UPDATE 12/09/2017:

The new methodology for hips and knees is currently being tested by a number of companies and will be used in parallel with the current system this Autumn, 2017.

ODEP Benchmark Developments 2017

Introduction

Since ODEP was introduced in 2002, the system has been regularly reviewed and upgraded. Three years ago, NICE reviewed their original directive with resulting changes to the benchmarks and, about that time, knees were successfully introduced into the system. ODEP has always made every effort to match the aspirations of everyone involved whilst at the same time, having patients and their safety at the forefront of our thoughts.

We have consulted widely with all our stakeholders and we propose to introduce the changes to our methodology, as set out in this paper, in a staged manner during the course of the next year, including the 13 year benchmarks.

We also expect to introduce ODEP for shoulders shortly and work is being undertaken on revision hip implants.

The new proposed methodology is outlined below.

The revised ODEP benchmarking methodology for hips and knees

The advice we have received

ODEP has received considerable input from all our stakeholders and, as many will know, we have held open, international meetings for industry and all interested parties during the last three British Orthopaedic Association (BOA) congresses and previously, before the ISAR meeting in May 2016. We have listened intently to all the comments that were made.

We have been fortunate to be part of the International Benchmarking Working Group (IBWG) led by Professor Steve Graves and indeed, a member of our own statistics group, Dr Robert Scott, has been a member of the statistics section of the IBWG. We have taken note of their output, which, whilst not completed at this time has been most helpful.

ODEP has always taken statistical advice, when appropriate. For the first few years, this was from Dr Jan Van de Meulen and his colleagues in the Clinical Effectiveness Unit at the Royal College of Surgeons of England. More recently, we have secured the services of Dr Erica Cooke (statistician at Bedford University and RNOH), Professor Andy Judge (University of Oxford) and Dr Robert Scott (independent statistician). They have worked

with Andy Smallwood (founder of ODEP), Dr Phil Lewis, Professor Sion Glyn Jones, Dr Martin Pickford and other members of the panel.

Below are details of the changes (including comments about the areas where we do not anticipate a change) that we are expecting to introduce. It is anticipated the finally agreed changes will be implemented during 2017.
As always, they will be introduced gradually and benchmarks that have already been awarded will be honoured until the date when they will normally be due for renewal. With the advent of the 13 year benchmark it is anticipated all 10 year ratings will need to be upgraded to the 13 year benchmark within 4 years of receiving their 10 year benchmark.

Key changes "in a nutshell"
The headline differences between the present ODEP system for hips and knees and the future changes are:

- The acceptable revision rates, at each time point, will depend on the type of implant involved
- Generally, they will be 2% above the National Joint Registry for England, Wales and Northern Ireland (NJR) revision rate for that type of device at a particular benchmark
- Significant changes in the cohort size at each benchmark.
- A more standard use of Kaplan Meier (KM) with regard to censoring
- The adoption of the 13 year benchmark
- The increased availability of data has allowed us to review our thresholds

End Point
ODEP feels that the only workable outcome for benchmarking at present, is all cause revision. PROMS has been discussed but found to be unworkable, certainly for hips and knees. There is an argument for PROMS to be used with shoulder ODEP as revision is not always a reliable end point.

Data sources
ODEP accepts that good quality registry data has a pre-eminent place in a list of data sources but not the sole source. At the same time, particularly early on in the life of an implant the use of well organised trials, RCTS etc, are encouraged, with the resulting data to be incorporated into benchmarking. It was noted that some implants could easily be developed in a part of the world where there is not yet an effective registry and that data should not be lost, such data might well be generated in China or India. We feel that collecting data from different countries, besides different units in the same country, added significantly to the richness of benchmarking, particularly for products that are being used worldwide. It has been suggested that bias that might occur in publications from originating centres but ODEP will only grant A and A* benchmarks when there is data from at least 3 centres.

ODEP has always accepted that manufacturers can aggregate data from more than one source to produce a cohort size for a particular benchmark, but using data from more than one source, where in fact the patients are the same, is not acceptable.

ODEP supports the practice of manufacturers obtaining data from multiple sources and where it comes from across national boundaries it adds a richness to the submission.

ODEP would like to encourage all the world registries to either allow manufacturers access to their 'raw' but anonymised data or tailored data, for ODEP purposes.

Combinations

ODEP has always benchmarked acetabular components separately from stems whilst benchmarking knees as one unit. ODEP agrees with the idea that it would certainly be possible to benchmark hip combinations but asks where would this end. Mix and match and the use of different materials in the bearing surface would make the benchmark more complicated than was probably worthwhile for the effort that would be required to make any sense of the data. Clearly, it is important that a cup will work with a stem and vice versa. Our recommendation is that the present policy of monitoring stems and cups separately should continue. We do take note if a stem is being assessed and the revisions are mainly due to cup failure and, vice versa, but it must be said that failure of a cup can contribute significantly to failure of a stem and vice versa. Thus, it is generally unwise to allow blame on one side of a hip to be responsible for high revision rates in the other side and so allowing a high rating on the "good" side. Future plans might include the monitoring of different bearing surfaces but this is certainly not a priority project.

This approach does not apply to knees which will be benchmarked as a whole including the femoral and tibial components.

Setting the standard for each benchmark

ODEP suggests there are three main issues to review; the timing, the standard required, and the method of calculation.

Up until now. we have tried to apply the same performance standard to all types of implant. This approach has considerable attraction but can cause difficulties with some implants such as unicondylar knee replacements, patella-femoral replacements and shoulder replacements. ODEP has decided that now is a time for change and that acceptable revision rates for each benchmark will be tailored to the type of device in question.

Timing

To a considerable extent, numeric timelines are based on worldwide experience of the historic failure of devices but they are also based on the ODEP philosophy of encouraging manufacturers to look at the data about their implants on a regular basis.

We did consider the 2 year rating but within the system of healthcare in the UK there are delays leading to revision being performed later than might be ideal. After discussing the issue with colleagues in Canada, New Zealand, Netherlands, Germany, Italy, Denmark and Spain we realise that the UK experience is almost certainly mirrored by a lot of healthcare systems throughout the world.

Thus, ODEP recommends that after the pre-entry benchmarks the 3, 5, 7 and 10 year time points should remain in force and the 13 year benchmark be added during 2017.

Loss to follow up
We realise that most of the data that will be submitted to ODEP will be from Registries where loss to follow-up does not apply. Where the data source is a non-registry study such as a publication, RCT podium presentation etc. we expect a loss to follow-up of less than 20% of the cohort.

Death rate
ODEP realises that it is unlikely that any of the prostheses that are submitted to ODEP will have the ability to cause death. However, ODEP would be concerned when the death rate at the 10 year benchmark is more than 35% (or a corresponding mortality rate at the other benchmarks). We acknowledge that the average age of the cohort, patient's co-morbidities and local factors have far more influence on death than the actual implant. It is also acknowledged that the death rate in patients having a total hip replacement (THR) in the treatment of their femoral neck fracture is higher than in patients undergoing THR for arthritis.

ODEP would expect a manufacturer who found that the death rate in the cohort of patients being submitted to be higher than expected, to verify their data from the data source and produce an explanation for this observation.

Number of centres and number of surgeons
ODEP do not propose any change to their present guidelines.

It has been argued, for a variety of reasons, that the first surgeons using a product will be better surgeons and would skew the results. It has also been suggested that, for personal reasons, they might be reluctant to revise an implant when revision might well be indicated. It can equally be argued that the early surgeons do not have the benefit of experience and initially they may have to cope with under designed instruments, lack of size options etc.

Thus, it can be said that data from these surgeons might skew the front end data for a benchmarking study either way and ODEP's view is that the front end data should always be considered in the context of the data coming up behind it.

ODEP's view is that, in an ODEP submission, there should always be data from more than 3 centres beyond the developing centre(s). It is nowadays quite common that besides the 'pioneer' surgeon, manufacturers may employ advising surgeons to help them with design/launch etc. ODEP's view is that, if these people are in receipt of royalties from the sale of the product, they will be working in developing centres.

Calculation of revision rates
ODEP agrees that Kaplan Meier (KM) survival analysis is an almost universally used system for predicting outcomes. It is used by most registries and revision based clinical studies. Many observers have suggested that competing risk adjustment answers the

wrong question and that unadjusted KM estimates of implant failure are more clinically meaningful and straightforward to interpret (Ranstam et al. 2011b).

KM confidence Interval

There are many ways to calculate confidence intervals in KM analysis. A common method in statistical software is the 'Greenwood formula' with a log-log transform (e.g. this is the default in SAS v9.2 onwards). This can give excessively narrow confidence intervals where there is heavy censoring between the time of the last revision and the end of the survival curve. Alternative methods have been suggested for calculating confidence intervals based on the number of implants still at risk at the time of interest, not the number at the time of the last revision (Ranstam et al. 2011a). However, these are not universally available in software and the problem of tail censoring can be mitigated by applying a minimum number at risk at the benchmark time. ODEP therefore recommends the 'Greenwood log-log' approach.

Adjustment for patient variables

It could be argued that some implant brands are preferentially used in high or low risk patients (e.g. younger or older than average) and, in order to make a fair assessment against a benchmark, it is necessary to adjust for patient variables. Although that may be desirable, ODEP's view is this would be difficult to apply in a consistent, transparent way across different sources of data and, in practice, is rarely a significant factor in benchmarking.

ODEP feels it will always be important for manufacturers to submit comprehensive data to include pre-operative demography and diagnostic details together with all reasons for revision, death etc. It is up to the Panel to interpret the data that is submitted and apply their clinical experience where necessary.

Acceptable revision rates

This is a change in the latest methodology that will have the most impact. The original National Institute for Health and Care Excellence (NICE) guidelines for ODEP were that a hip replacement could only be regarded as acceptable, for patients in the UK, if it had a revision rate of less than 10% at 10 years. An entry benchmark at 3 years with a correspondingly less high revision rate was also recommended by NICE.

Shortly after being established, ODEP introduced the 5 and 7 year benchmarks and stressed that an implant could not "sit still" at any of the benchmarks under 10 years but should progress. In fact, it was decided that the maximum failure rates at a particular benchmark would be numerically the same as the "year". At that time, NICE separated surface replacements and standard hips.

It should be noted that the introduction of ODEP for knees was not initiated by NICE guidelines. The introduction of ODEP for knees was encouraged by the British Association for Surgery of the Knee (BASK) and industry, and was subsequently entirely organised by ODEP. ODEP for shoulders, which is presently in the testing stage, has also not originated from NICE guidance.

Thus, the early benchmarks were decided on a pragmatic basis and have served the orthopaedic community well since their implementation. They have the advantage of being simple but effective enough to make sure implants come with data to support their continuing use.

In February 2014, the original NICE TA2 guidelines were superseded by TA304. One of the main changes to the guidelines was a reduction in the maximum revision rate at 10 years from 10% to 5%. This decision was based upon the fact that the majority of hip submissions at 10 years reported a revision rate of <5%, with most of the supporting data coming from registries. Thus, the ODEP A* rating was introduced, requiring a revision rate of less than 5% at 10 years. The 10% threshold for a 10A rating was kept for backwards compatibility.

Since the revision rate of total knee replacement (TKR) is similar to that of total hips (excluding metal on metal), the same criteria were applied for submissions for benchmarks for knees.

The 2% cushion

Following an in-depth analysis of registry data for both hips and knees, we have found that the 5% maximum revision rate at 10 years is approximately 1.5%-2% higher than the overall revision rate for the majority of total knees and for non metal-on-metal hips (surface replacement of the hip and unicondylar knee replacement has also been excluded for the present). We are cognisant that some registries report overall higher revision rates than others but we feel that the 2% cushion is a workable way forward.

Thus application of a 5% threshold for a 10 year benchmark effectively applies a 2% non-inferiority margin to the observed average revision rate from real world data. By using a non-inferiority comparison we seek to demonstrate that an implant brand is not unacceptably worse (by a defined margin) than the average performance of its class.

It is therefore proposed that for an A* rating the maximum revision rate is based on a 2% margin above the observed revision rate at that time point.

In addition to specifying an acceptable revision rate for each time point, it is obviously necessary to take account of the confidence interval surrounding the observed revision rate estimate for the brand. To give a reasonable degree of confidence that the underlying revision rate for the brand is less than the benchmark, it is proposed that the 95% confidence interval for the Kaplan Meier revision rate lies below the specified level.

Criteria for the other benchmark times have been calculated in a similar way. As previously noted, the maximum demonstrated revision rate at 10 years of 5% is approximately 2% above the average revision rate for all non metal-on-metal hip replacements in the NJR at 10 years. A similar margin of approximately 2% has been applied to the revision rate at earlier time points. Again, this is effectively a 2% non inferiority margin proposal and the proposed criteria are tabulated in the next section.

Sample size requirement
There are three options for setting a minimum number of implants for a particular benchmark:

1. Total cohort size, i.e. the number of implants in the whole data set
2. Total number of implants surviving at the benchmark time i.e. the patient is still alive and original implant still in place. In KM analysis, this is the "number at risk" at the benchmark time
3. Total number of devices implanted more than "n" years ago, where "n" is the benchmark time. i.e. for a 10 year benchmark for implants with follow up to the end of December 31st 2016, this is the number implanted in 2006 or before

The ODEP submission grid or rating criteria table (vide infra) includes minimum number requirements for 1 and 3 years, though all three are reported on the ODEP clinical data sheets. The differences between these will become more critical with benchmarks past 10 years when the number of deaths will become much more significant.

The introduction of a minimum number past the benchmark was introduced with the A* ratings, since several submissions were based on KM analysis in which only a very few implants had actually been in the patient for the prescribed length of time and this was thought to be unsatisfactory.

The minimum number required to be submitted for a particular cohort is a compromise, as with any clinical trial involving a new medical product or procedure. Numbers should be sufficient to give an appropriate degree of assurance that the product has adequate performance whilst a precise estimate of performance with a high degree of certainty inevitably requires a large number, which means exposing a large number of patients to the new product.

This compromise is particularly acute at the 3 year benchmark where our statistics colleagues would perhaps like to have a large number of patients in the cohort to analyse. They make the point that they can have little confidence in the data as it is immature, large numbers would increase their confidence. Many manufacturers would find it impossible to put together a cohort that met the demands of the 'stats' people within a reasonable time period.

One can look at this dichotomy of views in a different way and ask the question "what are we trying to achieve with an early benchmark"? It could be said that, early on, all we really can assess is whether the procedure is possible and little more. The early benchmarks can therefore be looked upon as evidence that there is no early catastrophic failure and the later benchmarks will be the ones that sort out the best and identify failure patterns of a more subtle nature.

The numbers at each benchmark
Until now, patients who have died by each of the ODEP benchmarks have been included in the cohort. This does of course mean that the patient has avoided a revision between the index operation and their demise but this arrangement does not

fit well with the supporters of the Kaplan Meier estimate for revision. Thus, ODEP is willing to consider making a minor change to the criteria for achieving a benchmark. Under the modified criteria, the Kaplan Meier survival analysis will be required to demonstrate (with 95% confidence) that the underlying revision rate for a brand is better than a specified level. For the 10A* rating, this level has been set by NICE at 5% revision at 10 years.

Thus a minimum number of patients with more than 10 years follow up will be retained, but to allow for losses due to death, the minimum number at risk at 10 years will be set at 400. For consistency with existing criteria at the start, the total cohort number must be at least 500 patients, though in practice a current implant with more than 400 at risk at 10 years will generally have many more than this size cohort.

The 13 year benchmark

ODEP considers it essential to develop the 13 year benchmark by the end of 2017 with a call for submissions within the next 3-6 months.

It would be inconsistent with ODEP's philosophy to have readily available 13 year data in registries and individual studies without using it and giving patients, surgeons and manufacturers it's obvious benefits. Several of the world's renowned registries have data well past 13 years and several more at the brink of this milestone.

ODEP believes that it befits anyone undertaking benchmarking that they should keep pace with the maturity of registries.

Already several manufacturers have asked when ODEP will be accepting submissions for the next benchmark.

ODEP has decided on 13 years for several reasons notwithstanding the maturity of NJR, after then, it will probably be 15 years but that is undecided at present. ODEP is aware that there will be a higher death rate at the later mark. It is suggested

- The acceptable revision rate with 95% confidence for an A* rating will be 6.5% and there should be 400 implants in the cohort
- For an A rating the revision rate will be < 8.5%

The A* rating

These criteria are summarised for the A* rating in the following tables:

Criteria – total hip replacement	3A*	5A*	7A*	10A*	13A*
Minimum number of centres outside development centre(s)	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	150	225	300	400	400
Maximum revision rate ‡	3.0%	3.5%	4.0%	5.0%	6.5%

‡ The upper 95% confidence interval for KM revision rate (1 - Survival) must be lower than the specified level.

Criteria – total knee replacement	3A*	5A*	7A*	10A*	13A*
Minimum number of centres outside originating centre(s)	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	150	225	300	400	400
Maximum revision rate ‡	3.5%	4.0%	4.5%	5.0%	6.0%

‡ The upper 95% confidence interval for KM revision rate (1 - Survival) must be lower than the specified level.

The A rating

Until now, the less demanding A rating was kept so that unicondylar knee replacements, surface replacement of hips and some other implants could be awarded a rating. Some of our partners needed the simple no star ratings. In this new and evolving methodology the criteria for an A rating are calculated in a similar way to the A*, but with a wider non-inferiority margin and smaller minimum numbers, reflecting a lower level of clinical evidence. The 1.5%-2% margin has been increased to 3.5%-4%, and the minimum numbers reduced.

The required number past the benchmark has been reduced to the bare minimum necessary to avoid anomalies where there are zero revisions. It has been calculated so that, if there are zero revisions, the upper confidence limit for the binomial exact confidence interval for the proportion revised will be lower than the benchmark maximum revision rate. This is necessary since the standard Greenwood method cannot calculate a KM confidence interval where there are zero revisions.

Criteria - total hip replacement	3A	5A	7A	10A	13A
Minimum number of centres	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	72	66	60	51	42
Maximum revision rate ‡	5.0%	5.5%	6.0%	7.0%	8.5%

Criteria - total knee replacement	3A	5A	7A	10A	13A
Minimum number of centres	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	66	60	55	51	45
Maximum revision rate ‡	5.5%	6.0%	6.5%	7.0%	8.0%

The B rating

ODEP has always seen fit to make a benchmark available for implants which are not bespoke but have a small market. A special type of hip replacement for a DDH (developmental dysplasia of the hip) patient would be a good example.

Thus, for boutique implants the following grid will be available. For these implants, it makes sense to take the opposite approach to the A* and A ratings and award a B rating to those implants which do not have demonstrably poor performance. For this, any implant with more than minimal usage and for which the lower confidence interval for the KM survival is no higher than a benchmark level, will be eligible for a B rating.

Criteria - total hip replacement	3B	5B	7B	10B	13B
Minimum number of centres	1	1	1	1	1
Minimum total cohort	100	100	100	100	100
Minimum at risk at benchmark time	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%

Criteria - total knee replacement	3B	5B	7B	10B	13B
Minimum number of centres	1	1	1	1	1
Minimum total cohort	100	100	100	100	100
Minimum at risk at benchmark time	40	40	40	40	40
Maximum value of 95% lower confidence limit for revision rate	3.5%	4.0%	4.5%	5.0%	6.0%

Example ratings for hips

Mark applied for	Total Cohort	Number at risk	Revision rate 95% CI	Rating	Reason
3	640	220	0.48% - 2.41%	3A*	Meets all benchmark criteria
10	4000	470	1.75% - 3.50%	10A*	Meets all benchmark criteria
10	5400	630	3.52% - 5.05%	10A	Upper CI > 5% but < 7%
10	1200	220	1.04% - 3.03%	10A	Insufficient numbers for A*
5	500	350	3.07% - 7.27%	5B	Upper CI > 5.5% but Lower CI < 3.5%
3	250	110	3.38% - 10.24%	U	Lower CI > 3.0%

NOTE: for the hip and knee draft submission grids please see appendix 1 and 2 respectively

The management of implants other than hips and knees

ODEP's view is that it would be best if the standard ODEP process could be applied to all joints and in all situations. Patients, surgeons, procurement officers and manufacturers all round the world are familiar with ODEP ratings and what they represent.

Unicondylar knee replacements

Up until now, it has been very difficult for unicondylar knee replacements to achieve a rating. It is generally agreed that these devices do have a place in modern day orthopaedics and although they generally have a significantly higher revision rate this is balanced by a lower death and complication rate together with faster rehabilitation.

With the new methodology they would be able to submit data with revision rates of 2% above the average revision rate at the standard time points. The revision rate has been calculated from the average revision rate for the unicondylar knees that are being marketed at present (2017). With time, it is likely that the revision rate will fall and effectively this will 'raise the bar'.

This approach does not take note of surgeons doing small numbers of procedures or extension of disease. Please see appendix 3 for a probable grid

Concluding remarks - methodology for ODEP hips and knees

After nearly 15 years of experience, ODEP is delighted to present its latest ideas for progressing and improving the process.

We are always open to suggestion but any change must be for the better. Changes can often lead to confusion and there will always have to be a significant benefit to everyone involved with the use of ODEP, to make a change worthwhile. In the quest for uniformity between ODEP and any other international benchmarking group, ODEP is prepared to consider compromise in some areas of their process provided they consider the change to be for the benefit of patients and equally fair for all the manufacturers.

In regard to further development, ODEP consider it essential that if changes are made then there should be a period of time for the effect of the change(s) to be noted.

Revision Hip

Introduction

Approximately nine months ago, ODEP decided to scope benchmarking for revision hip. ODEP suspected that the relatively high incidence of revision for infection would skew the results and make assessment of the implant very difficult but, so far, have found this not to be the case. Preliminary work has shown that by using NJR data, meaningful outcomes can be obtained.

It was realised that a lot of implants can be used both in the primary and revision setting but it is the implants that are primarily used for revision that are not being assessed at present. Certainly, there is a lot of heterogeneity in the data.

ODEP intend to take the scoping process further with hips and then look at knees.

Is there any point in undertaking these benchmarks?

- Not if everyone thinks it is a waste of time!
- It will mean that data is collected and analysed for revision products and this is not being done in a systematic way at present
- Surgeons will have a better idea of what works and what does not
- Some of these products might be harmful

The data now available is significant. Just in the NJR there are large numbers to analyse although small numbers are also helpful. Sometimes there is a good reason for a particular brand being unpopular!

Summary
We think there is now sufficient data for worthwhile development of an ODEP system for revision hip replacement. Manufacturers should be able to obtain meaningful data from more than one registry.

Shoulders

Introduction

Approximately two years ago, representatives from the British Elbow and Shoulder Society (BESS) approached ODEP to scope an ODEP service for shoulders. Interestingly, they cited the same argument that was used to initiate ODEP for hips and knees, that they wanted a system in place that encouraged manufacturers to look at their data and measure the performance of implants against benchmark standards. Subsequently, ODEP and BESS met with the Association of British Healthcare Industries (ABHI) to agree that a scoping exercise should be undertaken. The development of ODEP for shoulders has moved on and has benefitted from the involvement of all parties

The table in Appendix 4 reflects what is currently being used and under testing. In view of the proposed changes in ODEP's methodology, it is likely this will significantly change over the next year but to get the system up and running, it is a start. The first submissions were received from several manufacturers in March this year and a 'pilot' review meeting was held on 6th April to actively assess the submissions received. On the whole, the submissions were very encouraging although ODEP only reviewed submissions for reverse shoulder replacements. The issues were as expected and some details are mentioned in more detail in this document although it should be noted that there was a limited amount of PROMS data available.

Revision as a discriminator

Whereas with hips and knees, revision is generally accepted as an acceptable end point, with shoulder replacement this is not always true. Some patients whose shoulder replacement has failed in terms of useful function are, in fact, left in the patient and not removed. Thus the trigger "revision" is not available to the benchmarker. BESS have been exploring the use of PROMS data to discriminate between the good and the bad.

The options

There are a large number of different types of shoulder prostheses. Decisions need to be made as to how granular any analysis should be. It is a balance between numbers and confidence intervals. This work is ongoing and clearly the breakdown into different types and attributes will, where possible, need to fit well with the way registries are collecting data.

Data availability

The registries rich in shoulder data include those from Australia, New Zealand, UK and some of the Scandinavian registries.

Development plan
ODEP does not envisage publishing any data on shoulders until all the stakeholders involved are confident that ODEP can deliver a robust, transparent and reliable system. ODEP will continue to invite manufacturers to submit their shoulder data and they (and only they) will get ODEP's view until all parties are agreeable for it to be published on the ODEP website.

Acknowledgements

We are particularly grateful to everyone who helped us with these developments including:

- Members of the "International Benchmarking Working Group"
- Members of EPRD particularly Oliver Melsheimer and Alexander Grimberg
- Members of LROI including Professor Rob Nelissen, Paul Bom, Taco Gosens, Gijs Van Hellemondt and Reinoud Brouwer
- Professor Andy Judge (Statistician) from Oxford University
- Erica Cooke (Statistician) from Bedford University

They have all contributed to a greater or lesser degree in one way or another.

Final thoughts

ODEP is busy! It continues to evolve and it will always welcome comment and advice. ODEP meetings are open to stakeholders who do not have any conflict of interest, fortunately it is run by a dedicated team of people many of whom are volunteers who are completely committed to its further development.

We will continue to welcome comments and advice from all who want to be in touch. Whether we wished it or not we have become a global influence and we take that responsibility very seriously. Below is a one month snapshot of the hits we receive on our website. We don't know much about Greenland!

Country	Sessions
United Kingdom	692
United States	139
Unknown	108
Netherlands	101
India	69
Switzerland	62
Germany	59
Italy	59
Japan	45
Ireland	38

Phil Lewis
Robert Scott
Richard Armstrong
Martin Pickford
Andy Smallwood
Tim Wilton
Keith Tucker (Chair)
Spring 2017

Ranstam, J. et al., 2011a. Statistical analysis of arthroplasty data. I. Introduction and background. *Acta orthopaedica*, 82(3), pp.253–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3235301&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2014].

Ranstam, J. et al., 2011b. Statistical analysis of arthroplasty data. II. Guidelines. *Acta orthopaedica*, 82(3), pp.258–67. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3235302&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2014].

Appendix 1

Draft ODEP submission Hip Grid May 2017

Criteria - Total Hip Replacement					
Criteria - A* Ratings	**3A***	**5A***	**7A***	**10A***	**13A***
Minimum number of centres outside development centre(s)	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	150	225	300	400	400
Maximum revision rate ‡	3.0%	3.5%	4.0%	5.0%	6.5%
Criteria - A Ratings	**3A**	**5A**	**7A**	**10A**	**13A**
Minimum number of centres	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	72	66	60	51	42
Maximum revision rate ‡	5.0%	5.5%	6.0%	7.0%	8.5%
‡ The upper 95% confidence interval for KM revision rate (1 - Survival) must be lower than the specified level					
Criteria - B Ratings	**3B**	**5B**	**7B**	**10B**	**13B**
Minimum number of centres	1	1	1	1	1
Minimum total cohort	100	100	100	100	100
Minimum at risk at benchmark time	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%

Appendix 2

Draft ODEP Submission Knee Grid 2017

Criteria - Total Knee Replacement					
Criteria - A* Ratings	**3A***	**5A***	**7A***	**10A***	**13A***
Minimum number of centres outside development centre(s)	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	150	225	300	400	400
Maximum revision rate ‡	3.5%	4.0%	4.5%	5.0%	6.0%
Criteria - A Ratings	**3A**	**5A**	**7A**	**10A**	**13A**
Minimum number of centres	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	66	60	55	51	45
Maximum revision rate ‡	5.5%	6.0%	6.5%	7.0%	8.0%
‡ The upper 95% confidence interval for KM revision rate (1 - Survival) must be lower than the specified level					
Criteria - B Ratings	**3B**	**5B**	**7B**	**10B**	**13B**
Minimum number of centres	1	1	1	1	1
Minimum total cohort	100	100	100	100	100
Minimum at risk at benchmark time	40	40	40	40	40
Maximum value of 95% lower confidence limit for revision rate	3.5%	4.0%	4.5%	5.0%	6.0%

Appendix 3

Draft ODEP submission Unicondylar Knee Grid 2017

Criteria - Unicondylar Knee Replacement					
Criteria - A* Ratings	**3A***	**5A***	**7A***	**10A***	**13A***
Minimum number of centres outside development centre(s)	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	150	225	300	400	400
Maximum revision rate ‡	6.0%	8.0%	10.0%	13.0%	17.5%
Criteria - A Ratings	**3A**	**5A**	**7A**	**10A**	**13A**
Minimum number of centres	3	3	3	3	3
Minimum total cohort	150	250	350	500	500
Minimum at risk at benchmark time	45	40	40	40	40
Maximum revision rate ‡	8.0%	10.0%	12.0%	15.0%	20.0%
‡ The upper 95% confidence interval for KM revision rate (1 - Survival) must be lower than the specified level					
Criteria - B Ratings	**3B**	**5B**	**7B**	**10B**	**13B**
Minimum number of centres	1	1	1	1	1
Minimum total cohort	100	100	100	100	100
Minimum at risk at benchmark time	40	40	40	40	40
Maximum value of 95% lower confidence limit for revision rate	6.0%	8.0%	10.0%	13.0%	17.5%

Appendix 4

Shoulders

Early version of ODEP for shoulders submission grid (this will be amended in view of the likely changes in ODEP's methodology as laid out above).

A	B	C	D	E
Pre-entry	**3 years**	**5 years**	**7 years**	**10 years**
Pre-entry A* Product launched under Beyond Compliance	**3A* rating** 50 implants (with data from beyond the developing centre submitted) with a minimum of 3 years follow up with actual revision rates of less than 3%. All deaths, loss to follow up, failures and indications for revisions recorded. Pre-op and interval PROMS data.	**5A* rating** 100 implants (with data from beyond the developing centre submitted) with a minimum of 5 years follow up with actual revision rates of less than 5%. All deaths, loss to follow up, failures and indications for revisions recorded. Pre-op and interval PROMS data.	**7A* rating** 150 implants (with data from beyond the developing centre submitted) with a minimum of seven years follow up with actual revision rates of less than 5%. All deaths, loss to follow up, failures and indications for revisions recorded. Pre-op and interval PROMS data.	**10A* rating** 200 implants (including three centres in cohort & including data from beyond the developing centres) with a minimum of ten years follow up with a with actual revision rates of less than 5 % at 10 years i.e. demonstrating survivors implant of better than 95%. All deaths, loss to follow up failures and indications for revision included in data
Pre-entry Products registered with NJR. All primaries and revisions monitored via supplier feedback.	**3A rating** 50 implants (with data from beyond the developing centre submitted) with minimum three years follow up demonstrating less than 3% revision rates at three years, with Kaplan - Meier survivors implant data showing	**5A rating** 100 implants (with data from beyond the developing centre submitted) with minimum five years follow up demonstrating less than 5% revision rates at three years, with Kaplan - Meier survivors implant data showing	**7A rating** 150 implants (with data from beyond the developing centre submitted) with minimum seven years follow up demonstrating less than 7% revision rates at three years, with Kaplan - Meier survivors implans data	**10A rating** 200 implants (with data from beyond the developing centre submitted) with minimum ten years follow up demonstrating less than 10% revision rates at three years, with Kaplan - Meier survivors implant data showing
	3B rating **mimimum 50** Data for a smaller cohort demonstrating less than 3% revision rates at three years, and PTIR or Kaplan-Meier survivorsimplants data showing confidence limits on the data	**5B rating** **mimimum 50** Data for a smaller cohort demonstrating less than 5% revision rates at five years, and PTIR or Kaplan-Meier survivorsimplants data showing confidence limits on the data	**7B rating** **minimum 50** Data for a smaller cohort demonstrating 7% at seven years, and PTIR or Kaplan-Meier survivorsimplants data showing confidence limits on the data	**10B rating** **minimum 50** Data for a smaller cohort demonstrating 10% at ten years, and PTIR or Kaplan-Meier survivorsimplants data showing confidence limits on the data