



Archna Suchak

Ama Johal and Angie Wade

Useful Concepts for Critical Appraisal: 1. Study Design

Abstract: There is an increasing volume of research undertaken within orthodontics and with this comes a need to evaluate what is available. This short series of three articles aims to help the orthodontist revise basic concepts of critical appraisal and pertinent statistics.

Clinical Relevance: Critical appraisal skills are valuable tools that can aid clinical decision-making. In this first of three articles, we discuss different study designs.

Ortho Update 2012; 5: 57–60

The basics

What is critical appraisal?

Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness, and its value and relevance in a particular context.¹ Critical appraisal skills can be applied to almost any published research study. There are many resources available to assist with critical appraisal including a variety of checklists. However, for a checklist to be effective, one has to have a good understanding of the points to be considered. Therefore, an understanding of the research process, the rationale, strengths and weaknesses of different study designs and statistical and clinical inference are all essential to enable the reader to appraise effectively.

The research question

Scientific research usually involves undertaking a methodological study in order to answer a specific question. The introduction of a study should ideally summarize the literature, highlighting any shortcomings and clearly defining the questions that the study is intending

to answer (its aims). These may include a hypothesis that will be tested, such as whether a treatment is effective (Box 1).

Validity

Assessment of a piece of research should always consider its validity, which can be split into two types:

- Internal validity – the degree to which results of a study are likely to approximate to an accepted truth. In other words, did the study properly assess what it intended to?
- External validity – this relates to generalizability, the extent to which the effects observed are applicable to the outside world. It is useful to look at the paper's inclusion/exclusion criteria to help judge this (Box 2).

Factors that can influence the interpretation of the results

Bias

Bias may be defined as a systematic (regular) difference between the results from the study and the true state of affairs. This is a very important concept to consider when critically appraising a paper. There are many different types of bias, which

include selection bias, measurement bias and publication bias. It can invalidate the study results if it is not taken into account properly when drawing conclusions.

The *Hawthorne effect* is a type of bias in which subjects in a study change their behaviour, usually positively, purely in response to being observed. Some types of subjects may be more likely to be *lost to follow-up* and this can also bias the results.

Confounding

The word 'confounding' originates from the latin 'confundere' which means 'to mix together'. Confounding factors are those which make it appear as though there is a direct relationship between the exposure and the outcome (positive confounding factors) or which mask an association which is truly present (negative confounding factors): that is, they 'get in the way' of the comparison between groups that are being investigated. Confounding occurs when there are differences in factors that we are *not interested in directly* between the groups being compared that also happen to affect the outcome we are interested in assessing (Box 3).

Archna Suchak, BSc(Hons), BDS(Hons), MFDS, MSc, MOrth RCS, FOrth RCS, Locum Consultant Orthodontist, Great Ormond Street Hospital, London, **Ama Johal**, BDS(Hons), MSc, PhD, FDS MOrth, FDS(Orth) RCS, Senior Lecturer/Honorary Consultant Orthodontist, Queen Mary University of London, Barts and the London School of Medicine and Dentistry, London and **Angie Wade**, BSc, MSc, PhD, CStat ILTM, Senior Lecturer in Medical Statistics, Institute of Child Health, University College London, London, UK.

A useful acronym that can be applied to help assess if a research question is well-defined is 'PICO':

- Population: What is/what are the group(s) being investigated? What is the patient or problem?
 - Intervention: What interventions are being considered?
 - Comparison: What is the intervention being compared to? Or who are the diseased individuals compared to (usually healthy controls)?
 - Outcome: What is the outcome measure? What are the aims trying to establish differences in?
- Note that a good question does not always need to have all the P,I,C and O elements, but it is useful to apply the acronym to determine whether there are four elements and to be clear of what these are.

For example, the acronym can be applied to an investigation of plaque scores (outcome measure) of 10 and 11 year-old schoolchildren (population), comparing those who have received oral hygiene instruction at a school dental visit (intervention) with those who have not yet received it (comparison).

Box 1. Defining a research question.

A study may wish to determine whether mandibular distraction osteogenesis is effective in reducing the overjet of patients with Pierre-Robin sequence and a Class II malocclusion. We may have a very well-designed and conducted study based on a sample of patients presenting to a tertiary referral centre. This study may have internal validity but not external (we can validly infer the effect of treatment for patients from a small subgroup of patients with Pierre-Robin sequence but the results are not valid externally for all patients with a Class II malocclusion). Alternatively, note that a random sample may be totally representative but the study may have flaws which invalidate the results and hence would not be internally valid.

Box 2. Example of an assessment of a piece of research considering its validity.

Approaches to reduce bias and confounding

It is helpful to be aware of ways that researchers have tried to reduce bias and confounding at the design stage or in the analyses by:

- Matching or stratification of study participants between groups (such as treatment/control or diseased/healthy) according to potential confounders;
- Randomization;
- Blinding;
- Allocation concealment;
- Use of placebo controls;
- Sham treatment periods, so that those in the intervention group do not receive more attention.

Study designs

Classifying the study

There are many different types of study design and the appropriate one to use depends on the type of the research question being posed. Some designs yield more valid results than others. Sometimes the ideal design cannot be undertaken because of a lack of resources, ethical or time restrictions. Even if a study is not optimal for a stated research question it can often still provide valuable information which will help to

answer the question posed.

Published checklists to aid critical appraisal are generally categorized according to study type. Hence, since the paper may not explicitly state the study type (or it may even be wrongly labelled), it is important to be able to distinguish different types of study design to select the correct checklist to use.

Although there are several ways to classify studies, it is often helpful to consider how the study is designed with respect to time. A study may be *prospective* (often considered advantageous as it can be planned, looking forward in time) or *retrospective* (looking at past events). A *cross-sectional study* looks at a single timepoint, whilst a *longitudinal* one extends over a period of time.

Although prospective studies are generally given more credence, it is worth noting that retrospective studies can still be valuable, although we do need to be aware of the potential for recall bias. Data that has been collected specifically for the research study and not based on retrospective recall will generally be more reliable.

It is also worth noting that any study may have *both* prospective and retrospective elements.

An *observational* study reports what already exists, but in contrast an *interventional* one can be considered as *experimental*. It is possible to ascribe causality directly from a well-designed experimental study but not from even the most perfect observational study.

Descriptions of study design²

Surveys

A survey describes a snapshot in time and so shows how things are at that timepoint.

Cohort studies

Cohort studies are observational, longitudinal, prospective studies which follow groups over time. Note that case reports and case series may also be regarded as uncontrolled cohort studies. Their prospective nature means that they can be used to assess the true sequence of events and outcomes. Other applications of cohort studies are to measure incidence or risk of disease and assess prognosis or infer (but not prove) causation. Since short durations of follow-up may not be long enough to identify meaningful changes, cohort studies may be costly, requiring individuals to be followed over long time spans. Furthermore, very large sample sizes are required for rare outcomes to be observed. Bias can be introduced in the form of the *healthy entrant effect* since the design requires individuals to be disease-free at the start and hence they may be unrepresentative of the target population. Cohort studies may also be susceptible to the *Hawthorne effect* and bias introduced by losses to follow-up.

Case-control studies

Case-control studies are observational, longitudinal, retrospective studies that can help reveal differences between groups. They are relatively quick, cheap and easy to perform and may be used to study rare conditions. A wide range of risk factors can be investigated in the same study. They can also be used to infer (but not prove) causation. As they are retrospective, there is no loss to follow-up. However, case-control studies are not suitable if the risk factor is rare and are also susceptible to *recall bias*. Furthermore, the onset of the disease may have preceded exposure to a potential risk factor and, since recording is retrospective, this may not be apparent.

Clinical trials

Clinical trials are prospective (therefore can be planned) and they generally evaluate an intervention or therapy. They allow a single variable to be evaluated by comparing groups, usually using an

If we want to compare overjet reduction between patients with Class II malocclusions using two different functional appliances but one group is older, we will not know whether any difference seen in overjet reduction is due to the patient's age or their type of appliance – thus, in this example, age is a confounding factor. Confounding may be avoided at the design stage by selecting patients of similar ages or by accounting for differences in the analysis by adjusting for age in the comparison of overjet reduction between the groups.

Box 3. Example of a confounding factor.

Ethical approval and patient consent are necessary before any trial is carried out, be it observational or experimental. With experimental trials there are more issues to consider and participants should also consent to being randomized if this is planned. They should understand the process whereby randomization will be performed and the possibility that they will obtain only one of two or more different treatments or no treatment. Researchers should not have any preference for either intervention (this is termed equipoise). Baseline information on the group characteristics should be displayed to demonstrate their comparability and assess if the results can be generalized to wider populations. Interim analyses are sometimes advocated to ensure that the trial does not enrol more patients than necessary to show whether or not the treatment is effective. If these are undertaken then they should have been planned prior to trial commencement with agreed unambiguous rules for stopping the trial. Any interim analyses should take into account multiple testing.

Box 4. Planning clinical trials.³

Whatever means of allocation is employed, details of the process should be stated. Off-site allocation is generally considered more secure and less open to abuse. Any software used for allocation and all confounders considered in the process should be noted.

- **Block randomization** ensures that there are equal numbers of patients in each arm, with the blocks comprising equal numbers of individuals who will go into the experimental and control groups. The order of this allocation within the block is randomly generated and a random number sequence is used to choose a particular block for each set of individuals.
- **Stratified randomization** is based on the block technique. If a potential confounding factor (such as age or gender) can be identified at the design stage, the data generated during the study can be separated into strata based on this confounder. This allows the characteristics of the participants to be kept as similar as possible across the study groups. Once these strata are identified, separate block randomization schemes are created for each factor to ensure that the groups are balanced within each stratum.
- **Minimization⁵** is an alternative allocation method that can be used to ensure equal-sized groups with similar distribution of multiple potential confounders. Randomization of each patient is biased towards the group allocation that would result in the greatest overall similarity between groups after randomization of that individual. The technique copes well even if there are many confounders to consider (unlike stratification). Some potential confounders may be given more weight than others and interactions between confounders are also considered. The biasing factor should be stated at the study onset. There is a trade-off between the biasing factor used and the capacity for the next patient allocation to be predicted.

Box 5. Commonly-used randomization techniques.

untreated control group or a group given standard therapy. They may investigate either *efficacy* (the impact of interventions under optimal conditions, showing that internal validity is present) or *effectiveness* (the impact of interventions under ordinary conditions, showing that external validity is present). Randomization is the preferred means of allocating individuals to treatment or control groups as the chances of bias and confounding are reduced. Such trials are ethically constrained if potentially damaging factors are being investigated. They may also be complicated to co-ordinate and therefore relatively costly. However, the gain in validity usually warrants the additional cost and effort where feasible (Box 4).

Important concepts relating to clinical trials

The following terms are in common parlance in the literature and their relevance is explained below:

- **Prospective:** Planning the study before the data is collected means steps can be taken to minimize bias from the outset.
- **Controlled:** Control subjects allow comparison – without them, one cannot be sure that any response is solely due to the treatment. The control may be positive (the standard treatment) or negative (placebo

or absence of treatment). Alternatives are historical controls or non-randomized controls, but both of these are susceptible to bias.

- **Blind:** Blinding individuals enrolled in a study or those involved in patient care or assessment helps to guard against bias as the composition of each treatment group is kept undisclosed.
- **Allocation concealment:** Prior to consent being obtained, the treatment arm to which the individual is to be allocated should not be known by anyone directly involved with the patient or their care to guard against allocation and assessment bias.
- **Parallel:** Parallel study designs describe those in which comparisons are made between groups of individuals over the same time span. They are sometimes known as *between-patient* trials.
- **Crossover:** In crossover trials, each subject acts as its own control by having treatments (active vs placebo or treatment 1 vs treatment 2) in random order, thus allowing *within-individual* comparisons as opposed to *between-individual*. Some potential confounders are therefore controlled for and sample size can be smaller than for a corresponding parallel trial. A washout period may be needed between the treatments in a

crossover trial to avoid the risk of any carry-over effects. Statistical analysis should be for paired data and should also account for order and carry-over effects. Crossover trials are sometimes known as *within-patient* trials. An enhanced design is to assess two or more different treatments at the same time within the same individual. In orthodontics, where there are differing areas of the mouth to be treated, this is often possible. In such cases, the crossover is usually done diagonally to balance out upper/lower and left/right effects.

- **Randomized:** Randomization helps ensure that both known and unknown confounders are equally distributed between treatment groups, so patients have an equal chance of receiving either treatment independent of their personal characteristics. A true random sequence (compared to quasi-random, which is allocation in some sort of systematic way such as by case note number or date of birth) is ideal (Box 5).

- **Intention to treat:⁴** Loss to follow-up is important as it can lead to errors in the interpretation of results. A follow-up of less than 80% of participants is generally considered unacceptable, although the bias introduced by non-random losses is more important than the percentage lost overall. The *intention to treat* principle protects randomization and minimizes any attrition bias. All subjects are analysed according to their original group regardless of what treatment they actually received. Confounding factors thus remain equally distributed and a realistic picture of the effectiveness of the intervention is given. The opposite is a *per-protocol analysis*, where only data from those who sufficiently complied with trial protocol is used. It is important to know the reasons for any loss to follow-up: the clinical and demographic characteristics of those lost should be compared to those completing to see if any differences can be identified and whether potential bias can be inferred. Subjects may withdraw because of improvement in symptoms thus giving an *underestimate* of effectiveness. However, withdrawals due to side-effects or other difficulties can result in an *overestimate* of effectiveness. Thus, the validity of a study may be dubious if there is a large number of patients lost to follow-up but a low incidence of undesirable effects is reported.

Systematic reviews

The advantages of a systematic review include increased power and precision in addressing the research question. A comprehensive objective search strategy is used to find all relevant studies addressing a specified research question. The source studies are methodically located, appraised

There is an accepted structured process for formally undertaking a systematic review:⁶

1. A well formulated question is proposed with a list of search words.
2. A comprehensive data search is undertaken, covering a wide range of sources to avoid reporting bias.
3. An unbiased selection/abstraction process (with at least two people) follows.
4. A validity assessment is carried out (ideally using a component approach to assess relevant methodological aspects) with results reflected in the analysis.
5. A study synthesis is displayed (in narrative form, detailing the individual studies included in the review).

The available data may be graded within a hierarchy of strength of evidence to assist interpretation of all evidence jointly.

Box 6. How to conduct a systematic review.

- A label displays what the comparison and outcome of interest are at the top of the plot.
- The scale of treatment effect is shown at the bottom.
- A vertical line of no effect is positioned in the middle to represent no difference between control and treatment.
- For each study, an identification number, trial data for experimental and control groups and the % weight of the study are shown. A box is placed to show the treatment effect and its area may be made to be proportional to the weight of the study. The horizontal line across each box shows the confidence interval (which is the range of population values with which the sample is compatible, usually 95%). If the confidence interval crosses the line of no effect, then no statistically significant difference in the effects of the two interventions was found.
- The pooled analysis is shown as a diamond shape. The widest point vertically represents the treatment effect and horizontally represents the confidence interval.

Box 7. Interpreting Forest plots.⁶

and synthesized to provide as reliable an overview as possible. Ideally, there should be at least two reviewers who are blind to study authorship (Box 6).

The Cochrane Collaboration⁶ is an independent, international, not-for-profit organization, dedicated to making up-to-date, accurate information about the effects of healthcare readily available worldwide. The major product of the collaboration is the Cochrane database of systematic reviews, published quarterly as part of the Cochrane library. They are prepared by volunteer healthcare professionals and overseen by editorial teams. Ideally, Cochrane reviews

are based on the inclusion of randomized controlled trials and, if there is a lack of such data, it is difficult to draw any meaningful conclusions.

Meta-analyses

A meta-analysis⁶ uses statistical methods to combine the results of different studies with the aim of integrating their findings, pooling data and identifying the overall trend of results. In this way, the results of a number of small but similar studies can be combined to achieve a large enough sample size to detect an effect and the power is increased. They are often associated with a

systematic review.

The studies are weighted according to how much information is provided. Those with more participants, more events and lower variance carry more weight. *Subgroup analyses* can be undertaken *post hoc* if it is suspected that certain features may alter the effect of an intervention (such as gender or disease subtype). *Sensitivity analyses* can also be carried out *post hoc* whereby the meta-analysis is repeated without including lower quality trials. Ultimately, a *single summary statistic* is calculated to represent the treatment effect found in each study and these are then combined to give an overall measure. *Forest plots* are usually used to display the results (Table 1, Box 7).

References

1. Burls A. What is critical appraisal? 2009 http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_critical_appraisal.pdf
2. Greenhalgh T. *How to Read a Paper: the Basics of Evidence Based Medicine* 3rd edn. Oxford: Blackwell, 2006.
3. O'Brien KD, Wright JL, Mandall NA. How to do a randomized controlled trial. *J Orthod* 2003; **30**(4): 337–341.
4. Sutherland SE. Papers about therapy. Evidence-based dentistry. Part V. *J Can Dent Assoc* 2001; **67**(8): 442–445.
5. Wade AM, Pan HQ, Eaton S, Pierro A, Ong EGP. An investigation of minimisation criteria. *BMC Med Res Method* 2006; **6**: 11. (<http://www.biomedcentral.com/1471-2288/6/11>).
6. www.cochrane.org (The Cochrane Collaboration).

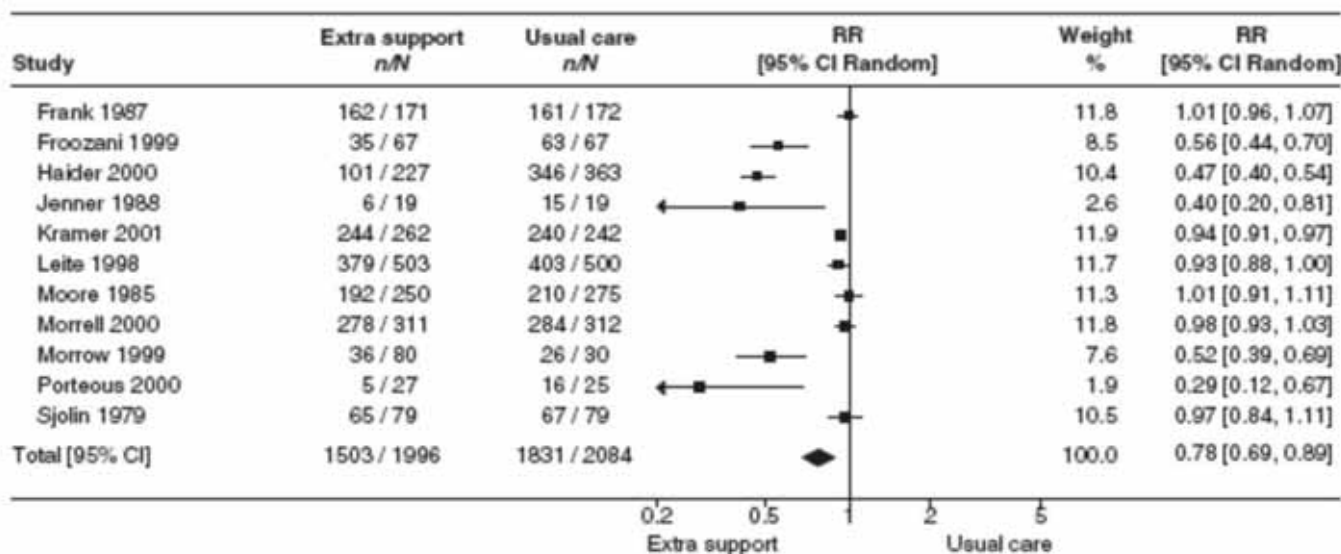


Table 1. Forest plot relating to a meta-analysis investigating the effect of extra occupational therapy support with oral hygiene for patients with rheumatoid arthritis affecting the hands. It can be seen that the overall treatment effect is in favour of extra support.