Quality of Datasets for Outcome Measurement

Quality of Publications regarding the Outcome of Revision Rate after Arthroplasty

Final Report of the QoLA* Project

Presented at the EFORT Congress 2011 in Copenhagen

Gerold Labek
On behalf of the QoLA* Study Group

*Quality of Literature in Arthroplasty

EAR
European Arthroplasty Register

The QoLA Research Group

Names in alphabetical order

J. Abel[1], M. Agreiter[1], D.N. Alegre[26], F. Arrich[1], A. Asparouhov[11], B. Atilla[3], E. Caceres-Palou[18], C. Berger[7], N. Böhler[13], E. Brabec[1], C. Delaunay[20], N.M. da Conceição Diogo[27], R. Dorninger[1], St. Frischhut[1], B. Haderer[1], F. Haselsteiner[1,13], C. Forster Horváth[5], J. Hassenpflug[2], R. Hochgatterer[13], P. Hoffmeyer[12], I. Hois[13], M. Hübl[1], L. Iovanescu[15], W. Janda[1], L. Kandel[9], D. Kendoff[6], C. Keurentjes[4], P. Kjaersgaard-Andersen[8], H. Klaus[1,25], M. Krismer[1], S. Krivanek[10], G. Labek[1], J. Lair[1], J. Leonhardt[8], M. Liebensteiner[1], T. Liebs[2], A. Lübbeke-Wolff[12], A. Matamalas Adrover[18], Y. Mattan[9], R. Mroz[2], L. Nečas[10], R. Nelissen[4], D. Neumann[16], C. Pabinger[22], W. Pawelka[1], K.-P. Pfeiffer[23], B. Pijls[4], R. Rauf[1], C. Ribeiro[26], R. Schlichtherle[1], R. Schuh[1], K. Sekyra[1], K. Specht[8], C.I. Stoica[17], B. Stöckl[24], M. Szendroi[5], A. Tarnau[15], M. Thaler[1], C. Thelu[19], S. Todorov[11], A. Williams[1], L. Zagra[14], A. Zahár[5,6], G. Ziernhöld[21]

1 Medical University of Innsbruck, Dept. of Orthopaedic Surgery, Innsbruck, Austria
2 University of Kiel, Dept. of Orthopaedic Surgery, Germany
3 Hacettepe University, Dept. of Orthopaedic Surgery, Turkey
4 University of Leiden, Dept. of Orthopaedic Surgery, The Netherlands
5 Semmelweis Univ. Budapest, Dept. of Orthopaedic Surgery, Hungary
6 Endoklinik, Hamburg, Germany
7 Sozialmedizinisches Zentrum Ost, Dept. of Orthopaedic Surgery, Vienna, Austria
8 South Danish University, Vejle, Denmark
9 Haddassah Univ. Hospital, Jewish University Jerusalem, Israel
10 University of Martin, Dept. of Orthopaedic Surgery, Martin, Slovakia
11 University of Pleven, Dept. of Orthopaedic Surgery, Pleven, Bulgaria
12 Geneva University Hospitals, Division of Orthopaedics and Trauma Surgery, Geneva, Switzerland
13 Allgemeines Krankenhaus Linz, Dept. of Orthopaedic Surgery, Linz, Austria
14 Istituto Galeazzi, Milano, Italy
15 University of Arad, Dept. of Orthopaedic Surgery, Arad, Romania
16 Medical University of Salzburg, Dept. of Orthopaedic Surgery, Salzburg, Austria
17 Clinica Foisor de Foc, Bucharest, Romania
18 Hospital del Mar, Dept. of Orthopaedic Surgery, Barcelona, Spain
19 University of Lille, Dept. of Orthopaedic Surgery, Lille, France
20 Clinique de l´Yvette, Paris, France
21 Dr. Günther Ziernhöld, Bolzano, Italy
22 Orthopädisches Physiotherapiezentrum, Graz, Austria
23 Joanneum Research, Graz, Austria
24 Landeskrankenhaus Klagenfurt, Dept. of Orthopaedic Surgery, Klagenfurt, Austria
25 Oberschwabenklinik, Wangen, Germany
26 Hospital Escala Braga, Orthopaedic Department, Porto, Portugal
27 Hospital de Curry Cabral, Lisbon, Portugal

Table of Contents

Introduction 5

Rationale 6

Materials and Methods 8
Methodology 8
Materials 9
Interpretation 10

Results 13
Evaluation of all Data 13
Implant Developers 18
Quality of the Literature: Analysis of the Region of Origin 23
- Quality of Literature from North America 23
- Outcome Literature by Implant Developers from the USA 26
- Quality of Literature from the UK 27
- Quality of Literature from Continental Europe 29

Numbers of Cases: Register Data and Clinical Studies 31
Underperformers in Registers 32
Outcome published in individual scientific Journals 33

Discussion and Summary 37
Statement to stimulate Discussion on the Results 41

Contact Information 42
Abbreviations 43
Acknowledgements 45
References 46

This final Report comprises the results of the QoLA (Quality of Literature in Arthroplasty) project, which was initiated by EFORT and EAR based on the results of the EUPHORIC project by the EU Commission's Directorate-General for Public Health and Consumer Protection (DG SANCO). The project was started in 2009 and led by EAR.

The methodology applied is largely based on the results of this previous project, which has meanwhile been completed and accepted by the EU Commission. Further information is available via the project website (www.euphoric-project.eu), as well as in a summary report presented during the 2009 EFORT Congress in Vienna. An electronic version of this report has been made accessible via the EAR webpage through the EFORT portal (www.ear.efort.org), section *EAR Publications*.

Rationale

It is a well-known fact that certain circumstances during the conduct of a clinical trial, e.g. patient selection, the surgeons' expertise and experience, or the study design, may have an impact on the results, and the question to what extent the results obtained are reproducible in the total patient population produces on-going critical discussions. This does not only apply to pharmaceutical studies but, of course, also to medical devices such as artificial joint implants.

In principle, two major datasets are available for the assessment of implants or surgical techniques:

1. Sample-based clinical studies that have been published in scientific journals
2. National and regional arthroplasty registers

Each of these datasets is characterised by specific priorities and requirements that should be taken into account in project planning and when interpreting the results.

Sample-based clinical studies:

- Try to extrapolate the results of a sample to the total patient population;
- Are usually applied to answer a particular question;
- Study design, measuring instruments or patient selection are therefore often very non-homogeneous, which is an essential advantage with regard to the precision in tuning the instruments but, in a meta-analysis, may lead to limitations.
- The characteristics of the collected data substantially affect the validity and possibilities of evaluations. Ordinal and nominal data, such as Yes/No decisions, or the formation of groups (for example: Excellent; Good; Fair; Poor) require relatively large numbers of cases in order to produce statistically significant differences and ensure sufficient statistical power. For example, the question of whether revision surgery has been performed falls into this category. In his PhD thesis Leif Ivar Havelin (Lit. 4) has shown that, to comply with the usual standards of a 95% confidence interval and a statistical power of 80%, a prospective study would require 13,474 patients in order to determine a 1% difference in outcome between two implants. Also, it would still require 3,008 patients to detect the relatively big difference of two percentage points. The execution of studies of that size quickly reaches organisational limits so that one must conclude by implication that the large majority of published studies might be statistically underpowered. Metric data, on the other hand, as used in most clinical scores, allow for reasonable evaluations already with a considerably lower number of patients.

Registers:

- Are designed to comprise all surgeries performed in a defined region, e.g. a state, thus providing a very realistic picture of the actual circumstances.
- To achieve completeness, the burden of documentation must not be too great. The questionnaires must therefore be confined to a relatively small core dataset.
- Apart from organisational difficulties, any modification to the dataset reduces the value of the data already collected for evaluations. Thus, registers are relatively inflexible.

- Data transferred to the register centre by the individual departments can thus be verified to a very limited extent, which should be taken into account when deciding on the contents to be recorded. Only objective and clearly-defined contents should be considered for the core dataset. However, many clinical scores contain a variety of data, such as pain or quality of life, that are strongly affected by subjective influences. These contents are less suited for regular data collection, but can be used successfully in projects including register datasets.

Thus, the two instruments and data sources do not compete with each other, but can sensibly complement one another. Registers offer advantages in recording and evaluation as regards revision rates and causes of revision. They are able to provide a realistic picture of the results in the area covered and considerably alleviate or eliminate effects arising in clinical studies, for example, due to patient selection or personal expertise. Completeness of collection is therefore an essential parameter for the quality of a register dataset. Clinical studies, on the other hand, have undeniable advantages in dealing with specific issues and subjectively-influenced answers.

Registers have been developed with great success in Scandinavia for more than 30 years, and impressive proof has been established of their usefulness for outcome measurement, quality control and quality improvement in many cases (1-15). During the past 10 years similar projects have been set up in quick succession in other countries so that an increasing number of datasets have become available for supranational analyses by now. These datasets can be used as reference values for comparative analyses regarding the reproducibility of published results. The quality, size, geographical distribution and length of follow-up periods of these datasets are only available in very few areas of medicine and for a very small number of indications. Thus, arthroplasty represents a positive exception in the medical field, and we have taken advantage of these positive circumstances for conducting a fundamental and critical analysis of those basic data that decisions have been based on worldwide.

Revision Rate is a recognised, well-defined and objective parameter after arthroplasty interventions that covers a variety of possible complications. The necessity for revision surgery has serious consequences for the patient's quality of life and causes high health-care expenditure. Decision-making largely follows standard procedures in diagnostic assessment and indication. This indicator is therefore well-suited for comparative analyses, and the conclusions are relevant for all major parties involved in the health-care system.

Materials and Methods

Methodology

When developing the methodology in the course of the EUPHORIC project the main question was how to summarise data based on different numbers of cases and follow-up periods in a single figure and make them directly comparable. We finally decided on the indicator 'Revisions per 100 observed component years', which was introduced in Orthopaedics by the Australian Joint Replacement Registry.

The formula for the calculation is:

$$\frac{\text{Number of cases of revision surgery for any reason}}{\text{Number of observed component years x 100}}$$

The concept of 'Revisions per 100 observed component years' is a recognised standard in epidemiology (16) and was, for example, used as early as the middle of the 20th century in providing evidence of the association between tobacco consumption and the incidence of lung cancer (17).

In principle, this method deals with calculating a correlation between the incidence of a potential risk exposure (e.g. cigarette smoking) and a consequential event (e.g. development of lung cancer). It also allows for considering essential influencing factors (e.g. smoking period or number of cigarettes) in the calculation.

Applied to arthroplasty, this means:

- There is a risk for revision from the moment of implantation. The total number of individual years from implantation (= observed component years) are counted.
- The total number of revisions (for any reason) as the failure end-point are documented and calculated in Revisions per 100 observed component years'.
- Back calculation of the calculated value into the usual way of presentation of Revisions/Time is possible by means of a linear function.
- A value of 1 represents a 1% revision rate at 1 year and a 10% revision rate at 10 years of follow-up.
- The advantage of this method is that it allows for comparison of datasets adjusted for the two main factors influencing the value of individual cohorts: number of cases and follow-up period.

This concept and the indicator can easily also be used for clinical studies.

Generous limits were defined for the definition of conspicuous datasets. In view of the multitude of potential influence factors, the possible uncertainty resulting from the calculation method, which will be discussed in more detail in the chapter Interpretation, has only a marginal effect.

Materials

A list was compiled of all implants for which data were available from arthroplasty registers, and that were suited for comparative analyses. The respective work packages were distributed among those partners who had evinced interest in collaboration after a call. The individual partners performed an analysis of clinical literature and the comparative values from registers were compiled by the EAR Scientific Office at the University Hospital of Orthopaedics in Innsbruck, Austria, which was also responsible for consolidation and comparable analysis of the data.

For the meta-analysis of peer-reviewed publications a structured literature review was performed based on electronic libraries such as Medline, followed by a manual literature research. Conventional meta-analyses were carried out from peer-reviewed journal publications in English and/or the native language of the partner in charge. The pooled results were stratified for potential influencing factors, such as the region of origin or whether the inventor of the respective implant had been part of the study team. The results of these investigations were compared with data from worldwide arthroplasty register reports.

Statistical analyses were performed calculating confidence intervals according to the current standards for meta-analyses.

These were the inclusion criteria for scientific articles to be considered in the subsequent evaluation:

- Unambiguous identification of the implant;
- Revision rate data (for any reason) either presented in the text or unambiguously calculable from the data contained. Unambiguous values were required for all items; an exception was only made in the case of follow-up times where also articles were accepted that merely indicated a time period. In that case a linear function was assumed for patient inclusion.
- Publications in Medline-listed, peer-reviewed journals.

Register data for calculations were obtained from annual reports or, if available, from journal publications. The most recent annual reports available were selected in all cases. In accordance with the register categorisation of the EUPHORIC project (18), only A.1.1.1 quality National reports were used. Thus, mainly National registers were included featuring a documentation completeness of more than 90% and published data validation. In the case of register datasets, precise values were strictly required.

Implant developers were identified through mentions in publications or manufacturers' documentations.

All papers were classified as developers' publications in which either the implant developer was listed as the author or as a co-author, or the developing institution's address was indicated for correspondence.

In cases where no developer was identified or no publications of the respective centre were available, the datasets concerned were not used for specific sub-group analyses.

In an additional step all studies dealing with Total Knee Replacement were analysed regarding the results published in individual journals.

Materials and Methods

Interpretation

Surgery outcomes are of course subject to certain fluctuations resulting from factors that are independent of the product used. They could be related to the profile of the patients treated in the respective department, the surgeons' expertise, specific surgical techniques, quality assurance measures, but also due to the influence of the particular public health system.

The maximum band-widths of the cumulative impact of these factors on the final outcome had to be calculated and evaluated. Various cross-sectional analyses of register data were conducted for this purpose.

A difference factor up to 3 (for instance, the revision rates of a dataset are three times as high as in the control group) between the datasets was considered to be explicable by individual expertise, circumstances in the particular hospital and other potential confounders. The value of 3 was chosen because this value covers the variability among individual hospitals in countries where National registers publish these data, such as the Swedish (Hip and Knee) Registers or the Danish National Arthroplasty Register, as well as the deviation from the mean of revision rates of individual implants in various National registers.

Calculating the deviation in outcomes achieved with the same implant in different countries covered by a National Arthroplasty Register from the worldwide average of the individual implant (as an estimate of non-implant-related impact factors) shows that the maximum outliers are also lower than a factor of 3:

	S	N	SF	DK	AUS	NZ	GB
AGC	0.94	0.56	0.76	2.39	0.77	0.38	
NexGen	0.37/2.71				1.55	1.66	1.27
Oxford Uni	0.86		1.17		0.97		
Duraloc	1.04			1.02	0.86		1.14
PFC	0.91			1.44	1.03	1.02	0.88

Even though the majority of datasets of both individual departments and individual implants show deviations that remain clearly below a factor of 3, the values of individual outliers are close to this cut-off point.

Particularly when analysing literature from centres of excellence it appears sensible to choose a generous limiting value to significance.

Therefore, to be rated as a significant value in the analysis, in terms of limited reproducibility in average patient treatment, the following criteria had to be fulfilled:

1. Deviations from the mean by a factor of 3, i.e. from 33% to 300%, as the measure of relevance;

2. Statistically significant deviation due to non-overlapping of confidence intervals in the main indicator 'Revisions per 100 observed component years' as a measure of the quality of datasets;

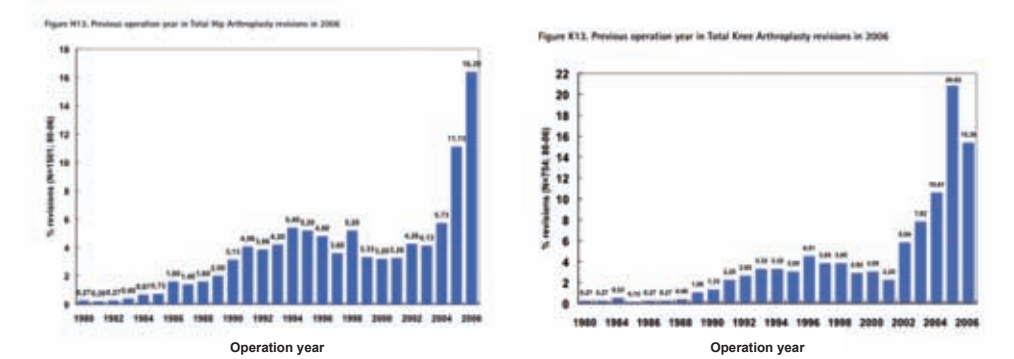
3. Or all studies included show a 100% survival rate, which means that not a single revision is documented. In this case it is mathematically impossible to compute confidence intervals, the deviation factor would be infinitely large.

Interpretation

Interpretation of data of the indicator 'Revisions per 100 observed component years':

The essential simplification behind the calculation of this indicator, which is mathematically inevitable, is the assumption that the distribution of revisions over time is linear.

This, however, does not correspond to the actual distribution. For example, the data of the Finnish register show that most revisions occur within the first years after primary surgery.



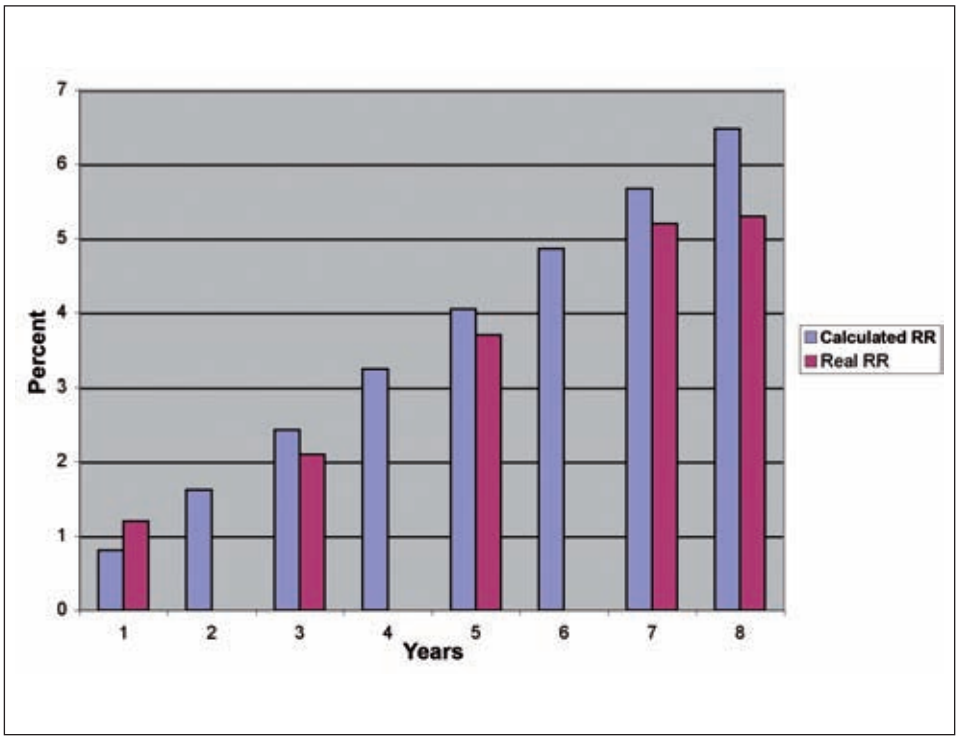
Total Hip Arthroplasty
Total Knee Arthroplasty
Finnish Arthroplasty Register, Annual Yearbook 2006: Year of primary surgery of cases revised in 2006

Interpretation

The Australian register publishes data for both 'Revisions per 100 observed component years' and the actual incidence of revisions so that the effect can be quantified.

In the first year after primary intervention the real values are slightly above the values calculated while they are slightly below in the long term.

In view of the broad scope allowed for the limit values and the large effects resulting from non-implant-associated impacts on the revision rate, the mathematical uncertainties appear very low and should not affect the overall result.



Average values of all implants of the Australian Joint Replacement Registry, comparing the back calculation of revision rates calculated based on 'Revisions per 100 observed component years' with the values actually measured. Basis: 2009 Annual Report

We would like to explicitly point out that this is a methodological study.
All values and factors refer to differences between datasets, i.e. to the inherent quality of the data.
They do not refer to absolute revision figures or the outcome of specific products assessed.

Evaluation of all Data

- For 21 implants (22 %) out of a total of 95 products examined not a single clinical study was available providing useful and workable data about the revision rate. Six of these implant systems – nearly one third and thus a remarkable proportion– are conspicuous as underperformers in registers.

Also implants manufactured by smaller, non-US based companies are strikingly often found in this group.

Implant	Manufacturer	
Natural Hip stem	DePuy	
RBK TKA	Global Orthopaedic Tech.	Small company
Option cup	Kinamed	Small company
F²L Multineck	Lima	Small company
Gemini TKA	Link	Small company
SPH-Blind cup	Lima	Small company
Journey TKA	S&N	New
Vanguard TKA	Biomet	New
Maxim TKA	Biomet	
Scan TKA	Biomet	
Citation stem	Stryker	
Scorpio TKA	Stryker	
ABG cup	Stryker	
Unix Uni	Stryker	
Anca-Fit stem	Wright	
Advantim TKA	Wright	
ZCA Uni	Zimmer	
ZUK Uni	Zimmer	New
Secur-Fit cup	Stryker	
Mitch Resurfacing	Stryker	
Freedom PKR Uni	Stryker	
GRU Uni	Global Orthopaedic Tech.	Small company

Legend: Red: underperformers
Green: innovative implants with short follow-up
Yellow: small companies

Evaluation of all Data

In the overall dataset 26 (35%) of the 74 implants and systems included in the comparative analysis showed statistically significant and relevant deviations from the benchmark, the respective outcome in register data. For all these implants also the sub-analysis of the publications authored or co-authored by developers revealed significantly better outcome.

- The highest value was found for the Optetrak Total Knee system, for which the average revision rates were 41.1 times higher in registers than in the clinical studies published.

- The vast majority of these systems, just like the respective studies, stem from North America, particularly from the USA.

- Developers' publications exhibit statistically significant and relevant deviations for another eight systems (10.6%). In three of these cases –the Taperloc stem, the Oxford Unicompartmental Knee Replacement, and the STAR Total Ankle Arthroplasty– this influence is as large as to lead to a significant and relevant impact on the overall dataset.

- In the case of two systems, Durom and the ABG stem, it is noticeable that the results published in clinical studies show a significantly and relevantly worse picture of the outcome than has been observed in registers. Possible reasons will be addressed in more detail in the Discussion chapter.

- For 43.2% of the implants examined, the clinical literature creates a significantly and relevantly too positive image of the outcome, in 2.7% of cases the revision rates seem to be clearly exaggerated.

Overall, the datasets of 45.9% of all implant systems examined are not in line with the outcome achieved in average patient service and could lead to wrong decisions because the basic data are affected by confounders.

Implant	Factor Difference between Outcome in Registers and comprehensive Publications in peer-reviewed Journals	Factor Difference between Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment	Region of Origin
Citation stem	Not a single revision published	n/a	n/a	n/a	n/a
Contemporary cup	Not a single revision published	n/a	n/a	n/a	n/a
Secur-Fit cup	Not a single revision published	n/a	n/a	n/a	USA
Summit stem	Not a single revision published	n/a	n/a	n/a	USA
Versys stem	Not a single revision published	n/a	n/a	n/a	USA
Secur-Fit stem	Not a single revision published	n/a	n/a	n/a	USA
Pinnacle cup	Not a single revision published	Not a single revision published	Only inventor studies published		USA
Vitalock cup	Not a single revision published	n/a	n/a	n/a	USA
Epoch stem	Not a single revision published	n/a	n/a	n/a	USA
Preservation Uni	Not a single revision published	n/a	n/a	n/a	USA
Secur-Fit cup	Not a single revision published	n/a	n/a	n/a	USA
Optetrak TKA	41.10	Not a single revision published	Yes	Yes	USA
Pappas-Büchel TAA	10.15	14.29	Yes	Yes	USA
Profemur Z stem	9.72	n/a	n/a	n/a	USA
C-stem	8.69	n/a	n/a	n/a	USA, D, GB
Corail stem	7.78	5.24	Yes	Yes	USA
CPT stem	7.33	n/a	n/a	n/a	USA
Synergy	6.79	n/a	n/a	n/a	USA
Charnley cup	5.28	n/a	n/a	n/a	GB
Trilogy	4.36	n/a	n/a	n/a	USA
AGC	4.01	4.15	Yes	Yes	USA
Genesis II	3.86	3.70	Yes	Yes	US, Can
Fitmore cup	3.22	n/a	n/a	n/a	EU

Results

Implant	Factor Difference between Outcome in Registers and comprehensive Publications in peer–reviewed Journals	Factor Difference between Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment	Region of Origin
Recap Resurfacing	3.17	n/a	n/a	n/a	n/a
Accolade Trident	3.17	n/a	n/a	n/a	USA
Natural Hip stem	3.13	n/a	n/a	n/a	USA
Taperloc	2.90	10.81	Yes	Yes	USA
Repicci Uni	2.89	n/a	n/a	n/a	USA
Bicontact	2.80	2.11	No	No	EU
Oxford Uni	2.71	4.37	Yes	Yes	GB
Link Uni	2.65	11.40	Yes	No	EU
Allofit cup	2.34	1.32	No	No	EU
Avon	2.18	2.17	No	No	GB
Charnley stem	2.17	n/a	n/a	n/a	GB
Spotorno CLS cup	2.11	9.05	Yes	No	EU
Cormet Resurfacing	2.1	n/a	n/a	n/a	EU
Definition stem	1.95	n/a	n/a	n/a	n/a
Hintegra	1.94	1.94	No	n/a	EU
Alloclassic	1.84	0.87	No	No	EU
Duracon TKA	1.71	1.48	No	No	USA
Durom Resurfacing	1.71	n/a	n/a	n/a	USA
Duracon Uni	1.59	n/a	n/a	n/a	USA
STAR	1.56	4.63	Yes	Yes	EU
Harris Galante cup	1.53	2.22	No	No	USA
ABG I cup	1.50	n/a	n/a	n/a	USA
Allegretto Uni	1.45	n/a	n/a	n/a	EU
LCS	1.46	1.17	No	No	USA
NexGen	1.45	n/a	n/a	n/a	USA
MG Uni	1.44	5.20	Yes	No	USA

Implant	Factor Difference between Outcome in Registers and comprehensive Publications in peer–reviewed Journals	Factor Difference between Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment	Region of Origin
Conserve Plus	1.43	1.47	No	No	USA
Advance TKA	1.41	n/a	n/a	n/a	USA
Profix	1.39	n/a	n/a	n/a	USA
BHR	1.33	4.33	Yes	No	GB
Triathlon TKA	1.29	n/a	n/a	n/a	USA
PFC Uni	1.26	n/a	n/a	n/a	USA
AML cement-less stem	1.22	4.74	Yes	No	USA
Duraloc	1.21	n/a	n/a	n/a	USA
Romanus cup	1.15	n/a	n/a	n/a	n/a
Natural-Knee	1.12	1.07	No	No	USA
ASR	1.06	n/a	n/a	n/a	USA
Agility	1.02	2.43	No	No	USA
SPII	0.99	n/a	n/a	n/a	EU
Spotorno	0.98	1.84	No	No	EU
Eius Uni	0.89	n/a	n/a	n/a	n/a
Kinemax TKA	0.83	2.75	No	No	USA
PFC	0.70	0.64	No	No	USA
Müller stem cemented	0.70	0.59	No	No	EU
Exeter stem	0.66	n/a	n/a	n/a	EU
RM cup	0.62	n/a	n/a	n/a	EU
Lubinus cup	0.58	n/a	n/a	n/a	EU
ABG stem	0.27	n/a	n/a	n/a	USA
Durom THA	0.25	n/a	n/a	n/a	USA

Results

Implant Developers

The analysis of publications by implant developers has yielded the following results:

- 14 out of 29 implants (48.3%) exhibit statistically significant and relevant deviations from register data and can therefore not be regarded as reproducible in average patient treatment.

- For nine of these 14 systems, the developer's influence on the results published in peer-reviewed journals is so large that it entails a statistically significant impact on the overall dataset.

- Nine of the developers concerned come from the USA, two from the UK, and another three from continental Europe. Significant bias of the overall dataset has been observed for seven developers from the USA, one from the UK, and one from continental Europe (Denmark).

Implant	Factor Difference between Outcome in Registers and comprehensive Publications in peer-reviewed Journals	Factor Difference between Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment	Region of Origin
Pinnacle cup	Not a single revision published	Not a single revision published	Only inventor studies published	---	USA
Optetrak TKA	41.10	Not a single revision published	Yes	Yes	USA
Pappas-Büchel TAA	10.15	14.29	Yes	Yes	USA
Corail stem	7.78	5.24	Yes	Yes	USA
AGC TKA	4.01	4.15	Yes	Yes	USA
Genesis II TKA	3.86	3.70	Yes	Yes	USA, Can
Taperloc stem	2.90	10.81	Yes	Yes	USA
Bicontact stem	2.80	2.11	No	No	EU
Oxford Uni	2.71	4.37	Yes	Yes	GB
Link Uni	2.65	11.4	Yes	No	EU
Allofit cup	2.34	1.32	No	No	EU
Avon	2.18	2.17	No	No	GB
Spotorno CLS cup	2.11	9.05	Yes	No	EU
Hintegra TAA	1.94	1.94	No	n/a	EU
Alloclassic stem	1.84	0.87	No	No	EU
Duracon TKA	1.71	1.48	No	No	USA
STAR TAA	1.56	4.63	Yes	Yes	EU
Harris-Galante cup	1.53	2.22	No	No	USA
LCS TKA	1.46	1.17	No	No	USA
MG Uni	1.44	5.20	No	No	USA
Conserve Plus	1.43	1.47	No	No	USA
BHR	1.33	4.33	Yes	No	GB
AML cement-less stem	1.22	4.74	Yes	No	USA
Natural-Knee TKA	1.12	1.07	No	No	USA
Agility TAA	1.02	2.43	No	No	USA
Spotorno stem	0.98	1.84	No	No	EU
Kinemax TKA	0.83	2.75	No	No	USA
PFC TKA	0.70	0.64	No	No	USA
Müller stem cemented	0.70	0.59	No	No	EU

Implant Developers

Influence of implant developers on the literature published in peer-reviewed journals:

- Implant developers generally have a large share in the literature published on their product.

- However, it is noticeable that, at least in the USA, developers who have published less than 25% of the total of observed component years for their product on average have published reproducible outcome.

- All developers publishing outcome that shows a significant and relevant discrepancy with register data –thus creating an unjustifiably positive picture– make up a proportion of more than 40% of all cases described for this product and hence dominate the publications about their own development.

- On the other hand, particularly in the case of specialty implants, such as total ankle arthroplasties or resurfacing implants, there are also examples where the developer exerts a similar influence while the published results are reproducible.

- A different situation is observed for continental Europe and the UK. Only the group from Oxford has been found to show a picture similar to the USA: non-reproducible results and a prevalence of the clinical literature.
  In the case of the Bicontact stem the developers, while having a comparably great influence on the literature about the product, published significantly higher revision rates than documented in registers.
  Otherwise developers publishing remarkable revision rates only account for a small proportion of the literature, or their results can be brought in line with the results from registers even though they dominate the publications.

Implant	Number Primaries	Number Revisions	Observed component years	Revisions per 100 observed component years	Ratio Difference Register vs. Clinical Studies	Ratio Difference Inventor vs. Clinical Studies	All Studies Ratio Difference to Register	% Primary Cases by Inventor
Optetrak TKA	448	1	2,283.50	0.04	infinite	infinite	41.10	74.78
Pinnacle cup	42	0	248.00	0.00	infinite	infinite	infinite	100.00
Pappas-Büchel TAA	517	36	3,152.60	1.14	1.92	14.29	10.16	57.25
Taperloc stem	1,929	36	16,114.86	0.22	5.50	10.81	2.90	44.53
Corail stem	214	2	2,507.50	0.08	5.24	5.24	7.78	68.69
AML cementless stem	577	23	5,992.17	0.38	9.92	4.74	1.22	62.43
AGC	30,596	571	310,872.85	0.18	2.38	4.15	4.01	85.93
Genesis II	15,049	136	95,433.36	0.14	1.15	3.70	3.86	47.51
Agility	682	82	2,917.33	2.81	4.03	2.43	1.02	31.96
Harris-Galante cup	7,352	393	68,481.41	0.57	1.60	2.22	1.53	31.46
Conserve Plus	2,023	140	10,134.00	1.40	1.42	1.47	1.43	96.24
LCS	14,196	863	162,271.22	0.53	0.79	1.17	1.46	5.67
Natural-Knee	1,514	68	10,847.70	0.63	1.12	1.09	1.12	91.88
PFC	14,363	617	88,090.00	0.70	0.92	0.64	0.70	7.21
MG Uni	449	29	3,639.00	0.80	5.20	1.16	1.44	27.17
Duracon TKA	472	13	3,537.00	0.37	1.95	1.48	1.71	23.73
Kinemax TKA	1,887	132	16,022.00	0.82	4.96	2.75	0.83	27.72

Relation between developers' share in the total volume of publications worldwide and the reproducibility of the data from developers from the USA

Implant	Number Primary Cases	Number Revision Cases	Observed component years	Revisions per 100 observed component years	Ratio Difference Inventor/ Register	Ratio Difference Inventor vs. Clinical Studies	All Studies Ratio Difference to Register	% Primary Cases by Inventor	% Revisions by Inventor	% Observed component years by Inventor
Oxford Uni	3,311	175	24,202.56	0.72	4.37	2.76	2.71	46.00	40.57	65.30
BHR Resurfacing	2,104	52	9,253.00	0.60	4.33	3.71	1.33	26.43	8.31	25.18
Avon retropatellar KA	663	30	3,231.00	0.93	1.27	0.91	2.18	93.06	96.67	96.36

Relation between developers' share in the total volume of publications worldwide and the reproducibility of the data from developers from the UK

Implant	Number Primary Cases	Number Revision Cases	Observed component years	Revisions per 100 observed component years	Ratio Difference Inventor/ Register	Ratio Difference Inventor vs. Clinical Studies	All Studies Ratio Difference to Register	% Primary Cases by Inventor	% Revisions by Inventor	% Observed component years by Inventor
Alloclassic	8,576	194	57,445.74	0.34	0.56	0.87	1.84	7.63	14.43	6.83
STAR	1,233	149	5,676.61	2.62	3.40	4.63	1.56	14.92	6.04	17.94
Hintegra	403	25	975.90	2.56	0.00	1.94	1.94	100.00	100.00	100.00
Bicontact	1,264	17	10,790.00	0.16	0.26	2.11	2.80	43.20	88.24	66.48
Spotorno CLS cup	3,833	90	31,387.80	0.29	2.03	9.05	2.11	7.80	1.11	4.76
Müller stem cem.	6,551	266	45,315.50	0.59	1.18	0.59	0.70	1.88	3.01	2.55
Link Uni	3,276	171	46,823.70	0.37	11.40	1.36	2.65	39.50	12.87	55.27

Relation between developers' share in the total volume of publications worldwide and the reproducibility of the data from developers from continental Europe

Quality of the Literature: Analysis of the Region of Origin

It must generally be stated that the majority of publications usually come from the developer's region, which is, of course, also reflected by the distribution of the respective implants in patient treatment. For the products developed in North America, a considerable number of publications are also available from other continents. Conversely, there are rarely any US publications dealing with European developments.

Quality of Literature from North America

- Of 44 products developed in the USA and Canada, 20 (44.5 %) show statistically significant and relevant deviations from register data in the overall dataset, presenting an overly positive picture of the outcome.
  For 14 (31.8%) of these 44 products no individual developer or clearly defined group of developers can be identified; hence the publications do not come from a particular group that is associated with specific circumstances (which will be addressed in more detail in the *Discussion* chapter).

- 1 product (2.3 %) shows increased revision rates in clinical studies as compared to register datasets.

- For another three products (6.8%) the developers' publications show statistically significant and relevant deviations, which leads to a statistically significant impact on the datasets in one case (Taperloc stem). Due to the influence of developer-independent studies, however, the overall datasets are within the limits where deviations are explicable by specific circumstances.

- It can only be attested for 20 (45.5%) of the datasets from North America that the published outcome is reproducible in routine patient treatment as is reflected in worldwide register data.

Implant	Factor Difference Outcome in Registers and comprehensive Publications in peer–reviewed Journals	Factor Difference Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment
Secur-Fit cup	Not a single revision published	n/a	n/a	n/a
Summit stem	Not a single revision published	n/a	n/a	n/a
Versys stem	Not a single revision published	n/a	n/a	n/a
Secur-Fit stem	Not a single revision published	n/a	n/a	n/a
Pinnacle cup	Not a single revision published	Not a single revision published	Only inventor studies published	n/a
Vitalock cup	Not a single revision published	n/a	n/a	n/a
Epoch stem	Not a single revision published	n/a	n/a	n/a
Preservation Uni	Not a single revision published	n/a	n/a	n/a
Secur-Fit cup	Not a single revision published	n/a	n/a	n/a
Optetrak TKA	41.10	Not a single revision published	Yes	Yes
Pappas-Büchel TAA	10.15	14.29	Yes	Yes
Profemur Z stem	9.72	n/a	n/a	n/a
C-stem	8.69	n/a	n/a	n/a
Corail stem	7.78	5.24	Yes	Yes
CPT stem	7.33	n/a	n/a	n/a
Synergy	6.79	n/a	n/a	n/a
Trilogy	4.36	n/a	n/a	n/a
AGC	4.01	4.15	Yes	Yes
Genesis II	3.86	3.70	Yes	Yes
Accolade Trident	3.17	n/a	n/a	n/a
Natural hip stem	3.13	n/a	n/a	n/a
Taperloc	2.90	10.81	Yes	Yes
Repicci Uni	2.89	n/a	n/a	n/a

Implant	Factor Difference Outcome in Registers and comprehensive Publications in peer–reviewed Journals	Factor Difference Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment
Duracon TKA	1.71	1.48	No	No
Durom Resurfacing	1.71	n/a	n/a	n/a
Duracon Uni	1.59	n/a	n/a	n/a
Harris-Galante cup	1.53	2.22	No	No
ABG I cup	1.50	n/a	n/a	n/a
LCS	1.46	1.17	No	No
NexGen	1.45	n/a	n/a	n/a
MG Uni	1.44	5.20	No	No
Conserve Plus	1.43	1.47	No	No
Advance TKA	1.41	n/a	n/a	n/a
Profix	1.39	n/a	n/a	n/a
Triathlon TKA	1.29	n/a	n/a	n/a
PFC Uni	1.26	n/a	n/a	n/a
AML cement-less stem	1.22	4.74	Yes	No
Duraloc	1.21	n/a	n/a	n/a
Natural-Knee	1.12	1.07	No	No
ASR	1.06	n/a	n/a	n/a
Agility	1.02	2.43	No	No
Kinemax TKA	0.83	2.75	No	No
PFC	0.70	0.64	No	No
Durom THA	0.25	n/a	n/a	n/a

Outcome Literature by Implant Developers from the USA

• Examining exclusively implants for which developers or groups of developers are documented yields a similar result: nine (53 %) out of 17 products show statistically significant and relevant deviations in the overall dataset or in publications authored by the developer.

• Overall, in North America 47.91 % of all primary surgeries reported on in sample-based journal publications were performed in developing centres, the value for observed component years is at 48.38 %. Thus, follow-up studies from developing centres on average do not have longer follow-up periods. The value for revision surgeries amounts to 30 %. On average developing centres hence publish slightly fewer revisions than independent studies.

Implant	Factor Difference Outcome in Registers and comprehensive Publications in peer–reviewed Journals	Factor Difference Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment
Pinnacle cup	Not a single revision published	Not a single revision published	Only inventor studies published	No
Optetrak TKA	41.10	Not a single revision published	Yes	Yes
Pappas-Büchel TAA	10.15	14.29	Yes	Yes
Corail stem	7.78	5.24	Yes	Yes
AGC	4.01	4.15	Yes	Yes
Genesis II	3.86	3.70	Yes	Yes
Taperloc	2.90	10.81	Yes	Yes
Duracon TKA	1.71	1.48	No	No
Harris-Galante cup	1.53	2.22	No	No
LCS	1.46	1.17	No	No
MG Uni	1.44	5.20	No	No
Conserve Plus	1.43	1.47	No	No
AML cementless stem	1.22	4.74	Yes	No
Natural-Knee	1.12	1.07	No	No
Agility	1.02	2.43	No	No
Kinemax TKA	0.83	2.75	No	No
PFC	0.70	0.64	No	No

Quality of Literature from the UK

- The analysis generally shows a non-homogeneous picture that is dominated by a few research groups.

- Presumably owing to the long history, no studies by Sir John Charnley could be included, although a multitude of studies with large numbers of cases were published about his developments. While in the publications by users the results of the stem on average did not show any irregularities, the revision rates described for the cup were considerably below those available from registers – in spite of the fact that registers also include recent cases and the data therefore more strongly reflect the further development in Orthopaedics.

- The publications about the Oxford unicompartmental prosthesis are largely dominated by the group of developers. The published results deviate from the comparable values in registers to a significant and relevant extent and have a statistically significant impact on the overall dataset. A detailed analysis regarding this product has been published in Acta Orthopaedica (19).

- McMinn's publications concerning the BHR system exhibit statistically significant and relevant deviations from register data and should therefore be subject to critical analysis. However, since only 20 % of the published cases stem from this group and independent literature on average even publishes slightly higher rates of revision than shown in the register dataset, the overall dataset on average shows reproducible values.

- Even though publications almost exclusively come from the developing hospital, the revision rates published for the Avon system are well-reproducible. The fact that the average revision rate documented in registers is approximately twice as high can be sufficiently explained by personal expertise and the learning-curve effect.

- Overall, 5.92 % of the primary cases and 4.87 % of the revision cases published for products from the UK stem from developing institutions. The value for observed component years is 3.14 %, which is due to the long follow-up periods of several large studies about the Charnley system.

- For those systems for which publications from developing institutions are available, the mean values are: 44.36 % of primary cases; 40.61 % of revisions; and 57.91 % of observed component years. However, the average values are strongly influenced by publications concerning the Oxford unicompartmental implant.

Quality of Literature from the UK

Implant	Factor Difference between Outcome in Registers and comprehensive Publications in peer–reviewed Journals	Factor Difference between Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment	Region of Origin
Charnley cup	5.28	n/a	n/a	n/a	GB
Oxford Uni	2.71	4.37	Yes	Yes	GB
Avon	2.18	2.17	No	No	GB
Charnley stem	2.17	n/a	n/a	n/a	GB
BHR	1.33	4.33	Yes	No	GB

Quality of Literature from Continental Europe

- Only one out of 16 systems developed in continental Europe shows significantly superior outcome compared to the benchmark. However, no publications by the developer are available for the system concerned (Fitmore).

- Another implant (ABG stem) on average shows significantly higher revision rates in clinical studies than in registers. A relevant number of these studies were published in the course of the controversial debate regarding the Robodoc system, which was used with this implant.

- Otherwise, all data show reproducible average values.

- For the CLS Spotorno cup, there is one developer study with a small number of cases that differs significantly and relevantly from all other data. These data therefore have no major impact on the overall dataset; the discrepancy would have been quickly noticed in a conventional meta-analysis.

- As regards the STAR Total Ankle Arthroplasty, the developer's publications also differ considerably from independent studies and register data. Here 14.9% of the cases published come from the developing hospital, which, together with the large difference, has a statistically significant impact on the average outcome of the overall dataset. However, there are also sufficient data available from independent studies to recognise the exceptional nature of the results of this single centre in a critical analysis.

- Regarding all datasets, approximately 8% of all published cases originate from developing hospitals (primary cases: 7.38%; revision surgeries: 8.60%; observed component years: 8.02%). Among the regions under examination, Europe is the only one where the proportion of revisions from developing institutions is slightly higher than the proportion of primary surgeries. This means that implant developers on average publish slightly higher rates of revision than ordinary users. However, the differences within Europe are not statistically significant.

- Analysis of those datasets for which developers' publications are available shows that for the implants in question on average 9.32% of primary surgeries, 10.15% of revision cases, and 9.48% of observed component years come from developing hospitals.

Quality of Literature from Continental Europe

Implant	Factor Difference between Outcome in Registers and comprehensive Publications in peer–reviewed Journals	Factor Difference between Inventor Outcome and Register Outcome	Inventor Bias	Inventor Bias leading to Bias in aggregated Assessment
Fitmore cup	3.22	n/a	n/a	n/a
Bicontact	2.80	2.11	No	No
Link Uni	2.65	11.4	Yes	No
Allofit cup	2.34	1.32	No	No
Spotorno CLS cup	2.11	9.05	Yes	No
Cormet Resurfacing	2.10	n/a	n/a	n/a
Hintegra	1.94	1.94	No	n/a
Alloclassic	1.84	0.87	No	No
STAR	1.56	4.63	Yes	Yes
Allegretto Uni	1.45	n/a	n/a	n/a
SPII	0.99	n/a	n/a	n/a
Spotorno	0.98	1.84	No	No
Müller stem cem.	0.70	0.59	No	No
RM cup	0.62	n/a	n/a	n/a
Lubinus cup	0.58	n/a	n/a	n/a
ABG stem	0.27	n/a	n/a	n/a

Numbers of Cases: Register Data and Clinical Studies

- With reference to all implants in the respective regions, the data shows that the number of primary and revision operations, as well as of observed component years recorded in the high-quality registers included in this study considerably exceed the cumulative number of cases treated in clinical studies worldwide.

- Although there is no operative National register in the largest markets worldwide, like the United States or Germany, and the British National Joint Registry was not included since it did not meet the strict inclusion criteria yet, 4.31 times as many primary surgeries, 4.5 as many revision surgeries, and 2.8 times as many observed component years are documented in worldwide Register datasets as by all clinical studies taken together.

- An overall analysis of a total of 42 implant datasets worldwide for which developers' publications have been identified shows that 26.55% of all published cases come from developing hospitals. The value for observed component years even reaches 24.18%, which is explicable by the longer follow-up periods of these studies. Moreover, 16.5% of all re-operations have been published by implant developers.

Underperformers in Registers

A special evaluation was performed regarding publications on products which showed statistically significantly higher revision rates in registers.

These are the results:

- There is not a single case in which potential problems can be derived from an analysis of the clinical literature. This is also true for recent cases such as ASR or Durom.
- In general, only very small numbers of cases were published for the products concerned.
- A developer's study was only identified for one product (Optetrak TKA).
- Not a single revision was published by developers.

Product	Number	Factor Difference between Inventor Outcome and Register Outcome	Inventor Bias
ASR	1092	49	No inventor study
Durom			No inventor study
Optetrak	449	1	
EIUS Uni	144	16	No inventor study
Recap	119	3	No inventor study
Profemur Z	2131	68	No inventor study
Preservation Uni	132	0	No inventor study

Outcome published in individual scientific Journals

One aspect in the assessment of the differences found in the published outcome of individual implants between clinical studies and register data is a potential publication bias. A special evaluation was performed to particularly analyse whether and to what extent a hypothetical accumulation of such studies in individual journals might affect the assessment of a product.

The analysis comprised all publications dealing with Total Knee Arthroplasty. The reason for choosing this particular product group was that the datasets available in this field are more homogeneous since they can be unambiguously assigned to a specific product. Free implant combinations, which are an option in Total Hip Arthroplasty in the case of cup and stem components – and implies that one implant could affect the revision frequency of the respective partner implant, are not possible for knee endoprostheses.

A total of more than 200 papers, which contained about 100,000 primary cases for 12 implants and were in published in 26 scientific journals, were included in the analysis.
The main criteria for evaluation were the distribution of studies by developers and independent literature, as well as the published outcome in individual scientific, Medline-listed journals. Furthermore, the data were analysed with respect to the region they were published.

<u>These are the most important results of the analysis:</u>

- The following journals have published the largest accumulated numbers of cases:

Journal of Arthroplasty (US)	33,728 primary cases
Clinical Orthopaedics and Related Research (US)	18,356 primary cases
Acta Orthopaedica (EU)	15,919 primary cases
JBJS-Br (EU)	6,625 primary cases
JBJS-Am (US)	5,967 primary cases

- Overall, about two thirds of cases were published in USA journals.
- The average revision rate published in USA journals is about 1.6 times lower than in European journals.

Outcome published in individual scientific Journals

- Compared to the worldwide benchmark from registers, the difference corresponds to a ratio of 3.43 and is therefore beyond the limit set in this project for differences explicable by usual influencing factors.

Asia	8 publications	842 cases	0.51 Revisions per 100 observed component years
EU	82 publications	31,217 cases	0.55 Revisions per 100 observed component years
USA	113 publications	67,397 cases	0.35 Revisions per 100 observed component years

- Worldwide about 30% of all studies dealing with total knee endoprostheses have been published by implant developers.
  The average revision rate of developers' publications corresponds to a survival rate of 98% after 10 years. The average revision rate of other users of the same implant is approximately 2.6 times higher.

Implant developer not identifiable	30 publications	3,922 cases	0.34 Rp100ocy
Developer	38 publications	29,148 cases	0.21 Rp100ocy
Independent	132 publications	56,383 cases	0.56 Rp100ocy

- No studies by developers have been identified in journals published in Asia. In European journals on average 7.5% of cases have been published by developers. The average outcomes published do not differ from other authors' results.

Implant developer not identifiable	17 publications	2,185 cases	0.41 Rp100ocy
Developer	7 publications	2,342 cases (7.5 %)	0.47 Rp100ocy
Independent	58 publications	26,739 cases	0.47 Rp100ocy

- In USA journals about 55 % of all primary cases have been published by implant developers. The average revision rate published is outstanding. It corresponds to a survival rate of 98 % or 0.19 revisions per 100 observed component years at 10 years. This value is 2.6 times lower than in comparable publications by

Implant developer not identifiable	11 publications	1,581 cases	0.28 Rp100ocy
Developer	31 publications	36,806 cases (54.6 %)	0.19 Rp100ocy
Independent	70 publications	29,010 cases	0.56 Rp100ocy

other users and approximately 10 times than the average outcome in registers – taken as a benchmark for worldwide average patient treatment.

Outcome published in individual scientific Journals

- The vast majority (96.8%) of all studies by developers identified in USA journals were published in The Journal of Arthroplasty (JOA) and in Clinical Orthopaedics & Related Research (CORR). 58.4% of cases were published in JOA, 38.5% in CORR.

- Based on the material analysed, 76.15% of all cases published in JOA and 63.04% of all cases published in CORR originate from developing hospitals.

- The average revision rates of the studies by developers published in CORR (0.14 revisions per 100 observed component years) and JOA (0.20 rp100ocy) are markedly better than in the other journals. The respective values are 1.22 rp100ocy for JBJS-Am 1.22 rp100ocy, and 0.48 rp100ocy for JBJS-Br. From Acta Orthopaedica no inventor studies were available about the implants examined.

- In summary, two journals, the Journal of Arthroplasty and Clinical Orthopaedics & Related Research show extraordinary results. The vast majority of developers' studies are published in these two journals. Within the category of outcome studies, developers –as a special interest group which is not representative for the average surgeon– account for a surprisingly high proportion of cases published in these journals.
  The published results show statistically significant and relevant deviations from the revision rates reported in registers or published in other journals. As to the causes for these observations, which may be related to various circumstances in the course of the publishing process, from manuscript submission to a particular journal via the review procedure through to editorial board decisions, no conclusions can be drawn based on the material available.

- The high influence of these two journals within the scientific community with respect to arthroplasty, as well as the extraordinary results published may have an impact on the assessment of implants that might remain undetected in the current procedures of structured decision-making by both physicians and public health authorities if no reference values are available.

- A large band-width has been observed as to the average values regarding the reproducibility of results published in clinical sample-based studies in peer-reviewed journals.

- There are considerable discrepancies at to the reproducibility of published results between North America, notably the USA, and Europe.

- More than 50% of the USA datasets, to a statistically significant and relevant extent, are not reproducible in average patient treatment and/or may lead to misinterpretations in meta-analyses performed according to the procedures currently applied.

- A relevant proportion of published outcome to a statistically significant extent shows overly positive results.

- The possible explanation that the general results after arthroplasty interventions are better in the US cannot be confirmed by comparative analyses between various countries. Compared to Europe, the average outcomes achieved in patient treatment are worse in the USA (18).

- On average, published revision rates are lower in clinical studies than in register data. However, this could, for example, be explained by the fact that clinical studies are usually conducted in centres of excellence, whereas register data also include small departments.

- However, no clear trend towards generally positive outcome publications can be derived. Compared to the proportion of datasets potentially compromised by relevant confounders, nearly just as many outcome results actually are reproducible. This is not easy to explain, and it must be doubted that there is a general reason.

- Regarding two products the studies published, contrary to the general trend, show markedly higher revision rates than registers.

    - Durom cup: Two publications met the inclusion criteria. The dataset is strongly influenced by a publication by Long et al, from the group led by L. Dorr. For some time now, there has been an intensive discussion about increased revision rates regarding this product. Registers also show increased rates of revision as compared to other products, but they are not as high as described in the cited article so that an impact by specific circumstances in the centre concerned cannot be excluded. There are relatively few publications dealing with this implant, and they are by no means adequate to allow for drawing final conclusions.

    - ABG stem: 24 European publications and one from New Zealand are available on this implant. Even though a direct relation to this product was only established in few publications, it was involved in a critical discussion concerning the use of the ROBODOC implantation system. Publication activity shows a peak around the time of this discussion and shortly afterwards. Historical cases have shown that the opinion prevailing on a certain product within the medical community may have an impact on scientific publications (20). It cannot be excluded that this phenomenon was also effective in this case. Apart from this, it is conspicuous that no publications are available from the USA.

- o It should generally be stated that scientific publications reporting on the occurrence of increased revision rates fulfil an important function for all users. In the majority of cases in which registers indicate existing problems with a product or its handling, such articles are missing completely. Moreover, it is striking that in the respective cases –for example, the ASR cup– mostly no publications are available from the developers at all.

- Contrary to the data from North America, the vast majority of European results are well-reproducible and show good validity. Even though there are individual groups of developers who draw an unjustifiable, overly-positive picture of their product, sufficient independent publications of good quality are available (except for the group from Oxford) to be able to recognise discrepancies even in a conventional meta-analysis based on scientific publications.

- Whereas outside of the United States the vast majority of developer-independent publications have shown reproducible results, this does not apply to the US. Here, even a considerable proportion of the independent literature presents significantly and relevantly better outcomes than are shown by the comparative values from worldwide registers.

- Striking differences have been observed in the published data and publication behaviour between the United States and Europe.

  - o Among the products for which developers have been identified, 58 % of published cases come from the developing hospital. In Europe this applies only to about 14 % of cases.

  - o The strong influence of developing institutions on scientific publications in the USA entails that results which are irreproducible in average patient treatment are hardly recognised because comparative data are often unavailable.

  - o Even publications by USA users who are not directly involved in product development cannot be reproduced to a relevant extent – a phenomenon that has been observed only in one single case in Europe.

  - o There are marked differences in the published number of cases for individual implant systems. In the USA, it is conspicuous that usually only few studies with low numbers of cases are available in the case of implants for which no developer has been identified.

  - o What is remarkable with respect to European products is that in the majority of cases relatively few studies have been published with low numbers of cases, particularly if the products have been developed in non-English-speaking countries and are not being marketed by a big international manufacturer. It should be investigated whether there are specific factors that negatively affect the chances of an article being accepted for publication.

  - o On average, the published literature on European products shows considerably better quality and reproducibility than the US literature.

- Publications from Europe are in general less influenced by particular groups, and convey a more democratic picture and wider scope of experiences.

- Nevertheless critical evaluation is recommended in individual cases.

• Implant developers have a strong influence in the published clinical literature and therefore, sometimes to a relevant extent, determine the users' assessment of the product as well as product-relevant administrative decisions, for example, in certification procedures, market monitoring, or regarding the choice of a certain system in tendering procedures.

The developers' influence in these procedures is by far larger with USA products and publications than it is in Europe. Usually both the users and public health authorities are interested in outcome data mainly to be able to estimate future quality in treatment or the complication rates to be expected for application in routine patient service.

However, centres and physicians involved in implant development are not, or only to a limited extent representative for average patient treatment with regard to several aspects.

- As a rule, the hospital concerned can rely on a high degree of expertise and a fundamental understanding of the product and its handling.

- High personal motivation can be assumed when it comes to the thorough investigation of potential, outcome-relevant flaws in the entire course of therapy, and drawing the consequences.

- The final result of a THA implantation depends on a variety of factors, such as the product, instrumentation, operating technique, patient selection, etc. Since every product is developed against a specific background and based on a specific set of experiences, the product might make particular allowance for the factors prevailing at this hospital.

- On the other hand, it would be absolutely conceivable that implant developers also test the limits of their products and have to accept revisions due to increased learning curves, for example, while defining the limits of potential product applications, whereas subsequent users profit from these findings in routine patient treatment. This would be a possible explanation why some implant developers on average even publish higher revision rates than are shown in independent studies or register data.

- Finally, one should also bear in mind that implant developers and manufacturers have a fundamental interest in the success of their product.

The mere fact that a certain author publishes data deviating from the benchmark does therefore not allow for drawing conclusions on the reasons for the discrepancy. However, one should critically consider the value of these data for one's own decisions.

• Arthroplasty registers can essentially support the evaluation of outcome data.

- They can serve as a benchmark in verifying the reproducibility of clinical studies. The methodology and basic concepts are being developed in the present project.

- Registers refer to all surgeries performed in a certain region and can therefore reduce or exclude several sample-based bias factors.

- The results of arthroplasty registers, just as every study, include the circumstances under which the data have been collected. Differences between countries, data collection and evaluation procedures may influence the results and conclusions drawn. This should be taken into account in the interpretation of results.

- Fluctuations in results from register datasets are generally considerably lower that in clinical studies.

- **In general, clinical studies seem to be affected by multiple confounders.**

- **Inferior outcome documented by Arthroplasty Registers would not be detected by analyses of clinical studies – on the one hand because only very little published material is often available about the products concerned, on the other hand because the publications often present the good results achieved in small series.**

- **It cannot be excluded that specific practices in the peer-review procedure and decisions made in the publication process may also affect the published results. This is essential also in view of the fact that current health technology assessment procedures, even though they allow for structured analysis of the published literature, cannot detect possible confounders in the process publishing.**

- **The results of the QoLA project indicate that there are hidden confounders in Arthroplasty that may influence the decisions of physicians, public health authorities and other stakeholders.**

Statement to stimulate Discussion on the Results

- As a general rule, analyses of clinical literature should only be conducted comprehensively and interpretation should be handled with care.

- Comparative data should be included whenever possible. If applicable, stratifications should be carried out according to the region of origin of the data and studies.

- In the case of contradictory results, register data should be rated as superior since they are less susceptible to sample-based confounders.

- Independent of the product, on average 1.2 to 1.3 revisions per 100 observed component years must be expected for total hip and knee endoprostheses. This would correspond to an average revision rate of about 6% at five years and of about 12% at 10 years. Data from studies that strongly deviate from this average value, i.e. by a factor of 3-5 or above, should be critically analysed and examined for signs of sample-based confounders, such as

    - o Patients lost to follow-up;
    - o Strict inclusion and exclusion criteria leading to the selection of patients with a favourable risk profile;
    - o Statistical power;
    - o Relation to implant developer;
    - o Specific expertise or the fact that the study centre mainly treats patients providing very favourable conditions for good outcomes.

- In view of the fact that implant developers have great influence on the published results, efforts should be made in the future to provide the reader with more transparent information on the specific circumstances under which the data have been obtained.

Since there are definitely implant developers in all continents who are renowned for their surgical skills while their published results achieved with their own developments are well-reproducible by other surgeons, stating a general reason alone, such as higher expertise, does not suffice to explain differences in outcome.

Contact Information

Dr. Gerold Labek
EFORT European Arthroplasty Register Coordinator
Vice President EFORT-EAR

Innsbruck Medical University
Department of Orthopaedic Surgery
Anichstrasse 35
A-6020 Innsbruck

Phone +43 512 504 81600
Fax +43 512 504 22701
e-mail gerold.labek@efort.org

Renate Fechter, M.A.
EAR Scientific Office
Project Management Assistant

Innsbruck Medical University
Department of Orthopaedic Surgery
Anichstrasse 35
A-6020 Innsbruck

Phone +43 512 504 22695
Fax +43 512 504 22701
E-mail renate.fechter@uki.at

Abbreviations

CORR	Clinical Orthopaedics and Related Research; US journal published by the Association of Bone and Joint Surgeons
EAR	European Arthroplasty Register, an EFORT-affiliated, non-profit scientific society focused on outcome research in arthroplasty and Arthroplasty Registers
EFORT	European Federation of National Associations of Orthopaedics and Traumatology
EUPHORIC	European Public Health Outcome Research and Indicators Collection; EU project funded by the EU Commission (Directorate General for Health and Consumers) under the Community Action Programme for Public Health 2003-2008 (Grant Agreement 2003134).
JBJS	Journal of Bone and Joint Surgery; British and American edition
JOA	Journal of Arthroplasty; official journal of the American Association of Hip and Knee Surgeons
KA	Knee Arthroplasty
RCT	Randomised Controlled Trial
Rp100ocy	Revisions per 100 observed component years
TAA	Total Ankle Arthroplasty
THA	Total Hip Arthroplasty
TKA	Total Knee Arthroplasty

Acknowledgements

We would like to thank the large number of voluntary collaborators from over 20 scientific centres in Europe, Turkey and Israel for their great and ongoing commitment and support in verifying the reproducibility of outcomes of the orthopaedic literature available.

Thanks to all who have endeavoured for many years to promote the development of Registers, and to all those countless colleagues worldwide who, by their daily documentation efforts during many years, have made these valuable datasets possible in the first place.

Our thanks are also due to Professor George Bentley for reviewing our report.

References

1. Furnes A, Lie SA, Havelin LI, Engesæter LB, Vollset SE. The economic impact of failures in total hip replacement surgery. The Norwegian Arthroplasty Register 1987-1993. Acta Orthop Scand. 1996; 67:115-21

2. Garellick G, Malchau H, Herberts P. Survival of hip replacements. A comparison of a randomized trial and a registry. Clin Orthop. 2000 Jun; (375):157-67

3. Havelin LI, Engesaeter LB, Espehaug B, Furnes O, Lie SA, Vollset SE. The Norwegian Arthroplasty Register: 11 years and 73,000 arthroplasties. Acta Orthop Scand 2000; 71-4:337-53

4. Havelin LI. Hip arthroplasty in Norway 1987-1994. The Norwegian Arthroplasty Register. Bergen, Norway: Thesis University of Bergen, 1995

5. Henricson A, Skoog A, Carlsson A. The Swedish Ankle Arthroplasty Register: An analysis of 531 arthroplasties between 1993 and 2005. Acta Orthopaedica 2007; 78(5):569-74

6. Herberts P, Malchau H. How outcome studies have changed Total Hip Arthroplasty Practices in Sweden. Clin Orthop. 1997; 344:44-60

7. Herberts P, Malchau H. Long-term registration has improved the quality of hip replacement: a review of the Swedish THR Register comparing 160,000 cases. Acta Orthop Scand. 2000; 71-2:111-21

8. Lucht U. The Danish Hip Arthroplasty Register. Acta Orthop Scand 2000; 71-5:433-9

9. Malchau H, Garellick G, Eisler T, Kärrholm J, Herberts P. Presidential guest address: the Swedish Hip Registry: increasing the sensitivity by patient outcome data. Clin Orthop. 2005; 441:19-29

10. Puolakka TJ, Pajamaki KJ, Halonen PJ, Pulkinen PO, Paavolainen P, Nevelainen JK. The Finnish Arthroplasty Register: report of the hip register. Acta Orthop Scand. 2001 Oct; 72(5):433-41

11. Robertsson O. Knee Arthroplasty Registers. J Bone Joint Surg 2007; 89-B:1-4

12. Robertsson O, Dunbar MJ, Knutson K, Lewold S, Lidgren L. The Swedish Knee Arthroplasty Register. 25 years experience. Bull Hosp Jt Dis 1999; 58-3:133-8

13. Robertsson O, Lewold S, Knutson K, Lidgren L. The Swedish Knee Arthroplasty Project. Acta Orthop Scand 2000; 71(1):7-18

14. Robertsson O, Ranstam JP. No bias of ignored bilaterality when analysing the revision risk of knee prostheses: Analysis of a population based sample of 44,590 patients with 55,298 knee prostheses from the national Swedish Knee Arthroplasty Register. BMC Musculoskelet Disord 2003; 4:1

15. Söderman P, Malchau H, Herberts P, Johnell O. Are the findings in the Swedish National Total Hip Arthroplasty Register valid? A comparison between the Swedish National Total Hip Arthroplasty Register, the National Discharge Register, and the National Death Register. J Arthroplasty 2000 Oct; 15(7):884-9

16. Lilienfeld AM, Lilienfeld DE. Foundations of Epidemiology, 2nd Edition, Oxford Univ. Press 1980, p. 245

17. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking. A second report on the mortality of British doctors. BMJ Nov. 10, 1956; 5001: 1071-1081

18. Labek G, Stöckl B, Janda W, Hübl M, Thaler M, Liebensteiner M, Williams A, Pawelka W, Frischhut S, Haselsteiner F, Schlichtherle R, Agreiter M, Abel J, Dorninger R, Sekyra K, Rosiek R, Mayr E, Hochgatterer R, Pfeiffer K-P, Krismer M, Böhler N. Quality of Datasets for Outcome Measurement, Market Monitoring and Assessment of artificial Joint Implants, available at: www.ear.efort.org

19. Labek G, Sekyra K, Pawelka W, Janda W, Stöckl B. Outcome and Reproducibility of Data concerning the Oxford Unicompartmental Knee Arthroplasty – a structured Literature Review including Arthroplasty Register Data. Acta Orthopaedica 2011 Apr; 82(2):131-5

20. Liebensteiner M, Janda W, Williams A, Pawelka W, Labek G. Erfassung von minderwertigen Produkten in der Endoprothetik und Umsetzung der Erkenntnisse. Eine retrospektive Analyse am Beispiel des Boneloc-Knochenzementes. ZorthopUnfall 2009; 147(6): 683-8 [German]

Contact:

Innsbruck Medical University
Dept. of Orthopaedic Surgery
Anichstrasse 35
6020 Innsbruck, Austria

Phone +43 (512) 504 22695
Fax +43 (512) 504 22701
gerold.labek@efort.org
www.ear.efort.org

EFORT Head Office
Technoparkstrasse 1
8005 Zurich, Switzerland

Phone +41 (44) 448 4400
Fax +41 (44) 448 4411

office@efort.org
www.efort.org