/G7.

Conclusion: The guideline can be implemented in the United States by a regional registry. Moreover, not all hip implants currently in use would receive an early benchmark. This raises concern as these implant combinations represent a significant number of cases in Michigan, some with increasing utilization.

Keywords: arthroplasty, hip, implant, benchmarking, revision

Plain Language Summary

Some total hip replacement implants are better than others. An international group has proposed a method for “benchmarking” implants, which means identifying which ones perform well. The Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI), which is a state-wide collaborative of hospitals and orthopaedic surgeons dedicated to improving the quality of care for hip and knee replacement patients in Michigan, has applied the proposed method to their data. They found that three implant combinations were not good enough to receive a benchmark based on their data. The results suggest health care quality can be improved by surgeons using implants that receive benchmarks.

Introduction

Elective total hip arthroplasty is a common in the United States, with over 522,000 hip replacements performed in 2014.¹ There is wide variation in revision risk

Correspondence: Richard E Hughes
Department of Orthopaedic Surgery,
University of Michigan, 2003 BSRB, 109 Zina
Pitcher Pl, Ann Arbor, MI, 48109-2200, USA
Fax +1 734 647-0004
Email rehughes@umich.edu

between total hip arthroplasty (THA) implants, with arthroplasty registry reports showing a range of 10-year revision risks for cemented implants from 1.03% to 36.2% and from 2.6% to 66.5% for uncemented fixation.² Voluntary implant product recalls by manufacturers are rare, and the Food and Drug Administration is reluctant to recall implants. It is imperative that arthroplasty registries play a public health role in providing information for surgeons and patients. Internationally, arthroplasty registries seek to reduce the number of revisions in three ways: (1) public reporting through annual reports, (2) identifying outlier implants, (3) implant benchmarking.

Benchmarking is a systematic process of determining whether an implant meets specified performance levels.^{3,4} There are currently three groups performing THA implant benchmarking: (1) Orthopaedic Data Evaluation Panel (ODEP) in the United Kingdom,^{5,6} (2) Prostheses List Advisory Committee (PLAC) in Australia, (3) Netherlands Orthopaedic Association Classification of Orthopaedic Implants (NOV).⁷ The International Prosthesis Benchmarking Working Group was established to review current systems and develop a global system proposal to evaluate and benchmark arthroplasty prostheses performance. The working group produced a guidance document in May of 2018 which focused on benchmarking hip and knee implants.⁸

The statistical subcommittee of the working group analyzed Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) data⁹ and determined that poor implant performance at two years is predictive of poor performance at ten years.⁸ Therefore, early benchmarking is extremely important, as devices with inferior performance at two years rarely recover.

The purpose of this project was to assess the feasibility of applying the proposal to arthroplasty registry data collected by a regional registry in the United States.

Materials and Methods

Utilizing the methodology detailed in the International Prosthesis Benchmarking Workgroup's proposal, benchmarking of prostheses in the MARCQI database was performed. All data collected by MARCQI is for the purposes of quality improvement. The Institutional Review Board of the University of Michigan's Medical School (IRBMED) provided a notice of determination of "not regulated" status for this project because it does not fit the definition of human subject research according to 45 CFR 46 and 21 CFR 56. "Not regulated" status is different than "exempt,"

and it reflects that the purpose of the data collection was quality or process improvement. This notice is available upon request. The details of MARCQI's organizational structure and methodology have been previously described, but in summary, MARCQI collects data on over 97% of all elective total hip and knee arthroplasty cases performed in the state of Michigan.¹⁰⁻¹² To qualify for inclusion in the MARCQI registry, a primary case must be elective, defined by a planned procedure, treating a non-emergent condition at the pre-planned surgical date. Hemiarthroplasty cases are excluded, as well as non-elective total arthroplasty cases such as for hip fracture. 91.9% of the cases were done for a diagnosis of osteoarthritis. All total hip and knee replacement revisions are captured and linked to the primary case. The linkage can occur across hospitals in the state of Michigan. Thus, a revision case performed at a different site than the primary is linked to the primary case. This enables MARCQI to conduct analyses based on time-to-revision for primary procedures.

Each site has free access to their individual data. MARCQI performs revision risk analyses of implants and publicly reports the results in an annual report which is readily available to the public online.¹²⁻¹⁴ The dataset used to generate the most recent annual report was used in this benchmarking analysis.¹⁴ In addition to demographic and clinical data collected on each case, catalog numbers of every device implanted are captured. These catalog numbers are converted to device descriptors (stem, cup, head, liner, product name, etc.) using a device library made available by Curvo Labs (Evansville, IN). The Curvo Labs library matches 98.5% of all devices used in MARCQI cases. The cup, stem, and product name fields were utilized in the benchmarking analysis.

Data for this study specifically was limited to total hip arthroplasty (THA) cases performed between 2/15/2012, and 12/31/2018. Due to the design of the MARCQI registry, all qualifying primary cases were elective, either conventional or conversion, and patients were at least 18 years old. The International Prosthesis Working Group proposed protocol describes benchmarking based only on cases performed for a diagnosis of osteoarthritis. Therefore, for this analysis, inclusion was restricted to a primary diagnosis of osteoarthritis. The clinical endpoint used was all-cause revision. The exclusion criteria were: (1) cases performed before 2/15/2012 or after 12/31/2018, (2) knee procedure, (3) resurfacing THA procedure, (4)

diagnosis other than osteoarthritis, or (5) otherwise non-qualifying MARCQI case.

The benchmarking guidelines proposed by the International Prosthesis Working Group focus on the 10-year time point. The philosophy of the working group was that implant combinations would receive early benchmarks by default at two and five years unless the device revision risk exceeded a predetermined standard. Thus “early” benchmark procedures at 2 and 5 years were described.

MARCQI began collecting data in 2012 and does not yet have data for a 10 year benchmark analysis.⁸ Therefore, the analysis focused on the early benchmark time points of 2 and 5 years. The proposed methodology was based on Kaplan-Meier estimates of time-to-revision following primary procedure and the associated 95% confidence intervals. The International Prosthesis Benchmarking Working Group document proposed that all prostheses combinations will receive an early benchmark if they have at least 250 procedures at risk and the lower 95% confidence limits of the revision rate does not exceed the proposed benchmark standard of 2% at 2 years and 3% at 5 years. The early benchmarking standard would not be provided if the lower 95% confidence interval exceeds the proposed benchmark standard. The benchmarking criteria were applied to each stem/cup combination separately. Based upon working group recommendations, the criteria were then applied to stems aggregated across cups and cups aggregated across stems.

The percentage of primary hip cases performed using implants that failed to receive early benchmarks was computed to provide a population-wide quality measure. This measure was computed using the total number of primary THA cases in the MARCQI database as the denominator. It was also done by year to provide a time trend.

The International Prosthesis Benchmarking Working Group proposed that benchmarks be determined regardless of specific patient characteristics. However, the working group recommended that patient characteristics, such as age and gender, be summarized with revision rates. Therefore, prostheses combination revision rates were also evaluated by gender and age groups (less than 65 years and 65 years and older).

All statistical analyses were performed in SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA) and mathematical modeling was done using Excel (Microsoft, Redmond, WA, USA).

Results

A total of 72,949 primary cases met the inclusion criteria for benchmarking (Figure 1). Of these, 1,369 had revisions in the database. At 2 years and 5 years respectively, twenty-nine and six stem/cup combinations satisfied the requirement that there be at least 250 at-risk cases (Table 1). Twenty-six individual femoral stem components satisfied the minimum at-risk case threshold at 2 years and seven met it at 5 years (Table 2). Fifteen individual acetabular cup components met the minimum at-risk case requirement at 2 years and seven met it at 5 years (Table 3).

The majority of stem/cup combinations and individual components achieved early benchmarks at the 2- and 5-year time points. At 2 years, twenty-six stem/cup combinations received a benchmark, while three prostheses combinations did not (Figure 2). The total number at risk at 2 years was 35,887, and the number at risk for the 3 combinations that did not receive a benchmark was 2,837. In Figure 2 the vertical dotted line denotes the 2% benchmark criteria at 2 years. Any combination where the lower confidence limit falls to the right of this line does not meet the pre-determined benchmark standard. The three combinations that do not receive an early benchmark (Secur-Fit /Trident, Anthology/Reflection 3, and Taperloc 133/G7) have lower confidence limits of 3.67%, 2.27%, and 2.06%, respectively. All other combinations had 95% confidence intervals whose lower limit was no greater than 2%. At 5 years all stem/cup combinations received a benchmark (Figure 3). The total number of risk at 5 years was 4,111. However, the three combinations that did not receive an early 2-year benchmark were not assessed, as they did not meet the minimum requirement of 250 at-risk cases in the MARCQI registry at the 5-year time point. The analysis of stem components in isolation at 5 years showed all would receive a benchmark (Figure 4), and there were 4,992 at risk at 5 years. Only six of the seven acetabular cup components aggregated across stems

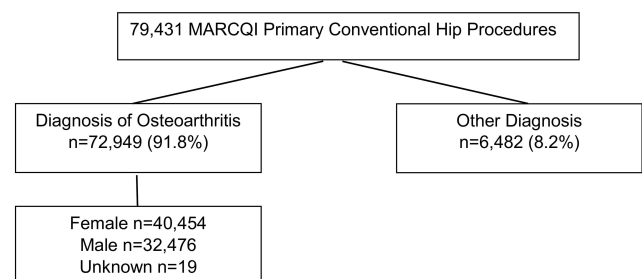


Figure 1 Case flow diagram.

Table 1 Characteristics of Stem/Cup Combination Analysis at 2 and 5 Years

Stem/Cup Combinations	2 Year Early Benchmark (2%)			5 Year Early Benchmark (3%)		
	No. at Risk	CPR Estimate (95% Confidence Interval)	Benchmark	No. at Risk	CPR Estimate (95% Confidence Interval)	Benchmark
Tri-Lock BPS/Pinnacle	1154	0.89 (0.54, 1.45)	Yes	759	1.88 (1.45, 2.44)	Yes
Tri-Lock BPS/Trident	256	1.02 (0.33, 3.15)	Yes			
Fitmore/Trabecular Metal	310	1.26 (0.47, 3.32)	Yes			
Citation TMZF/Trident	294	1.32 (0.50, 3.47)	Yes			
Summit/Pinnacle	3522	1.45 (1.14, 1.84)	Yes			
Avenir Muller/Continuum	401	1.49 (0.71, 3.10)	Yes			
Taperloc 133 Microplasty/G7	734	1.58 (1.02, 2.45)	Yes			
Accolade II/Restoration ADM	316	1.62 (0.73, 3.58)	Yes	1051	2.69 (2.32, 3.12)	Yes
Accolade II/Trident	9186	1.64 (1.44, 1.86)	Yes			
Fitmore/Continuum	1790	1.66 (1.22, 2.25)	Yes			
Corail/Pinnacle	1132	1.67 (1.17, 2.38)	Yes			
Accolade TMZF/Trident	824	1.72 (1.04, 2.83)	Yes			
M/L Taper/Continuum	4600	1.80 (1.49, 2.19)	Yes			
Secur-Fit Plus Max/Trident	1450	2.04 (1.47, 2.83)	Yes	479	2.66 (1.91, 3.70)	Yes
Trabecular Metal/Continuum	472	2.04 (1.16, 3.56)	Yes			
Polarstem/Reflection 3	282	2.05 (1.24, 3.36)	Yes			
M/L Taper/Trilogy	1096	2.07 (1.41, 3.03)	Yes			
SROM/Pinnacle	700	2.08 (1.30, 3.33)	Yes			
AML/Pinnacle	261	2.09 (1.07, 4.03)	Yes			
M/L Taper/Trabecular Metal	440	2.14 (1.25, 3.67)	Yes			
M/L Taper/G7	439	2.17 (1.42, 3.31)	Yes	307	3.24 (2.43, 4.30)	Yes
Taperloc 133/RingLoc+	975	2.33 (1.60, 3.37)	Yes			
Secur-Fit Max/Trident	1286	2.42 (1.81, 3.24)	Yes			
Taperloc 133/G7	937	2.63 (2.06, 3.36)	No			
Echo Bi-Metric/Regenerex RingLoc+	251	2.68 (1.29, 5.54)	Yes			
Synergy/Reflection 3	531	2.79 (1.82, 4.25)	Yes			
Anthology/Reflection 3	1270	2.95 (2.27, 3.82)	No			
Taperloc 133/Regenerex RingLoc+	348	2.97 (1.74, 5.07)	Yes			
Secur-Fit/Trident	630	4.87 (3.67, 6.46)	No			

would receive a benchmark at 5 years (Figure 5). There were 6,706 cases at risk at 5 years for the cup analysis.

Specific age and gender requirements are not given for conventional hip replacement; however, benchmarking may have clinical indications following appropriate stratification. Revision rates with 95% confidence intervals, stratified by gender and age group for prostheses combinations, provide additional information about the performance of an implant. Applying the 2% pre-determined benchmark criteria at the 2-year time point, three stem/cup combinations perform better in one gender group and one combination does not perform well in males or females (Table 4). Likewise, five stem/cup combinations perform better in one age group and one does not perform

well in either age group, below 65 years or 65 years and above (Table 5).

The proportion of cases in Michigan utilizing implant combinations which did not receive a 2-year benchmark was 8.6% of primary THA cases from 2/15/2012 through 12/31/2018. Moreover, some combinations show an increasing utilization trend over time (Figure 6).

Discussion

The purpose of this project was to assess the feasibility of applying the implant benchmarking methodology developed by the International Prosthesis Benchmarking Working Group to a regional arthroplasty registry in the United States. The result was that there were sufficient

Table 2 Characteristics of Femoral Stem Analysis at 2 and 5 Years

Femoral Stem	2 Year Early Benchmark (2%)			5 Year Early Benchmark (3%)		
	No. at Risk	CPR Estimate (95% Confidence Interval)	Benchmark	No. at Risk	CPR Estimate (95% Confidence Interval)	Benchmark
Tri-Lock BPS	1415	0.92 (0.58, 1.44)	Yes	765	1.93 (1.49, 2.49)	Yes
Avenir Muller	494	1.25 (0.65, 2.39)	Yes			
Citation TMZF	294	1.32 (0.50, 3.47)	Yes			
Summit	3538	1.47 (1.16, 1.86)	Yes			
Trabecular Metal	617	1.59 (0.94, 2.67)	Yes			
Natural	299	1.64 (0.69, 3.91)	Yes			
Taperloc	413	1.64 (0.83, 3.26)	Yes			
Taperloc 133 Microplasty	1112	1.66 (1.16, 2.38)	Yes			
Corail	1134	1.67 (1.17, 2.38)	Yes	1109	2.79 (2.42, 3.21)	Yes
Accolade II	9615	1.69 (1.50, 1.91)	Yes			
Accolade TMZF	824	1.72 (1.04, 2.83)	Yes			
Echelon	265	1.88 (0.85, 4.14)	Yes			
Fitmore	2351	1.90 (1.49, 2.43)	Yes			
M/L Taper	6665	1.93 (1.66, 2.24)	Yes			
Secur-Fit Plus Max	1453	2.02 (1.46, 2.81)	Yes			
SROM	716	2.03 (1.27, 3.25)	Yes			
AML	263	2.08 (1.07, 4.01)	Yes	309	3.22 (2.41, 4.28)	Yes
Polarstem	290	2.14 (1.32, 3.45)	Yes			
M/L Taper Kinectiv	457	2.20 (1.25, 3.86)	Yes			
Secur-Fit Max	1291	2.41 (1.80, 3.22)	Yes			
Synergy	673	2.41 (1.59, 3.65)	Yes			
Taperloc 133	3087	2.51 (2.12, 2.97)	No			
Echo Bi-Metric	416	2.52 (1.64, 3.87)	Yes			
Anthology	1466	2.74 (2.14, 3.52)	No	325	3.28 (2.69, 4.00)	Yes
Versys	261	3.20 (1.68, 6.06)	Yes			
Secur-Fit	632	4.70 (3.54, 6.24)	No			

numbers of implants in the MARCQI registry to conduct benchmarking at the early time points (2 and 5 years), but the registry has not been in existence long enough to conduct a later assessment (10 year). While the majority of implants received a benchmark, some did not. MARCQI's application of the proposed benchmarking methodology revealed that 8.6% of primary THA cases captured by MARCQI across the state of Michigan were done with an implant combination that would not receive an early benchmark. The rising use of these non-benchmarked implants may increase the risk of revision among patients and merits continued surveillance.

It is important to note that one limitation of benchmarking is the difficulty to detect early impact of small changes in a prosthesis until a sufficient number of cases (250) are performed. There is ongoing debate to "lump" similar prostheses together for larger numbers and statistical significance, or "split" prostheses with minor changes

into smaller groups for analysis which spreads out the time to achieve statistical significance. At this time, there are no established guidelines to categorize a new change as significantly different to "lump" or "split." Splitting may have some benefit in the interest of promoting innovation.

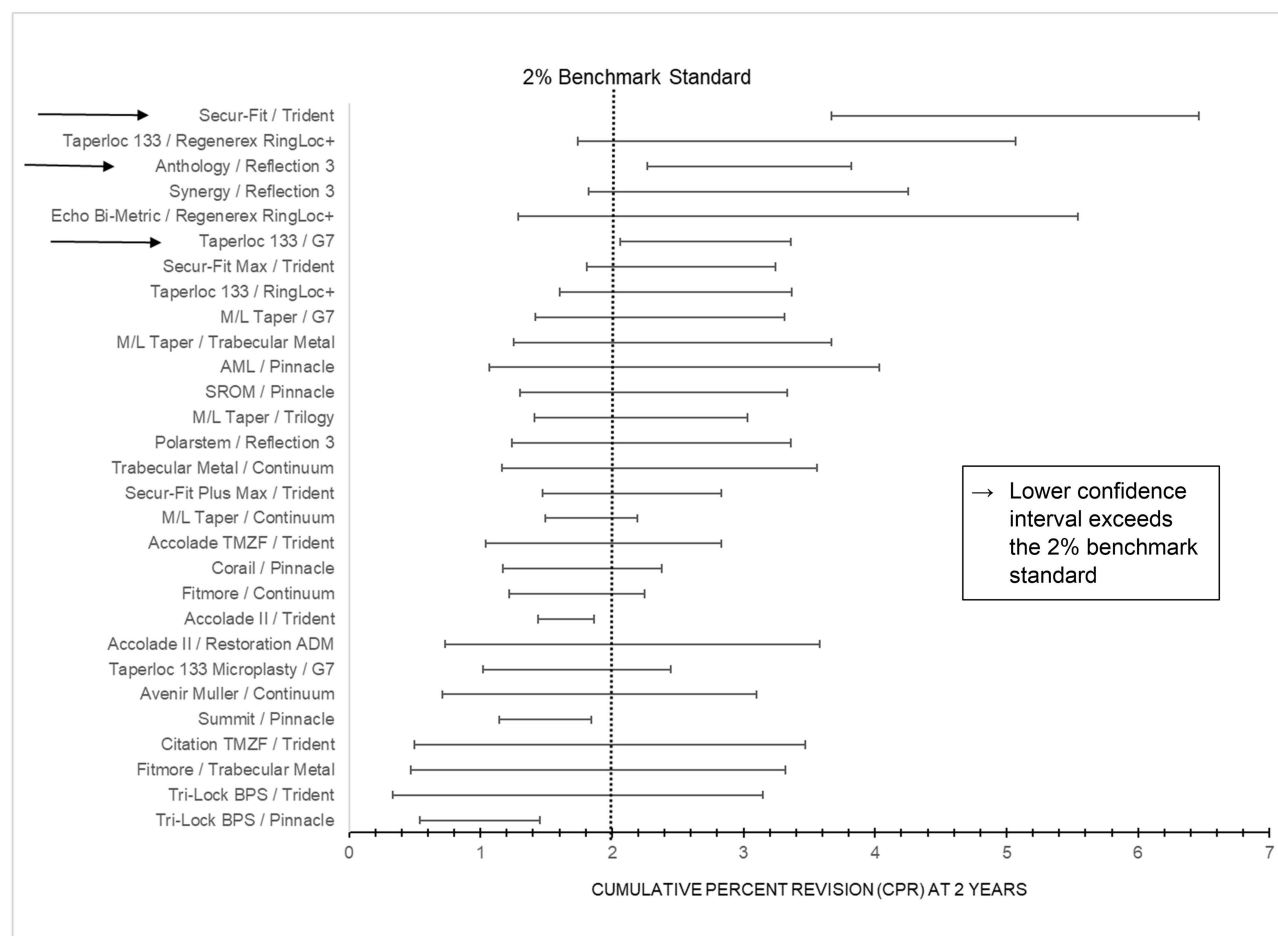
An additional limitation of the early benchmarking methodology proposed by the International Prosthesis Benchmarking Working Group is that benchmarks are based on a non-inferiority analytical framework rather than superiority. In simplistic terms, a superiority analysis requires that the upper end of a 95% confidence interval be less than a pre-specified threshold. In a non-inferiority analysis, a margin is added to the threshold to obtain a new non-inferiority threshold. Non-inferiority is determined if the upper end of the 95% confidence interval is no greater than the non-inferiority threshold. Applying a clinically accepted non-inferiority margin of 20% to the pre-determined criteria of 2% at 2 years sets

Table 3 Characteristics of Acetabular Cup Analysis at 2 and 5 Years

Acetabular Cup	2 Year Early Benchmark (2%)			5 Year Early Benchmark (3%)		
	No. at Risk	CPR Estimate (95% Confidence Interval)	Benchmark	No. at Risk	CPR Estimate (95% Confidence Interval)	Benchmark
Ranawat-Burstein	411	0.48 (0.12, 1.91)	Yes	1145	2.01 (1.67, 2.42)	Yes
Mallory-Head	388	0.96 (0.36, 2.54)	Yes			
Pinnacle	7079	1.41 (1.20, 1.66)	Yes			
Restoration ADM	323	1.56 (0.70, 3.46)	Yes			
Converge	447	1.74 (0.91, 3.31)	Yes			
Reflection	425	1.87 (0.98, 3.56)	Yes	2285	2.88 (2.60, 3.20)	Yes
Trident	14,329	1.92 (1.74, 2.11)	Yes			
Continuum	8585	1.97 (1.73, 2.25)	Yes			
Trabecular Metal	1063	2.09 (1.43, 3.03)	Yes			
Trilogy	1288	2.13 (1.51, 3.00)	Yes			
RingLoc+	1443	2.2 (1.60, 3.02)	Yes			
G7	2798	2.25 (1.93, 2.63)	Yes			
Regenerex RingLoc+	760	2.38 (1.56, 3.63)	Yes			
RingLoc	284	2.51 (1.34, 4.66)	Yes			
Reflection 3	2462	2.74 (2.27, 3.30)	No	260	3.80 (3.00, 4.81)	No

the non-inferiority threshold at 2.4%. A non-inferiority analysis finds that of the three combinations that would not receive an early benchmark, Secur-Fit/Trident is

classified as inferior, but the evidence against Anthology/Reflection 3 and Taperloc 133/G7 is inconclusive (Figure 7).

**Figure 2** Benchmarking stem/cup combinations at 2-year time point.

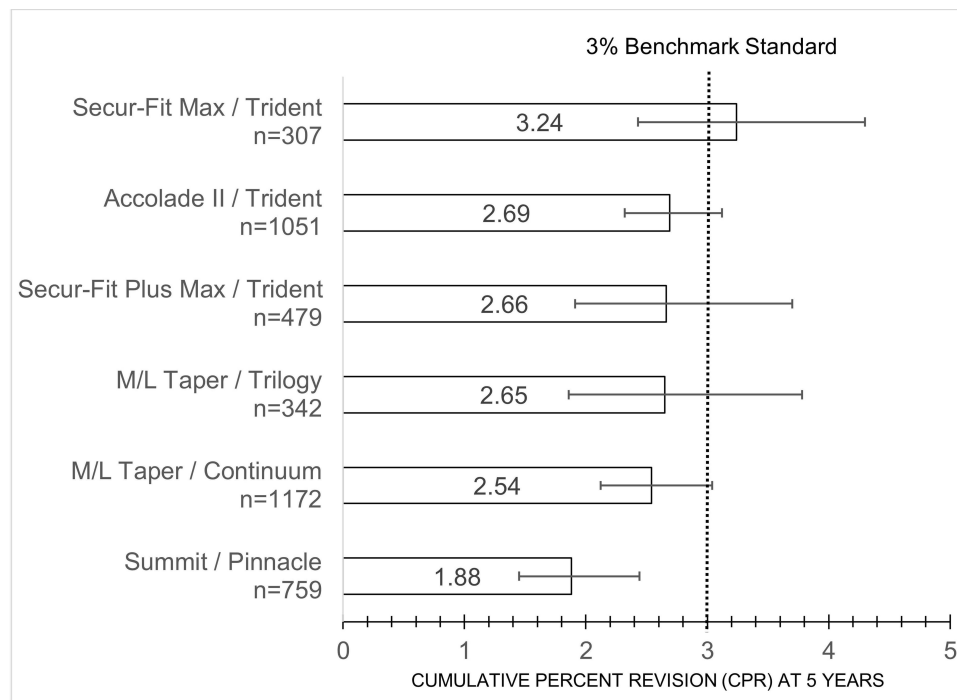


Figure 3 Benchmarking stem/cup combinations at 5-year time point.

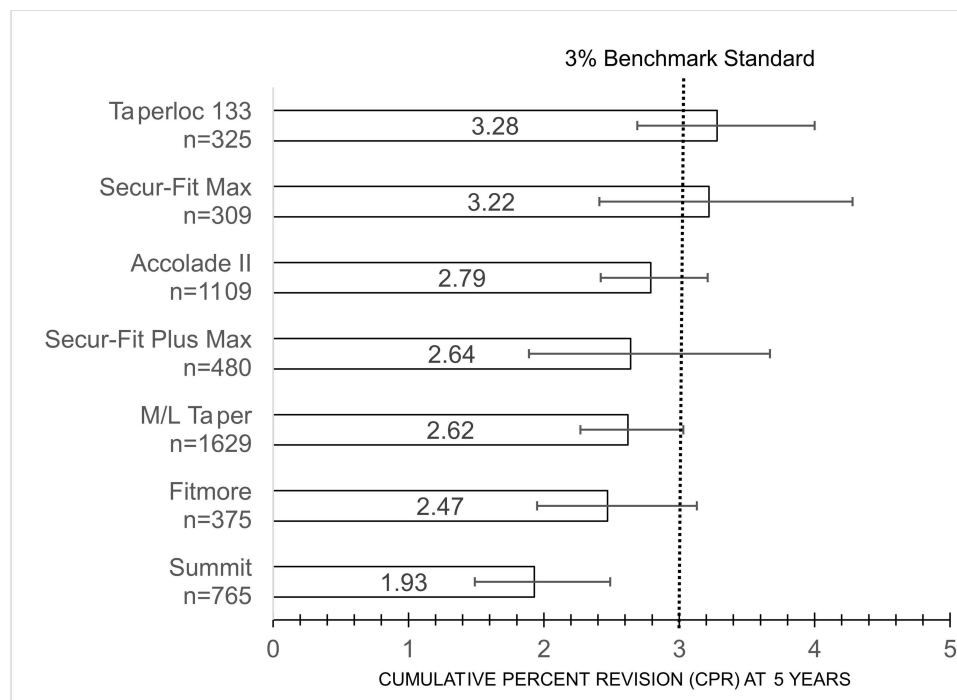


Figure 4 Benchmarking femoral stems at 5-year time point.

The working group proposed a superiority approach at 10 years, which is a more definitive statement that an implant performs well. In contrast, the group's proposal

for earlier benchmarks gives a benchmark by default, and it is only withheld if the implant proves to be inferior with respect to the 2- and 5-year pre-determined criteria of 2%

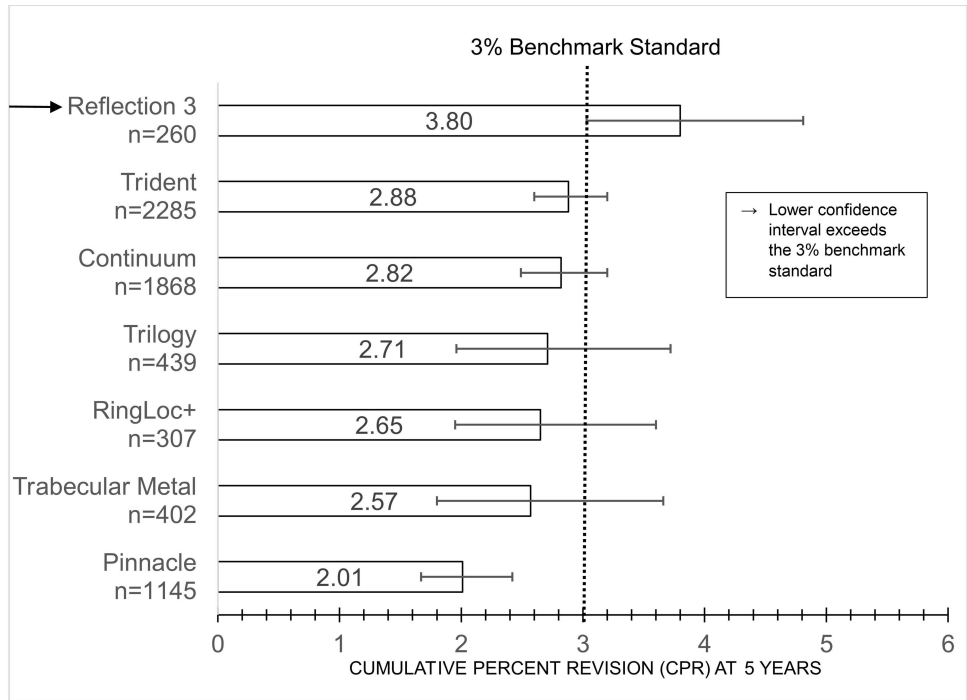


Figure 5 Benchmarking acetabular cups at 5-year time point.

and 3%, respectively. This approach may allow a mediocre product to initially be portrayed as an acceptable product. Differences between the two approaches at the early time points and the 10 year time point appear to be a compromise between the competing interests of innovation and public health.

Table 4 Characteristics of Stem/Cup Combination Analysis at 2 and 5 Years by Gender

Device Combinations	Gender	2 Year Time Point		5 Year Time Point	
		No. at Risk	CPR Estimate (95% Conf. Interval)	No. at Risk	CPR Estimate (95% Conf. Interval)
AML/Pinnacle	Female	154	2.06 (0.84, 4.98)		
	Male	107	2.11 (0.78, 5.62)		
Accolade II/Restoration ADM	Female	161	0.57 (0.08, 4.01)		
	Male	155	2.74 (1.14, 6.48)		
Accolade II/Trident	Female	5032	1.68 (1.41, 1.98)	586	2.80 (2.33, 3.37)
	Male	4146	1.59 (1.31, 1.94)	464	2.56 (2.01, 3.26)
Accolade TMZF/Trident	Female	456	2.06 (1.11, 3.79)		
	Male	368	1.29 (0.54, 3.07)		
Anthology/Reflection 3	Female	683	2.61 (1.80, 3.77)		
	Male	587	3.35 (2.32, 4.84)		
Avenir Muller/Continuum	Female	225	1.85 (0.77, 4.39)		
	Male	176	1.00 (0.25, 3.94)		
Citation TMZF/Trident	Female	154	1.89 (0.61, 5.74)		
	Male	140	0.68 (0.10, 4.76)		

(Continued)

Table 4 (Continued).

Device Combinations	Gender	2 Year Time Point		5 Year Time Point	
		No. at Risk	CPR Estimate (95% Conf. Interval)	No. at Risk	CPR Estimate (95% Conf. Interval)
Corail/Pinnacle	Female	666	1.56 (0.96, 2.52)		
	Male	466	1.85 (1.10, 3.13)		
Echo Bi-Metric/Regenerex RingLoc+	Female	143	2.03 (0.66, 6.15)		
	Male	108	3.54 (1.34, 9.16)		
Fitmore/Continuum	Female	869	1.93 (1.26, 2.94)		
	Male	921	1.42 (0.91, 2.23)		
Fitmore/Trabecular Metal	Female	178	1.10 (0.28, 4.31)		
	Male	132	1.48 (0.37, 5.79)		
M/L Taper/Continuum	Female	2479	2.00 (1.56, 2.56)	595	2.79 (2.20, 3.53)
	Male	2121	1.58 (1.16, 2.14)	577	2.25 (1.69, 2.97)
M/L Taper/G7	Female	252	2.15 (1.29, 3.57)		
	Male	187	2.26 (1.11, 4.57)		
M/L Taper/Trabecular Metal	Female	236	2.71 (1.41, 5.16)		
	Male	204	1.45 (0.54, 3.83)		
M/L Taper/Trilogy	Female	570	2.56 (1.60, 4.09)	185	3.16 (2.04, 4.87)
	Male	526	1.52 (0.79, 2.90)	157	2.09 (1.13, 3.84)
Polarstem/Reflection 3	Female	131	2.86 (1.46, 5.55)		
	Male	151	1.45 (0.69, 3.04)		
SROM/Pinnacle	Female	361	1.96 (0.98, 3.88)		
	Male	339	2.21 (1.16, 4.21)		
Secur-Fit/Trident	Female	386	5.42 (3.82, 7.67)		
	Male	243	4.11 (2.53, 6.64)		
Secur-Fit Max/Trident	Female	662	2.34 (1.54, 3.56)	174	3.55 (2.41, 5.22)
	Male	624	2.51 (1.68, 3.73)	133	2.85 (1.87, 4.35)
Secur-Fit Plus Max/Trident	Female	733	1.52 (0.89, 2.61)	253	2.06 (1.18, 3.58)
	Male	717	2.54 (1.68, 3.84)	226	3.24 (2.15, 4.88)
Summit/Pinnacle	Female	2006	1.75 (1.32, 2.33)	410	2.52 (1.83, 3.46)
	Male	1515	1.05 (0.68, 1.60)	349	1.05 (0.68, 1.60)
Synergy/Reflection 3	Female	286	3.50 (2.08, 5.86)		
	Male	245	1.94 (0.93, 4.03)		
Taperloc 133/G7	Female	563	3.07 (2.26, 4.16)		
	Male	374	2.00 (1.33, 2.99)		
Taperloc 133/Regenerex RingLoc+	Female	189	3.37 (1.70, 6.63)		
	Male	159	2.50 (1.05, 5.91)		
Taperloc 133/RingLoc+	Female	573	2.54 (1.59, 4.06)		
	Male	396	2.05 (1.11, 3.78)		
Taperloc 133 Microplasty/G7	Female	326	2.03 (1.16, 3.53)		
	Male	406	1.19 (0.58, 2.44)		

(Continued)

Table 4 (Continued).

Device Combinations	Gender	2 Year Time Point		5 Year Time Point	
		No. at Risk	CPR Estimate (95% Conf. Interval)	No. at Risk	CPR Estimate (95% Conf. Interval)
Trabecular Metal/Continuum	Female	267	3.32 (1.85, 5.91)		
	Male	205	0.38 (0.05, 2.70)		
Tri-Lock BPS/Pinnacle	Female	662	1.17 (0.66, 2.06)		
	Male	491	0.53 (0.20, 1.42)		
Tri-Lock BPS/Trident	Female	219	0.80 (0.20, 3.19)		
	Male	37	2.50 (0.36, 16.45)		

Note: Bold items do not meet pre-determined standards.

Another obvious limitation of this work arises from the structure of MARCQI, which is limited to the state of Michigan. MARCQI does receive full abstraction on over 97% of all primary and revision total hips in the state and performs audits to ensure that all primary and revision surgeries are captured at each site. While MARCQI

identifies revision surgeries that occur in the state, it has no mechanism for finding revision cases performed outside Michigan. However, Etkin et al¹⁵ reported that only 4.1% of patients having primary THA or TKA migrate out of Michigan within 5 years based on Medicare claims data between 2004 and 2016. While this only represents the

Table 5 Characteristics of Stem/Cup Combination Analysis at 2 and 5 Years by Age Group.

Device Combinations	Age Group	2 Year Time Point		5 Year Time Point	
		No. at Risk	CPR Estimate (95% Conf Interval)	No. at Risk	CPR Estimate (95% Conf Interval)
AML/Pinnacle	Below 65 years	108	0.49 (0.07, 3.40)		
	65 years or Above	153	3.29 (1.63, 6.57)		
Accolade II/Restoration ADM	Below 65 years	153	2.19 (0.82, 5.73)		
	65 years or Above	163	1.07 (0.26, 4.24)		
Accolade II/Trident	Below 65 years	4547	1.45 (1.19, 1.77)	521	2.58 (2.08, 3.19)
	65 years or Above	4639	1.81 (1.53, 2.13)	530	2.79 (2.27, 3.42)
Accolade TMZF/Trident	Below 65 years	440	1.08 (0.45, 2.58)		
	65 years or Above	384	2.43 (1.31, 4.47)		
Anthology/Reflection 3	Below 65 years	602	2.65 (1.76, 3.98)		
	65 years or Above	668	3.20 (2.28, 4.49)		
Avenir Muller/Continuum	Below 65 years	177	1.97 (0.74, 5.17)		
	65 years or Above	224	1.12 (0.36, 3.42)		
Citation TMZF/Trident	Below 65 years	127	2.27 (0.74, 6.86)		
	65 years or Above	167	0.58 (0.08, 4.05)		
Corail/Pinnacle	Below 65 years	499	1.54 (0.87, 2.70)		
	65 years or Above	633	1.78 (1.13, 2.80)		
Echo Bi-Metric/Regenerex RingLoc+	Below 65 years	109	2.61 (0.85, 7.87)		
	65 years or Above	142	2.74 (1.04, 7.13)		
Fitmore/Continuum	Below 65 years	938	1.58 (1.02, 2.44)		
	65 years or Above	852	1.74 (1.13, 2.69)		

(Continued)

Table 5 (Continued).

Device Combinations	Age Group	2 Year Time Point		5 Year Time Point	
		No. at Risk	CPR Estimate (95% Conf Interval)	No. at Risk	CPR Estimate (95% Conf Interval)
Fitmore/Trabecular Metal	Below 65 years 65 years or Above	116 194	0.83 (0.12, 5.77) 1.51 (0.49, 4.61)		
M/L Taper/Continuum	Below 65 years 65 years or Above	2202 2398	1.72 (1.28, 2.29) 1.88 (1.45, 2.43)	551 621	2.31 (1.76, 3.04) 2.74 (2.15, 3.48)
M/L Taper/G7	Below 65 years 65 years or Above	179 260	2.23 (1.17, 4.24) 2.11 (1.21, 3.67)		
M/L Taper/Trabecular Metal	Below 65 years 65 years or Above	195 245	1.90 (0.86, 4.18) 2.34 (1.12, 4.85)		
M/L Taper/Trilogy	Below 65 years 65 years or Above	423 673	2.11 (1.14, 3.89) 2.04 (1.25, 3.31)	136 206	2.54 (1.38, 4.64) 2.70 (1.74, 4.17)
Polarstem/Reflection 3	Below 65 years 65 years or Above	141 141	0.76 (0.25, 2.37) 3.29 (1.89, 5.69)		
SROM/Pinnacle	Below 65 years 65 years or Above	407 293	1.89 (0.99, 3.61) 2.32 (1.17, 4.60)		
Secur-Fit/Trident	Below 65 years 65 years or Above	295 335	4.88 (3.20, 7.41) 4.86 (3.31, 7.11)		
Secur-Fit Max/Trident	Below 65 years 65 years or Above	571 715	1.89 (1.16, 3.07) 2.86 (2.00, 4.09)	138 169	2.75 (1.68, 4.48) 3.64 (2.56, 5.16)
Secur-Fit Plus Max/Trident	Below 65 years 65 years or Above	773 677	1.60 (0.95, 2.68) 2.50 (1.64, 3.81)	270 209	1.89 (1.16, 3.08) 3.53 (2.25, 5.51)
Summit/Pinnacle	Below 65 years 65 years or Above	1567 1955	1.36 (0.94, 1.97) 1.52 (1.12, 2.07)	337 422	2.02 (1.31, 3.12) 1.77 (1.30, 2.42)
Synergy/Reflection 3	Below 65 years 65 years or Above	200 331	2.14 (0.96, 4.73) 3.18 (1.92, 5.22)		
Taperloc 133/G7	Below 65 years 65 years or Above	489 448	2.15 (1.46, 3.16) 3.11 (2.26, 4.27)		
Taperloc 133/Regenerex RingLoc+	Below 65 years 65 years or Above	188 160	3.95 (2.07, 7.45) 1.88 (0.71, 4.92)		
Taperloc 133/RingLoc+	Below 65 years 65 years or Above	427 548	1.19 (0.53, 2.62) 3.19 (2.09, 4.86)		
Taperloc 133 Microplasty/G7	Below 65 years 65 years or Above	429 305	1.28 (0.67, 2.41) 2.02 (1.10, 3.71)		
Trabecular Metal/Continuum	Below 65 years 65 years or Above	196 276	2.53 (1.14, 5.55) 1.71 (0.77, 3.77)		
Tri-Lock BPS/Pinnacle	Below 65 years 65 years or Above	535 619	0.39 (0.12, 1.24) 1.29 (0.75, 2.22)		
Tri-Lock BPS/Trident	Below 65 years 65 years or Above	85 171	0.00 (0.00, 0.00) 1.50 (0.48, 4.61)		

Note: Bold items do not meet pre-determined standards.

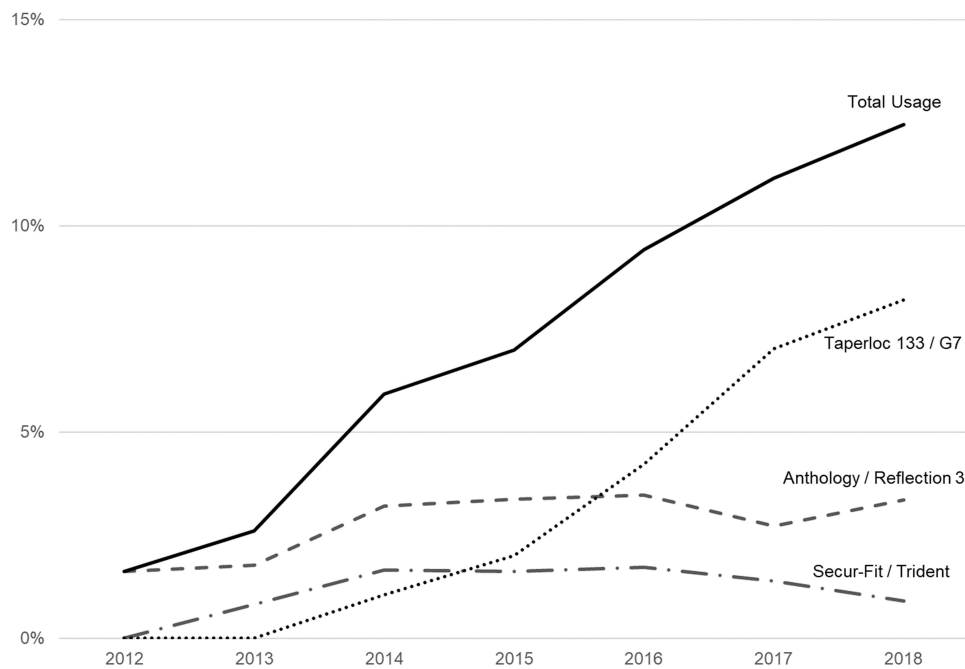


Figure 6 Percent of MARCQI total hip arthroplasty cases using implant combinations that would not receive an early (2 year) benchmark over time.

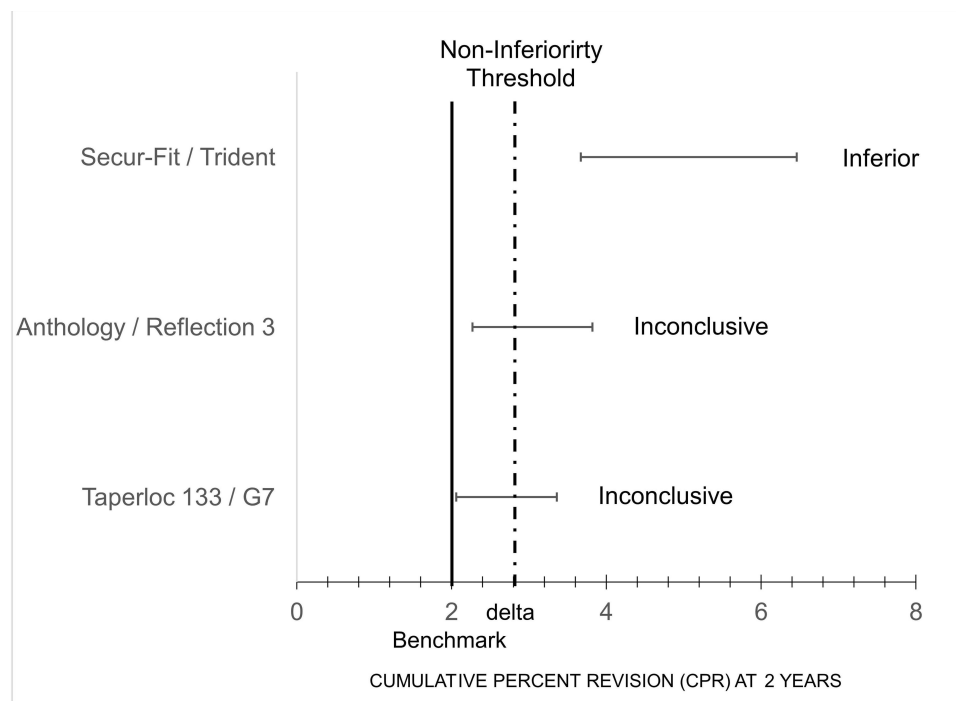


Figure 7 Non-inferiority analysis at 2-year time point.

over 65 year-old population, it suggests a low fraction of patients would be lost due to the inability to follow-up outside the state.

Despite its limitations, the International Prosthesis Benchmarking Working Group proposal has major strengths.

Among these strengths is the belief that the preferred data source for benchmarking is accurate and complete registry data. The combination of data from multiple sites in a registry environment allows benchmarking to be based on statistically significant numbers. This is an advantage over

analysis from the scientific literature where studies generally have small numbers, and those from the developers of an implant have better outcomes than demonstrated in national registry data.^{16,17} An additional strength of the proposal is that it was developed by a broad group of stakeholders around the globe. The adoption of a global methodology for benchmarking would serve to make benchmarks more transparent to payers, hospitals, surgeon, regulators, and patients. A single accepted methodology would also benefit implant manufacturers by reducing the cost of preparing and submitting data for benchmarking organizations and regulatory bodies. Such efficiencies would be advanced if additional methodology were developed to aggregate data from multiple registries into a single world-wide benchmark. However, accomplishing this would require adapting the benchmarking proposal to include sound meta-analysis methods for analyzing data from multiple sources. While sponsors, medical device manufacturers, registries and organizations currently involved in benchmarking were the intended audience, the group recognized their proposal would receive interest from additional stakeholders with the potential to be broadened for consideration in other joints as well. MARCQI's application of the benchmarking proposal reflects community use performance in real-world settings and hopes to strengthen the arthroplasty and scientific communities in registry involvement.

It is important to differentiate benchmarking from implant outlier detection. The two processes use different analytics, thresholds, and metrics.¹⁸ The most important difference may possibly lie in whether outlier surgeons and hospitals are analyzed. The benchmarking process does not control for confounding at the site or surgeon level. It is based on the analysis of the cumulative percent revision and number at risk at each benchmarking time point. The outlier detection model is based on component time incidence rate and allows registries to develop a standardized process in which to identify outliers and determine possible reasons for any difference, including device and non-device concerns. This opens the possibility that poor performance indicated by not receiving a benchmark at two or five years could be due to the implant performing poorly in the hands of only a few surgeons or a few sites. An additional difference between early benchmarking and outlier detection is that early benchmarking uses a non-inferiority analysis and outlier detection seeks to determine inferiority. Investigating the gap between benchmarking and outlier detection might prove useful for future implant performance detection.

Conclusion

The International Prosthesis Benchmarking Working Group protocol for benchmarking THA implants was found to be applicable to a regional arthroplasty registry in the United States. We found three implant combinations that did not perform sufficiently well to receive a benchmark at 2 years. Due to the fact that MARCQI is a young registry, we did not have sufficient numbers at risk at 5 years to conduct a benchmark assessment of these combinations. 8.6% of MARCQI cases were done with implant combinations that did not receive a 2-year benchmark. Moreover, the number of cases done with these non-benchmarked implant combinations is increasing over time in the state of Michigan. This presents a significant opportunity for quality improvement.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by Blue Cross and Blue Shield of Michigan and Blue Care Network as part of the BCBSM Value Partnerships program. Although Blue Cross Blue Shield of Michigan and the Michigan Arthroplasty Registry Collaborative Quality Initiative work collaboratively, the opinions, beliefs and viewpoints expressed by the author do not necessarily reflect the opinions, beliefs and viewpoints of BCBSM or any of its employees.

Disclosure

Heather A Chubb receives full salary support from Blue Cross Blue Shield of Michigan as a lead statistician in Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI). Brian R Hallstrom and Richard E Hughes receive partial salary support from Blue Cross Blue Shield of Michigan as co-directors of MARCQI. Eric Cornish declares that he has no conflicts of interest. None of the co-authors have financial relationships with the medical device industry.

References

1. Healthcare Cost and Utilization Project. HCUP fast stats - Most common operations during inpatient stays; 2019. Available from: <https://www.hcup-us.ahrq.gov/faststats/NationalProceduresServlet?year1=2014&characteristic1=0&included1=1&year2=&characteristic2=0&included2=1&expansionInfoState=hide&dataTablesState=hide&definitionsState=hide&exportState=hide>. Accessed January 28, 2019.
2. Hughes RE, Batra A, Hallstrom BR. Arthroplasty registries around the world: valuable sources of hip implant revision risk data. *Curr Rev Musculoskelet Med*. 2017;10(2):240–252. doi:10.1007/s12178-017-9408-5
3. Deere KC, Whitehouse MR, Porter M, Blom AW, Sayers A. Assessing the non-inferiority of prosthesis constructs used in total and unicompartmental 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;9(4):e026736. doi:10.1136/bmjopen-2018-026736
4. Sayers A, Crowther MJ, Judge A, Whitehouse MR, Blom AW. Determining the sample size required to establish whether a medical device is non-inferior to an external benchmark. *BMJ Open*. 2017;7(8):e015397. doi:10.1136/bmjopen-2016-015397
5. Orthopaedic Data Evaluation Panel; 2020. Available from: <http://www.odep.org.uk/>. Accessed May 20, 2021.
6. Tucker K. ODEP. The Parliamentary Review Web site; 2018–2019. Available from: <https://www.theparliamentaryreview.co.uk/organisations/odep>. Accessed June 9, 2020.
7. 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;25(2):131–137. doi:10.5301/hipint.5000209
8. International Prosthesis Benchmarking Working Group. Guidance document: hip and knee arthroplasty devices; May, 2018. Available from: <https://www.isarhome.org/publications>. Accessed April 22, 2020.
9. Australian Orthopaedic Association National Joint Replacement Registry. Australian orthopaedic association national joint replacement registry annual report: 2016 annual report. Available from: <https://aoanjrr.sahmri.com/documents/10180/275066/Hip%2C%20Knee%20%26%20Shoulder%20Arthroplasty>. Accessed June 1, 2020.
10. Hughes RE, Hallstrom BR, Cowen ME, Igrisan RM, Singal BM, Share DA. Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) as a model for regional registries in the United States. *Orthop Res Rev*. 2015;7:47–56. doi:10.2147/ORR.S82732
11. Hughes RE, Zheng H, Igrisan RM, Cowen ME, Markel DC, Hallstrom BR. The Michigan arthroplasty registry collaborative quality initiative experience: improving the quality of care in Michigan. *J Bone Joint Surg Am*. 2018;100(22):e143. doi:10.2106/JBJS.18.00239
12. Hughes RE, Hallstrom BR, Zheng T, Kabara J, Igrisan R, Cowen M. *Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) Report: 2012–2016*. Ann Arbor: Michigan Arthroplasty Registry Collaborative Quality Initiative; 2017.
13. Hughes RE, Zheng H, Hallstrom BR. *Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) Report: 2012–2017*. Ann Arbor: Michigan Arthroplasty Registry Collaborative Quality Initiative; 2018.
14. Hughes RE, Zheng H, Hallstrom BR. *2019 Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) Annual Report (Updated February 2020)*. Ann Arbor: Michigan Arthroplasty Registry Collaborative Quality Initiative; 2019.
15. Etkin CD, Lau EC, Watson HN, et al. What are the migration patterns for U.S. primary total joint arthroplasty patients? *Clin Orthop Relat Res*. 2019;477(6):1424–1431. doi:10.1097/CORR.0000000000000693
16. Labek G, Frischhut S, Schlichtherle R, Williams A, Thaler M. Outcome of the cementless Taperloc stem: a comprehensive literature review including arthroplasty register data. *Acta Orthop*. 2011;82(2):143–148. doi:10.3109/17453674.2011.570668
17. Labek G, Sekyra K, Pawelka W, Janda W, Stockl B. Outcome and reproducibility of data concerning the Oxford unicompartmental knee arthroplasty: a structured literature review including arthroplasty registry data. *Acta Orthop*. 2011;82(2):131–135. doi:10.3109/17453674.2011.566134
18. de Steiger RN, Miller LN, Davidson DC, Ryan P, Graves SE. Joint registry approach for identification of outlier prostheses. *Acta Orthop*. 2013;84(4):348–352. doi:10.3109/17453674.2013.831320

Orthopedic Research and Reviews

Dovepress

Publish your work in this journal

Orthopedic Research and Reviews is an international, peer-reviewed, open access journal that focusing on the patho-physiology of the musculoskeletal system, trauma, surgery and other corrective interventions to restore mobility and function. Advances in new technologies, materials, techniques and pharmacological agents are particularly

welcome. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/orthopedic-research-and-reviews-journal>