

Levels of evidence in orthopaedics

Rob Middleton,¹ Andrew Price²

In modern surgical practice, the use of evidence based medicine (EBM) to support the delivery of clinical care has become increasingly prevalent. EBM is an approach to decision-making in which a practitioner considers the available evidence and its quality, including the potential for bias, in answering a specific clinical question. Alongside the individual clinician, EBM is used by institutions, public health agencies and regulatory agencies in the creation of protocols, guidelines and approval of interventions on a societal level. As a result, the ability to critically appraise the available evidence is a crucial skill for those in the healthcare arena. *Journal of ISAKOS* is an exciting new journal that has an important role in presenting evidence to the global Orthopaedic community. In doing this it is important to reflect on the levels of evidence available and their relative merits.

As part of the EBM process, there is a concept of ordering evidence according to its perceived 'quality', based on study design and potential for bias. This has resulted in a 'hierarchy of evidence', with higher quality sources of evidence at the apex and those more prone to bias at the base. Since the late 1970s, a number of different hierarchies have been published, as well as tools to rate the quality of evidence of a particular report. These have been introduced in response to concerns that the basic hierarchy did not reflect the varying quality of evidence at each level. Across these various hierarchies the overall order is reasonably consistent, with systematic reviews and meta-analysis at the top and expert opinion at the bottom.

It is important to understand the different types of evidence that make up the hierarchy of evidence. Each type has particular strengths and weaknesses, and so each may be more suitable to answering a particular clinical question than others, and this should be appreciated when assessing the evidence base. Discussed below are some of these types of evidence

and situations where they have been applied.

Systematic reviews (SR) are a structured and critical appraisal of the available evidence to answer a specific clinical question. The clinical question and review methodology are determined *a priori*. Meta-analysis (MA) refers to the aggregation of data from multiple studies and subsequent quantitative analysis of results. SRs and MAs occupy the top level of the hierarchy of evidence and represent a comprehensive answer to a given clinical question. SRs include an assessment of the quality of evidence of available studies and a summary of the overall results. Inclusion of MA involves quantitative analysis in answering the given question to summarise the available data and to present a numerical effect size for the exposure in question. Given the rate at which new research data is published and the time required to critically appraise each study individually, SRs are a powerful tool to summarise available evidence for use by clinicians, when properly performed. There exist guidelines on both the production of a SR, as well as in appraising a given SR for quality. A SR and MA of multiple randomised controlled trials with homogeneity is currently considered the top of the hierarchy of evidence.

A randomised controlled trial (RCT) is an experimental study in which participants are enrolled with explicitly stated inclusion/exclusion criteria and are subsequently randomly allocated to a treatment or non-treatment or alternative treatment group. As part of the process to minimise bias within the study, blinding is an important consideration, and may be of the participant, the individual delivering treatment, the individuals analysing the data, and those responsible for interpreting and reporting the results.¹ Comparisons are then made according to a set of defined outcomes of interest. There are a number of RCT study designs, descriptions of which are beyond the scope of this editorial.²⁻⁴

The advantages of a well-designed RCT, with appropriately performed enrolment, randomisation and blinding are numerous.^{2 5 6} RCTs provide the strongest evidence of a causal relationship between the intervention of interest and measured outcome. Appropriate randomisation of participants prior to treatment ideally results in an unbiased distribution

of potentially confounding variables between study groups, reducing the risk of systemic bias between the groups.⁷ The use of blinding as described above is an important factor in minimising observer bias.

However, while RCTs can generate robust evidence, they are not without difficulties. The development and delivery of an RCT can be a financially expensive undertaking, particularly where the expected effect size is small (thus requiring large numbers of participants) or where the outcomes of interest require a long study follow-up period. Appropriate sample size calculations must be part of an RCT's design to minimise the risk of an inconclusive study outcome (or a Type II error), while limiting the number of participants exposed to potential risk.⁸ RCTs can also pose ethical challenges, for example, the use of a placebo arm potentially exposes patients to harm the treatment group is not at risk of.⁹ In surgical RCTs involving 'sham' surgery, individuals in the control group are exposed to the risks of surgery, without the potential for treatment benefit.¹⁰

While RCTs can be an effective method to limit potential sources of bias, they are not immune. Volunteer bias is one such problem, in which those who volunteer to be included in an RCT differ from those declining (or differences in retention during the study period). This introduces a type of systematic bias to the study and can render the study results non-generalizable to the population originally of interest. In surgical settings it is often not possible to run a fully blinded study, introducing the risk of observer bias. It is simply not possible to blind the surgeon to the intervention — and potentially the patient or care team.¹ The challenges in delivering RCTs in surgery are highlighted by Solomon and McLeod, who estimated that only 40% of surgical treatment questions could be addressed in this way.¹¹ Despite these challenges, when appropriately designed, RCTs can be used to answer surgical questions.

A cohort study is a type of observational longitudinal study in which a group of participants (the 'cohort'), with an exposure or intervention of interest, are followed up over time and the incidence of events measured. Typically cohort studies are used to monitor outcomes of interest that may be rare or require a long follow-up period. Cohort studies may be performed prospectively or retrospectively.

In common with RCTs, cohort studies can determine the timing of outcomes,

¹Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, University of Oxford Nuffield, Oxford, UK

²Nuffield Orthopaedic Centre, Oxford University, Oxford, UK

Correspondence to Dr Andrew Price, Nuffield Orthopaedic Centre, Oxford University, Oxford OX3 7HE, UK; andrew.price@ndorms.ox.ac.uk

and as such, a degree of causal evidence. However, as an observational study this evidence is not as strong as that from a RCT. Cohort studies also avoid the ethical challenges presented by RCTs, as there is no placebo group or sham treatment group for example. The delivery of a cohort study is also less challenging than a RCT, particularly retrospective cohort studies, with regards to time and cost. Prospective cohort studies do, however, require significant infrastructure where large numbers of patients are followed up over many years. It is also possible to investigate multiple outcomes after a given intervention of interest.

Conversely, cohort studies lack some of the advantages of RCTs. The participants in the study are not blinded and are not randomised, presenting possible sources of bias. Lack of randomisation increases the risk of confounding variables impacting on the outcome. Where confounders are recognised, attempts can be made to control for them, but hidden confounders may remain. As mentioned above, for the investigation of rare outcomes, large numbers of participants, or long follow-up period, is required.

A case-control study is a type of observational retrospective study. A case group of individuals with the outcome of interest are compared with a control group without the outcome of interest. The past exposures of the two groups are then compared with identify potential causative factors.¹²

Case-control studies offer advantages over RCTs or cohort studies, as by definition, case-control studies are retrospective, faster to perform, and as a result cheaper. It is also a more appropriate study design for investigating outcomes that are very rare or with long lag times. The ethical considerations are less than for RCTs or cohort studies, as the study does not impart any risk to the participants beyond that from reviewing their medical history or responding to a questionnaire.

However, case-control studies exist lower in the hierarchy of evidence for a number of reasons. The lack of randomisation can result in an unequal distribution of confounding factors between the groups, despite attempts at matching the groups. There is also a risk of bias (for example, recall, selection or survival bias). Unlike

RCTs and cohort studies, case-control studies cannot determine causality, prevalence or incidence of the outcome. Also in contrast to the previous study types, case-control studies cannot investigate multiple outcomes of interest for a given exposure, but rather the opposite (one outcome of interest may highlight multiple potential causative exposures).

Case reports and case series are considered of lesser quality than the studies above and report the outcomes of a given population after exposure or intervention. They lack a comparator group and so are at greater risk of bias. Case reports are a description of a single clinical problem, its treatment and subsequent outcome. A case series is similar, except that it consists of a number of patients treated in a similar way, and the group outcomes reported may be retrospective or prospective. They can be used to report outcomes of new interventions or rare treatment options and to provide baseline data prior to the undertaking of a formal comparative study. They are also significantly less difficult to undertake than these larger study types.

Expert opinion lies at the bottom of the hierarchy of evidence and represents the knowledge of specialists in the relevant field to the question of interest. Such opinions are usually based on the clinical experience of said specialist, previously published texts and mechanistic reasoning based on laboratory studies. There is no case group, or control group, and as such there is a significant risk of bias. However, such reports can lead to research questions to be addressed by higher level studies.

Hierarchies of evidence have been created to aid in the interpretation of available evidence in addressing a clinical question. These hierarchies broadly order sources of evidence according to the validity of their results and their potential for bias. However, each level of evidence has its own strengths and weaknesses. Some clinical questions simply cannot be addressed by a particular method, and as such the highest level of evidence available may not be the highest possible in the hierarchy. Furthermore, the quality of studies is an important consideration during the EBM process, rather than simply their level in the hierarchy. The results of a poor-quality RCT do not outweigh the

results of a high-quality cohort study for example, and as such the use of appraisal tools in assessing study quality is encouraged.

Competing interests None declared.

Provenance and peer review Commissioned; internally peer reviewed.

Author note *Journal of ISAKOS* has an important role in presenting all levels of evidence to its members and beyond. The importance of reflecting all levels of evidence available is critically important as we support the evidence base for Orthopaedics.

© International Society of Arthroscopy, Knee Surgery and Orthopaedic Sports Medicine (unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.



To cite Middleton R, Price A. *JISAKOS* 2017;2:295–296.

JISAKOS 2017;2:295–296.
doi:10.1136/jisakos-2017-000188

REFERENCES

- Karanicolas P, Farrokhyar F, Bhandari M. Blinding: Who, what, when, why, how? *Can J Surg* 2010;53:345–8.
- Levin KA. Study design VII. Randomised controlled trials. *Evid Based Dent* 2007;8:22–3.
- Whelan DB, Dainty K, Chahal J. Efficient designs: factorial randomized trials. *J Bone Joint Surg Am* 2012;94:34–8.
- Wellek S, Blettner M. Vom richtigen Umgang mit dem Crossover-Design in klinischen Studien: Teil 18 der Serie zur Bewertung wissenschaftlicher Publikationen. *Dtsch Arztebl Int* 2012;109:276–81.
- Kendall JM. Designing a research project: randomised controlled trials and their principles. *Emerg Med J* 2003;20:164–8.
- Mundi R, Chaudhry H, Mundi S, et al. Design and execution of clinical trials in orthopaedic surgery. *Bone Joint Res* 2014;3:161–8.
- Altman DG, Bland JM. Treatment allocation in controlled trials: why randomize? *Br Med J* 1999;318:1209.
- Zhong B. How to calculate sample size in randomized controlled trial? *J Thorac Dis* 2009;1:51–4.
- Savulescu J, Wartolowska K, Carr A. Randomised placebo-controlled trials of surgery: ethical analysis and guidelines. *J Med Ethics* 2016;42:776–83.
- Wolf BR, Buckwalter JA. Randomized surgical trials and “sham” surgery: relevance to modern orthopaedics and minimally invasive surgery. *Iowa Orthop J* 2006;26:107–11.
- Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery* 1995;118:459–67.
- Busse JW, Obrebsky WT. Principles of designing an orthopaedic case-control study. *J Bone Joint Surg Am* 2009;91:15–20.