Research methodology for orthopaedic surgeons, with a focus on outcome

LUEBBEKE-WOLFF, Anne

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

Reference
LUEBBEKE-WOLFF, Anne. Research methodology for orthopaedic surgeons, with a focus on outcome. EFORT Open Reviews, 2018, vol. 3, no. 5, p. 160-167

DOI : 10.1302/2058-5241.3.170064
PMID : 29951252
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. 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). 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). 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. 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, 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...
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’. 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 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 Assessment (COMET), the European Clinical Research Infrastructures Network (ECRIN) Database, the International Consortium for Health Outcomes Measurement (ICHOM), as well as the core sets of domains from the World Health Organization’s International Classification of Functioning, Disability and Health. 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; Outcome assessment in fracture healing trials: a primer) as well as an excellent overview for researchers and clinicians of outcome definition and measurement in observational comparative effectiveness research.

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); (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...
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 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’. 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’. 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 ‘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’. 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) and biomechanical parameters (e.g. knee instability, gait biomechanics in knee osteoarthritis). 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.

Another group of measures often encountered in the surgical literature and elsewhere are structural and process
measures. According to Donabedian, 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. 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 (e.g. 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. 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 ($2 \times 2$ 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. Moreover, other time-to-event analyses such as survival analysis or Cox regression analysis need to be considered.

Absolute and relative measures convey different and complementary information. 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. 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, 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. The topic of PROs has been extensively covered. There are also publications specifically on types, selection, interpretation, quality criteria (such as validity, reliability, responsiveness) and pitfalls of PROs in orthopaedics, as well as an example of their pre- and post-operative use after knee replacement. Moreover, retrospectively assessing PROs in emergency admissions may be feasible and relevant, and should be considered.

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, the minimal clinically important difference (MCID) and other related metrics.
the patient-acceptable symptom state (PASS)\textsuperscript{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.\textsuperscript{51,52} Finally, many PROs
are constructed in a way that allows obtaining both sum-
mary 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 suffi-
ciently well-known in either situation. However, experience
with these outcome tools is rapidly evolving.\textsuperscript{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 reli-
ably’. The researcher intending to perform a randomized
controlled trial (RCT), which compares the risk of disloca-
tion 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 dif-
fERENCE. 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.\textsuperscript{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 replace-
ment registry) or other large database studies (e.g. national
inpatient sample in the US).\textsuperscript{55} However, in the planning
phase of every study it is essential to reflect on the out-
come difference that is perceivable and important to the
patient and/or relevant for public health. Finally, the rela-
tionship between a statistically significant difference and
an important difference is clearly explained in the publica-
tions by Ranstam and Cook.\textsuperscript{30,56} Further relevant readings
on the use or non-use of p-values are provided by Wasser-
stein et al,\textsuperscript{57} and on big data and p-values by Kaplan et al.\textsuperscript{58}

**How to present outcomes in the paper**

Chan et al\textsuperscript{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 recom-
mented.’ Outcome presentation is also part of reporting
guidelines, such as CONSORT for clinical trials or STROBE
for observational studies,\textsuperscript{60} and other orthopaedics-specific
publications.\textsuperscript{56,58} All clinically relevant outcomes that have
been assessed in the study need to be reported, whether
the result is negative or positive. Selective outcome report-
ing (e.g. choosing to indicate only the statistically signifi-
cant findings) can lead to incomplete and biased results as
shown for RCTs,\textsuperscript{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 fre-
cently presented in orthopaedic papers). Continuous
outcomes compared with categorical outcomes are less
intuitively understood when only presented in tables. Cool et al\textsuperscript{62} have summarized graphical presentations of
outcome data in orthopaedics including box plots, histo-
grams and scatter plots. When the sample size is small,
the use of dot plots is recommended.\textsuperscript{54} When the sample
size is large, Kernel density estimation is a useful non-
parametric technique for visualizing the underlying distri-
bution of a continuous variable.\textsuperscript{63,64}

**Bias in outcome determination**

All research studies are susceptible to random error and
systematic error, albeit to a different degree, and to differ-
ent types of errors.\textsuperscript{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 (attri-
 tion bias, non-responder bias) or through insufficient con-
trolling for confounding factors.\textsuperscript{7} RCTs and observational
studies can both be afflicted by random error. However,
systematic error (especially selection bias and confound-
ing) is a particular problem in observational studies. A
common bias affecting the outcome is the measurement
or information bias. Detection bias (= systematic differ-
ences between the groups in how outcomes are deter-
mined; blinding (or masking) of outcome assessors can
reduce the risk) and recall bias (= differences in the accu-
racY 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.
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.

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?

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.

LICENSE
© 2018 The author(s)
This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) licence (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed.

REFERENCES


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


33. Sedgwick P. How to read a Kaplan-Meier survival plot. BMJ 2014;349:g6688.


50. Kviën 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):i140-1.


