EOR | VOLUME 3 | MAY 2018
DOI: 10.1302/2058-5241.3.170064
www.efortopenreviews.org

EFORT open reviews

Research methodology for orthopaedic surgeons, with a focus on outcome

Anne Lübbeke

- Since improving the patient's condition is the ultimate goal of clinical care and research, this review of research methodology focuses on outcomes in the musculoskeletal field.
- This paper provides an overview of conceptual models, different types of outcomes and commonly assessed outcomes in orthopaedics as well as epidemiological and statistical aspects of outcomes determination, measurement and interpretation.
- Clinicians should determine the outcome(s) most important to patients and/or public health in collaboration with the patients, epidemiologists/statisticians and other stakeholders.
- Key points in outcome choice are to evaluate both the benefit and harm of a health intervention, and to consider short- and longer-term outcomes including patient-reported outcomes.
- Outcome estimation should aim at identifying a clinically important difference (not the same as a statistically significant difference), at presenting measures of effects with confidence intervals and at taking the necessary steps to minimize bias.

Keywords: outcomes; epidemiology; statistics

Cite this article: *EFORT Open Rev* 2018;3
DOI: 10.1302/2058-5241.3.170064

Introduction

Why focus on outcome?

The outcome of a treatment is what matters most to the patients, and to improve the patient's condition is the ultimate goal of clinical care and clinical research.[1-3] To identify the most relevant outcome and to measure it precisely is challenging. This review is intended to facilitate that task. The spectrum of outcomes measured in routine healthcare and for research purposes has increased substantially over the past decades. One reason is the more and more widespread use of patient-reported outcomes (PRO).[4,5] Another reason is the increasing ability to routinely collect large amounts of diverse data in clinical and administrative databases and in electronic health records (e.g. deep infection after joint replacement[6]).

Orthopaedic surgeons, together with the patients and other healthcare providers (e.g. physiotherapists) involved in the treatment of a condition, are specialists in the clinical course of this specific condition. As a consequence, they are well positioned to determine the most important outcomes of a healthcare decision/intervention. In addition, collaboration with academics in the relevant field and with methodologists is beneficial, particularly for advice on how to measure outcomes. The outcome is at the centre of almost every research question, whether you want to determine the outcome of a treatment, assess the prognosis for a specific subgroup of patients, determine the cause of an adverse event, and even for diagnostic evaluations where you attempt to improve the diagnosis of a condition in order to positively influence its outcome. Outcomes are closely linked to the study's exposure of interest. The latter is the primary explanatory variable of interest and may be a risk factor (e.g. smoking status) or treatment type (e.g. surgical *versus* non-surgical treatment) or other. The conceptual framework of a particular study is determined by its specific exposure–outcome relation and also takes into account variables/factors, which are potential confounders or effect modifiers.[7] Outcome considerations determine much of the analysis plan, such as important difference determination, sample size calculation or length of follow-up, and the choice of analytic tools.

This review of research methodology focusing on the outcome is directed at orthopaedic and trauma surgeons and other healthcare professionals working in this field, who do research and/or are readers of the scientific literature. Numerous publications on the aspects of research methodology have been written – for the general medical audience and more specifically for those working in the musculoskeletal field,[8-11] and I will indicate them where appropriate. I will start by introducing two conceptual models of outcomes. I will then mention the most commonly assessed outcomes in our field and clarify the role

of biomarkers, surrogates, process and structural measures. Furthermore, I will describe how to measure the impact of an intervention on a given outcome while concentrating mainly on absolute and relative measures of effect, PROs and the target difference, and how to present the outcomes in the manuscript. Finally, I will describe systematic errors one needs to be aware of in outcomes determination.

The conceptual model of patient outcomes

In 1995, Wilson and Cleary described a five-level model of patient outcomes in their paper 'Linking clinical variables with health-related quality of life: The conceptual model of patient outcomes'.[3] Biological and physiological variables constitute level one of the model, symptom status level two, functional status level three, general health perceptions level four, and overall quality of life constitutes level five of the model. These levels are under the influence of characteristics of the patients and characteristics of the environment (e.g. social and economic support) as well as non-medical factors, all of which cannot be controlled by the physician. The influence of these factors and the complexity and difficulty in measuring the outcome increase from level one (biological and physiological variables) to level five (overall quality of life).

The outcome measures hierarchy

The 'Outcome Measures Hierarchy' proposed by Michael E. Porter in 2010[1] is based on the principles that multiple outcomes, most relevant to the patient, and including the short- and longer-term should be measured in healthcare evaluation. The hierarchy consists of three tiers of outcome measures applying to any medical condition: (1) health status achieved or retained; (2) process of recovery; and (3) sustainability of health. Each of the tiers consists of two levels of outcomes. Tier 1 'Health status achieved or retained' is characterized, first, by the proportion of patients who survive and, second, by the patients' degree of health or recovery (e.g. pain reduction and functional improvement achieved after joint replacement, ability to return to work). Tier 2 'Process of recovery' measures, first, the time to recovery and time to return to normal activities and, second, disutility of care and treatment process (e.g. diagnostic errors, ineffective care, treatment-related discomfort, complications, adverse effects). Tier 3 'Sustainability of health' evaluates the sustainability of the health status achieved and nature of recurrences (e.g. revision after joint replacement) as well as the long-term consequences of therapy (e.g. stiff knee after knee replacement, susceptibility of deep infection after joint replacement).

Common outcomes in orthopaedics and traumatology

There are an increasing number of collaborative efforts to define sets of outcomes or core outcome measures for specific conditions. Thus, in the design phase of every study, it is useful to search for and consider already existing outcome recommendations. Their aim is to standardize and harmonize outcome assessment and reporting between centres and countries. Examples include the Core Outcome Measures in Effectiveness Trials (COMET), the European Clinical Research Infrastructures Network (ECRIN) Database, the International Consortium for Health Outcomes Measurement (ICHOM),[12] as well as the core sets of domains from the World Health Organization's International Classification of Functioning, Disability and Health.[13] Moreover, there are numerous publications recommending sets of outcomes for specific conditions in orthopaedic or trauma surgery (e.g. Outcome measures for orthopaedic interventions on the hip;[14] Outcome assessment in fracture healing trials: a primer[11]) as well as an excellent overview for researchers and clinicians of outcome definition and measurement in observational comparative effectiveness research.[15]

Commonly assessed outcomes in orthopaedic and trauma surgery include: (1) mortality; (2) post-operative medical complications (e.g. deep vein thrombosis, pulmonary embolism, cardiovascular and gastrointestinal complications, anaemia, delirium); (3) infections (wound, implanted material, urinary, pulmonary, other); (4) peri-operative peripheral nerve injury; (5) post-operative orthopaedic complications (e.g. dislocation, peri-operative fracture near the implanted material, implant breakage or cut-out, loss of reduction, implant mal-positioning); (6) degree of bone healing such as osseointegration, loosening, deformity, mal-union, nonunion, osteonecrosis or heterotopic bone formation; (7) clinical outcomes assessed by the physician through history and clinical examination, such as pain, function, activity, range of motion or muscle strength; (8) performance testing (e.g. get-up and go, gait analysis, activity assessment with body-worn sensors[16,17]); (9) PROs such as pain, function, sleep, ability to live independently, return to work, recreational and daily living activities, general physical/mental health, quality of life or satisfaction; (10) concomitant treatment need (e.g. analgesia usage, physiotherapy); (11) subsequent surgery, such as re-operation, revision, implant removal, closed reduction of dislocation or arthrodesis; and (12) long-term implant-related systemic reactions such as allergy, adverse local tissue reactions or metal-ion induced systemic adverse events.

It is crucial to include both safety and efficacy (= ability to produce an expected result under ideal circumstances)/effectiveness (= ability to produce an expected result in the real-world clinical setting) or in other words benefit and harm of a health intervention. Moreover, the following points should also be considered: think of both the short-term and long-term if applicable; choose complementary outcomes measures such as objective (e.g. revision, gait

Table 1. Outcomes after hip replacement and ORIF proximal humerus fracture based on the outcome measures hierarchy

Outcome measures hierarchy	Hip replacement	ORIF proximal humerus fracture
Health status achieved or retained		
Survival	Mortality, short-term (surgery-related) and long-term (complication-, implant-related)	Mortality, short-term (surgery-related)
Degree of recovery/health	Symptom reduction (pain, function)	Symptom reduction (pain, function, range of motion)
	Degree of return to activities of daily living, work, sports	Degree of return to activities of daily living, work, sports
Process of recovery		
Time to recovery or return to normal activities	Time to being symptom-free	Time to being pain-free
	Time to return to physical activities	Time to healing (clinical, radiological)
	Time to return to work	Time to return to independence
		Time to return to work, recreational activities
Disutility of care/treatment process	Residual pain/analgesic use	Residual pain/analgesic use
	Length of stay in hospital	Length of stay in hospital
	Reduced range of motion	Reduced range of motion/reduced muscle strength
	Medical complications post-surgery	Medical complications post-surgery
	Infection (prosthesis, wound, urinary, pulmonary)	Infection (material, wound)
	Dislocation	Nerve lesion
	Peri-prosthetic fracture	Implant breakage/cut-out
	Impingement due to implant mal-positioning	Nonunion/deformity
Sustainability of health		
Sustainability of recovery or health over time	Maintenance of activity level over time, quality of life	Ability to live independently, quality of life
	Revision-/Re-operation-free interval	Maintenance of shoulder function over time
Long-term consequence of therapy	Risk of haematogenous deep infection	Re-fracture risk around implant
	Aseptic loosening/wear	Shoulder stiffness
	Peri-prosthetic fracture	Osteonecrosis
	Adverse local tissue reaction/metal ion allergy	Heterotopic bone formation
	Systemic effects of metal ions	

analysis) and subjective measures (e.g. PRO measures); and a clinically relevant outcome/endpoint is superior to a surrogate measure/endpoint (see also below).

In Table 1, the outcome measures hierarchy is used to guide the choice of outcomes after two different interventions: hip replacement and open reduction with internal fixation of a proximal humerus fracture. Wilson and Cleary's model[3] is complementary. It expands the range of health outcomes from the classical outcomes (biological and physiological variables, symptoms and functional status) to general health perceptions and quality of life and underlines the importance of considering different outcome levels in a study. General health perception and quality of life are patient-reported. Symptoms such as pain, functional and activity limitations - in the past obtained as part of physician-assessed outcome scores together with measures of range of motion, alignment or strength (e.g. Harris hip score, Constant-Murley shoulder score) - are now mainly assessed with PROs (e.g. HOOS, Oxford knee score). The topic of PROs is discussed in more detail below.

Biomarkers, surrogates, process and structural measures

The model by Wilson and Cleary defines biological and physiological variables as level one. They are part of so-called biomarkers defined as 'anatomical, physiological, biochemical, molecular, or genetic parameters associated with the presence, absence, or severity of a disease process'.[18] Another definition describes them as 'a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention'.[19] Biomarkers are especially useful as surrogates for clinical outcomes that are rare or occur very late (e.g. revision of joint replacement, mortality). According to Rigatto and Barret[18] 'a surrogate outcome is defined as a (bio)marker that is intended to serve as a substitute for a clinically meaningful end point and is expected to predict the effect of a therapeutic intervention'. It also has to 'be predictive of clinically important outcomes, of corresponding changes in clinically important outcomes when itself changed by therapy; be able to explain, at least partly, how therapy affects the surrogate outcome and how this affects the clinically relevant outcome; and in the case of a surrogate for drug effects, have a similar dose response for the surrogate and the clinical effects'.[18]

In orthopaedics, frequently used surrogates for outcomes are imaging findings (e.g. osteolysis or tip-apex distance on radiographs), biological markers (e.g. current status of biomarkers for osteoarthritis and their use in approval studies[20]) and biomechanical parameters (e.g. knee instability,[21] gait biomechanics in knee osteoarthritis[22]). The use of surrogates instead of clinical outcomes is common but associated with risks; notably, their validity in predicting future clinically relevant outcomes remains largely untested.[23]

Another group of measures often encountered in the surgical literature and elsewhere are structural and process

measures.[24,25] According to Donabedian,[2] structural measures assess how care was organized whereas process measures assess what was done in the care process and outcome measures what happened to the patient. Examples of structural measures in orthopaedics include hospital and surgeon volume and the presence/absence of an implemented care pathway. Examples of process measures include the proportion of patients receiving prophylactic antibiotics before surgery, admitted with proximal femur fractures operated within a certain time frame, or operated upon with use of computer navigation.

How to measure the impact of an intervention on outcome

This topic has been largely covered by publications and textbooks for the general medical audience and also specifically for the musculoskeletal field.[8-11,26-28] I will thus mainly highlight aspects which in my opinion are important for orthopaedic surgeons and will facilitate the conduct of their research. There are mainly three types of outcome variables: continuous (e.g. scores), categorical (e.g. infection yes/no), and time-to-event data (e.g. time to revision of a hip replacement, time to re-operation after fixation failure of proximal humerus fracture).

Absolute and relative measures of effect

Estimating and comparing the effects of two different treatments (e.g. ORIF *versus* hemiarthroplasty for proximal humerus fracture) or patient characteristics (ever- *versus* never-smokers undergoing hip replacement) on a specific outcome (e.g. post-operative infection) is the aim of many clinical studies. As a consequence, it is crucial to present these effects in the results section and the abstract of the manuscript.[29,30] Absolute and relative effect measures are called measures of association, since they measure or summarize the association of two point estimates. Absolute measures include absolute risk reduction or risk difference, number needed to treat (= inverse of the risk difference), and incidence rate difference. Relative measures are relative risk or risk ratio, relative risk reduction, incidence rate ratio, odds ratio and hazard ratio. All measures of effect need to be presented with confidence intervals to indicate the precision.

Effect measures are mainly risk-based or rate-based measures. Risk-based measures of effect are typically obtained in a study comparing two or more groups with regard to the occurrence of a categorical event/outcome (e.g. infection yes/no) occurring in a relatively short follow-up time. The event can be either an adverse event (= complication, harm, safety concern) or a desirable event (= reduction of a risk of a certain complication, benefit). Absolute and relative effect measures involving risks are easily calculated from the numbers of events and the number of patients at risk in both treatment groups (2×2 table). When the study follow-up is longer and/or patients are lost over the follow-up time due to competing risks such as death, the use of rate-based measures (incidence rate difference or incidence rate ratio) is indicated as illustrated.[31] Moreover, other time-to-event analyses such as survival analysis or Cox regression analysis need to be considered.[32-34]

Absolute and relative measures convey different and complementary information.[29,30,35,36] Absolute measures allow calculating numbers needed to treat (NNT). As an example, risks of a given event of 50% *versus* 25% in two treatment groups correspond to a large risk difference of 25%, a number of patients needed to treat with the low-risk instead of the high-risk treatment of 4, and a relative risk of 2. This is in contrast to a situation in which the risks of a given event in the two treatment groups are 5% *versus* 2.5%. Here, they correspond to a much smaller risk difference of 2.5%, a much larger number of patients needed to treat of 40, but similarly to a relative risk of 2.

Measures of effect, whether absolute or relative, should be presented both unadjusted and – if applicable – adjusted for confounding factors.[37] To take into account the confounding factors, there is a variety of regression models such as multiple logistic or linear regression models, proportional hazard models (Cox) or non-linear multiple regression models,[38] among others.

Patient-reported outcomes

To ask the patient which is/are the most important symptom(s) for her/him is crucial in the process of choosing the outcomes of interest for the study. Patient-reported outcome measures (PROMs) are important tools in clinical care and research. Both generic and disease-specific PROMs should be used since they provide complementary information. The *AO Handbook* provides an extensive overview of both clinician- and patient-reported outcome measures and instruments in the field of musculoskeletal diseases.[39] The topic of PROs has been extensively covered.[40-43] There are also publications specifically on types, selection, interpretation, quality criteria (such as validity, reliability, responsiveness) and pitfalls of PROs in orthopaedics,[5,44,45] as well as an example of their pre- and post-operative use after knee replacement.[46,47] Moreover, retrospectively assessing PROs in emergency admissions may be feasible and relevant, and should be considered.[48]

Typically, PROs are measured on a continuous scale. If applicable, both absolute values (e.g. PRO at baseline and PRO at one year after a health intervention such as hip replacement) and the change value (e.g. difference between baseline and one year) should be reported. There are several ways of assessing whether the observed difference between PRO at baseline and PRO at one year is perceivable and important for the patient. These include, among others, the effect size,[49] the minimal clinically important difference (MCID) and other related metrics,

the patient-acceptable symptom state (PASS)[50] and the categorization into patients who achieved a better, a similar and a worthy PRO result after the health intervention of interest. The publications by Katz et al and Maltenfort et al provide useful information on MCIDs of currently used PROs in the musculoskeletal field.[51,52] Finally, many PROs are constructed in a way that allows obtaining both summary scores as well as sub-scores for specific domains such as pain, function, physical activity or other.

In clinical care, physicians evaluate an individual patient's score, whereas in clinical research or public health, population-based average scores are assessed. The optimal way of interpreting and presenting PROs is still not sufficiently well-known in either situation. However, experience with these outcome tools is rapidly evolving.[43,53]

Target difference and sample size calculation

Once the outcome(s) have been chosen, the next step is to determine the sample size necessary for the study. To be able to do so, the researcher needs to specify the target difference. The target difference is the 'difference in the primary outcome that the study is designed to detect reliably'.[54] The researcher intending to perform a randomized controlled trial (RCT), which compares the risk of dislocation within the first six months after two types of surgical approaches for hip replacement, needs to anticipate the risk of dislocation in both groups and derive the target difference. Cook et al describe two bases for determining the target difference: one is the difference that is considered important by the stakeholders; the other is the 'realistic' difference based on available difference estimates from the literature.[54]

Sample size considerations are particularly relevant for RCTs, where the researcher needs to minimize the number of patients exposed to the experimental character of this type of study and to contain the costs. It is also relevant for observational studies but for cost reasons, whereas it is much less an issue in registries (e.g. national hip replacement registry) or other large database studies (e.g. national inpatient sample in the US).[55] However, in the planning phase of every study it is essential to reflect on the outcome difference that is perceivable and important to the patient and/or relevant for public health. Finally, the relationship between a statistically significant difference and an important difference is clearly explained in the publications by Ranstam and Cook.[30,56] Further relevant readings on the use or non-use of p-values are provided by Wasserstein et al,[57] and on big data and p-values by Kaplan et al.[58]

How to present outcomes in the paper

Chan et al[59] have recently defined standard protocol items for clinical trials, which also apply to observational studies, as follows: 'Primary, secondary, and other outcomes, including the specific measurement variable (e.g. systolic blood pressure), analysis metric (e.g. change from baseline, final value, time to event), method of aggregation (e.g. median, proportion), and time point for each outcome should be presented. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended.' Outcome presentation is also part of reporting guidelines, such as CONSORT for clinical trials or STROBE for observational studies,[60] and other orthopaedics-specific publications.[26,56] All clinically relevant outcomes that have been assessed in the study need to be reported, whether the result is negative or positive. Selective outcome reporting (e.g. choosing to indicate only the statistically significant findings) can lead to incomplete and biased results as shown for RCTs,[61] especially to overestimation of treatment effects due to the non-reporting of negative findings.

Graphical representation of the main finding(s) is often desirable and helps to underline their importance and improve their understanding (e.g. survival curves frequently presented in orthopaedic papers). Continuous outcomes compared with categorical outcomes are less intuitively understood when only presented in tables. Cool et al[62] have summarized graphical presentations of outcome data in orthopaedics including box plots, histograms and scatter plots. When the sample size is small, the use of dot plots is recommended.[56] When the sample size is large, Kernel density estimation is a useful non-parametric technique for visualizing the underlying distribution of a continuous variable.[63,64]

Bias in outcome determination

All research studies are susceptible to random error and systematic error, albeit to a different degree, and to different types of errors.[7,9,37,65] Random error or error due to chance can be reduced by increasing the sample size of the study, whereas systematic error cannot be eliminated this way. Systematic error, also called bias, can occur in patient selection (selection bias), in the measurement of the study variables such as the outcome, in patient follow-up (attrition bias, non-responder bias) or through insufficient controlling for confounding factors.[7] RCTs and observational studies can both be afflicted by random error. However, systematic error (especially selection bias and confounding) is a particular problem in observational studies. A common bias affecting the outcome is the measurement or information bias. Detection bias (= systematic differences between the groups in how outcomes are determined; blinding (or masking) of outcome assessors can reduce the risk) and recall bias (= differences in the accuracy or completeness of the patient's memory of past events, leading to overestimation or underestimation) are types of information bias.

The internal validity of a study depends on the degree of systematic error present. Thus, bias minimization is crucial.

Table 2. Key questions

Think about:
Is the outcome important for the patient?
Is the outcome relevant for clinical care and public health?
Are both efficacy/effectiveness AND safety outcomes included?
Is the whole care cycle represented in the choice of outcomes?
Is improvement achievable? / Can we really get better?
Are your data able to show this (target difference that matters, sample size sufficient, comparator group(s) included)?
Are surrogate outcomes used and how robust is the association with a clinically relevant outcome?
Is the outcome instrument (PRO) appropriate and sensitive to change?
Is accurate outcome measurement realistically achievable (time, money, staff)?
Are the outcomes clearly presented (absolute measures of effect included) and their clinical relevance outlined?
Have relevant potentially confounding factors been considered?
May the results be due to bias in outcomes assessment?
Are the reported outcomes generalizable?

Conclusions

The outcome is at the centre of almost every research question and the clinician's expertise is crucial. She or he should understand and guide the reflections surrounding outcomes in collaboration with the patient, other stakeholders and epidemiologists/statisticians (Table 2). Key messages are:

1) For all research projects, it is strongly advised to perform a thorough review of the literature and other public resources and to involve a statistician in the design, analysis and interpretation phase of a study.
2) Reflect on the outcome and the expected (target) difference that is perceivable and important to the patient and/or relevant for public health in the study design phase.
3) An important difference is not the same as a statistically significant difference.
4) Evaluate both benefit and harm of a health intervention.
5) Include short- and longer-term outcomes.
6) Evaluate complementary types of outcomes measures and include whenever possible PROs (general health and disease-specific).
7) Present measures of effects with confidence intervals in results and abstract. Absolute measures are often more informative.
8) Spend time on choosing the best way to graphically present the main outcome(s).
9) Bias in outcome measurement is a threat to the internal validity of the study.

AUTHOR INFORMATION
Division of Orthopaedic Surgery and Traumatology, Geneva University Hospitals, Switzerland; Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK.

Correspondence should be sent to: A. Lübbeke, Division of Orthopaedic Surgery and Traumatology, Geneva University Hospitals, 4, Rue Gabrielle-Perret-Gentil, 1205 Geneva, Switzerland.
Email: anne.lubbekewolff@hcuge.ch

ICMJE CONFLICT OF INTEREST STATEMENT
None declared.

FUNDING STATEMENT
No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

REFERENCES

1. **Porter ME.** What is value in health care? *N Engl J Med* 2010;363:2477-81.
2. **Donabedian A.** Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;44 (Suppl.):166-206.
3. **Wilson IB, Cleary PD.** Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;273:59-65.
4. **Epstein AM.** The outcomes movement – will it get us where we want to go? *N Engl J Med* 1990;323:266-70.
5. **Gagnier JJ.** Patient reported outcomes in orthopaedics. *J Orthop Res* 2017;35:2098-108.
6. **Sips ME, Bonten MJM, van Mourik MSM.** Semiautomated surveillance of deep surgical site infections after primary total hip or knee arthroplasty. *Infect Control Hosp Epidemiol* 2017;38:732-5.
7. **Rothman K.** *Epidemiology: an introduction*. Second ed. Oxford: Oxford University Press, 2012.
8. **de Moraes VY, Ferrari PM, Gracitelli GC, Faloppa F, Belloti JC.** Outcomes in orthopedics and traumatology: translating research into practice. *Acta Ortop Bras* 2014;22:330-3.
9. **Kocher MS, Zurakowski D.** Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. *J Bone Joint Surg [Am]* 2004;86-A:607-20.

10. Bhandari M, Petrisor B, Schemitsch E. Outcome measurements in orthopedic. *Indian J Orthop* 2007;41:32-6.

11. Kooistra BW, Sprague S, Bhandari M, Schemitsch EH. Outcomes assessment in fracture healing trials: a primer. *J Orthop Trauma* 2010;24(Suppl. 1):S71-5.

12. Agency for Healthcare Research and Quality. *Outcome measures framework: literature review, findings and implications. AHRQ publication no. 16*. Rockville, MD: Agency for Healthcare Research and Quality, 2016.

13. World Health Organization. *International classification of functioning, disability and health*. ICF, 2017. http://www.who.int/classifications/icf/en/ (date last accessed 14 March 2018).

14. Ashby E, Grocott MP, Haddad FS. Outcome measures for orthopaedic interventions on the hip. *J Bone Joint Surg [Br]* 2008;90-B:545-9.

15. Velentgas P, Dreyer NA, Wu AW. Outcome definition and measurement. In: Velentgas P, Dreyer NA, Nourjah P, et al., eds. *Developing a protocol for observational comparative effectiveness research a user's guide*. Rockville, MD: Agency for Healthcare Research and Quality, 2013:71-92.

16. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage* 2013;21:1042-52.

17. Grimm B, Bolink S. Evaluating physical function and activity in the elderly patient using wearable motion sensors. *EFORT Open Rev* 2017;1:112-20.

18. Rigatto C, Barrett BJ. Biomarkers and surrogates in clinical studies. *Methods Mol Biol* 2009;473:137-54.

19. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89-95.

20. Bay-Jensen AC, Reker D, Kjelgaard-Petersen CF, et al. Osteoarthritis year in review 2015: soluble biomarkers and the BIPED criteria. *Osteoarthritis Cartilage* 2016;24:9-20.

21. Amano K, Li Q, Ma CB. Functional knee assessment with advanced imaging. *Curr Rev Musculoskelet Med* 2016;9:123-9.

22. Kumar D, Manal KT, Rudolph KS. Knee joint loading during gait in healthy controls and individuals with knee osteoarthritis. *Osteoarthritis Cartilage* 2013;21:298-305.

23. Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. *BMJ* 2011;343:d7995.

24. Auerbach A. Healthcare quality measurement in orthopaedic surgery: current state of the art. *Clin Orthop Relat Res* 2009;467:2542-7.

25. Birkmeyer JD, Dimick JB, Birkmeyer NJ. Measuring the quality of surgical care: structure, process, or outcomes? *J Am Coll Surg* 2004;198:626-32.

26. Dorey F, Amstutz HC. Discrepancies in the orthopaedic literature: why? A statistical explanation. *Instr Course Lect* 1993;42:555-64.

27. Petrie A. Statistics in orthopaedic papers. *J Bone Joint Surg [Br]* 2006;88:1121-36.

28. Greenfield ML, Kuhn JE, Wojtys EM. A statistics primer. *Am J Sports Med* 1996;24:393-5.

29. Rothman K. Measuring disease occurrence and causal effects. In: Rothman K, ed. *Epidemiology: an introduction*. Second ed. Oxford: Oxford University Press, 2012:38-68.

30. Cook JA, Ranstam J. Statistical analyses that provide an effect size are to be preferred. *Br J Surg* 2016;103:1365.

31. Lübbeke A, Rothman KJ, Garavaglia G, et al. Strong association between smoking and the risk of revision in a cohort study of patients with metal-on-metal total hip arthroplasty. *J Orthop Res* 2014;32:762-8.

32. Ranstam J, Cook JA. Kaplan-Meier curve. *Br J Surg* 2017;104:442.

33. Sedgwick P. How to read a Kaplan-Meier survival plot. *BMJ* 2014;349:g5608.

34. Columbia University Mailman School of Public Health. Time-to-event data analysis. https://www.mailman.columbia.edu/research/population-health-methods/time-event-data-analysis (date last accessed 7 March 2018).

35. Vetter TR, Jesser CA. Fundamental epidemiology terminology and measures: it really is all in the name. *Anesth Analg* 2017;125:2146-51.

36. Jaeschke R, Guyatt G, Shannon H, et al. Basic statistics for clinicians: 3. Assessing the effects of treatment: measures of association. *CMAJ* 1995;152:351-7.

37. Cook JA, Ranstam J. Statistical models and confounding adjustment. *Br J Surg* 2017;104:786-7.

38. Kleinman LC, Norton EC. What's the risk? A simple approach for estimating adjusted risk measures from nonlinear models including logistic regression. *Health Serv Res* 2009;44:288-302.

39. Suk M, Hanson B, Norvell D, Helfet D. *Musculoskeletal outcomes measures and instruments*. Second ed. Stuttgart: Thieme, 2009.

40. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *N Engl J Med* 1996;334:835-40.

41. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2-assessing respondent understanding. *Value Health* 2011;14:978-88.

42. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1-eliciting concepts for a new PRO instrument. *Value Health* 2011;14:967-77.

43. Greenhalgh J, Dalkin S, Gooding K, et al. *Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care*. Southampton, UK: Health Services and Delivery Research, 2017.

44. Poolman RW, Swiontkowski MF, Fairbank JC, et al. Outcome instruments: rationale for their use. *J Bone Joint Surg [Am]* 2009;91(Suppl. 3):41-9.

45. Ramkumar PN, Harris JD, Noble PC. Patient-reported outcome measures after total knee arthroplasty: a systematic review. *Bone Joint Res* 2015;4:120-7.

46. Ayers DC. Implementation of patient-reported outcome measures in total knee arthroplasty. *J Am Acad Orthop Surg* 2017;25(Suppl. 1):S48-S50.

47. Walker R, Gough AT, Williams DH. Patient-reported outcome measures (PROMs): enhancing decision making and follow-up. *BMJ Case Rep* 2017;2017:bcr-2017-221172.

48. Kwong E, Black N. Retrospectively patient-reported pre-event health status showed strong association and agreement with contemporaneous reports. *J Clin Epidemiol* 2017;81:22-32.

49. Cohen D. *Statistical power analysis for the behavioral sciences*. Second ed. Hillsdale, NJ: Erlbaum, 1988.

50. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Ann Rheum Dis* 2007;66(Suppl. 3):iii40-1.

51. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *J Orthop Surg* 2015;10:24.

52. Maltenfort M, Díaz-Ledezma C. Statistics in brief: minimum clinically important difference-availability of reliable estimates. *Clin Orthop Relat Res* 2017;475: 933-46.

53. Raine R, Fitzpatrick R, de Pury J. Challenges, solutions and future directions in evaluative research. *J Health Serv Res Policy* 2016;21:215-6.

54. Cook JA, Hislop J, Altman DG, et al; DELTA group. Specifying the target difference in the primary outcome for a randomised controlled trial: guidance for researchers. *Trials* 2015;16:12.

55. Alluri RK, Leland H, Heckmann N. Surgical research using national databases. *Ann Transl Med* 2016;4:393.

56. Ranstam J, Cook JA. Considerations for the design, analysis and presentation of in vivo studies. *Osteoarthritis Cartilage* 2017;25:364-8.

57. Wasserstein A, Lazar N. The ASA's statement on p-values: context, process, and purpose. *Am Stat* 2016;70:129-33.

58. Kaplan RM, Chambers DA, Glasgow RE. Big data and large sample size: a cautionary note on the potential for bias. *Clin Transl Sci* 2014;7:342-6.

59. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;158:200-7.

60. Simera I, Moher D, Hoey J, Schulz KF, Altman DG. A catalogue of reporting guidelines for health research. *Eur J Clin Invest* 2010;40:35-53.

61. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.

62. Cool P, Ockendon M. The use of plots in orthopaedic literature. *Bone Joint J* 2015;97-B:1593-603.

63. Rosenblatt M. Remarks on some nonparametric estimates of a density function. *Ann Math Stat* 1956;27:832-7.

64. Franklin PD, Li W, Ayers DC. The Chitranjan Ranawat Award: functional outcome after total knee replacement varies with patient attributes. *Clin Orthop Relat Res* 2008;466:2597-604.

65. Lambert J. Statistics in brief: how to assess bias in clinical studies? *Clin Orthop Relat Res* 2011;469:1794-6.