Journal of Orthodontics, Vol. 30, 2003, 159–164

FEATURES SECTION

How to … interpret the orthodontic literature

A-M. Glenny
Cochrane Oral Health Group, University Dental Hospital of Manchester, UK

J. E. Harrison
Liverpool University Dental Hospital, UK

Introduction

In a previous article in this series, Bickley and Harrison discuss the issue of searching for evidence.[1] Within dentistry alone, there are around 500 journals publishing over 43,000 research articles a year. Given that a large proportion of these papers are of limited relevance to everyday practice and may be of poor quality, how do you know which of these articles you should read to inform your practice and which you can disregard? In this article we aim to provide information on how you can appraise the research to identify articles that are both of a high quality and relevant to our clinical practice.

Most published papers appearing in the medical and dental journals follow the IMRAD format (**I**ntroduction, **M**ethods, **R**esults **a**nd **D**iscussion).[2] Published papers will often begin with an abstract that summarizes the key elements from each section. It is very tempting, when reading a paper of interest, to focus on the abstract and the results or the conclusions of the study. However, to decide whether a paper is truly worth reading, attention should be given primarily to the methods section to establish whether the study design was appropriate and valid. Consideration should then be given to what the paper says (the results of the study) and whether it helps your clinical practice (the relevance, or applicability, of the paper).[3]

Is the study design appropriate?

The initial step in assessing a research paper is to determine what study design has been used and whether or not it was appropriate for the question being asked.

Primary studies are often graded into a hierarchy of evidence according to their design. Studies least susceptible to bias are placed at the top of the hierarchy. For example, experimental studies or clinical trials (those studies in which certain conditions, in particular assignment of study participants (or teeth) to intervention groups, are under the control of the investigator) are placed above observational studies. Observational studies are those in which natural variations in exposure or interventions among study participants are investigated to explore the effect of the intervention or exposure on health outcomes.[4] The strength of evidence decreases from the controlled observational studies to those without controls, as the susceptibility to bias increases (Table 1).

Such hierarchies provide a useful 'rule of thumb' against which to grade studies. However, it must be noted that different clinical research questions require evaluation through different study designs. Randomized controlled trials (RCTs) may well be the 'gold standard' upon which to base decisions on the effectiveness of interventions, but they are not necessarily appropriate or ethical to answer other questions.

The key features of different study designs have been described previously.[5] Table 2 illustrates the types of questions that can be addressed by the various study designs. When assessing the research literature it is important to identify whether or not the highest, appropriate level of evidence has been used to answer the research question.

Table 1 An example of a hierarchy of evidence (adapted from the NHS Centre for Reviews and Dissemination[4])

STRONG	Experimental studies/clinical trials
	Randomized controlled trials
	Non-randomized controlled clinical trials
	Observational studies
	Cohorts
	Case-controls
	Cross-sectional surveys
	Case series
	Case reports
WEAK	Expert opinion, consensus

Address for correspondence: A-M. Glenny, Cochrane Oral Health Group, University Dental Hospital of Manchester, Higher Cambridge Street, Manchester M15 6FH, UK. Email: a.glenny@man.ac.uk

Table 2 Study designs and the types of questions they address

Definition of study design	Used for
Experimental studies	
Randomized controlled trial – a group of participants (or other unit of analysis, e.g. teeth) is randomized into different treatment groups. These groups are followed up for the outcomes of interest	Evaluating the effectiveness of an intervention *What is the most effective adhesive for fixing brace attachments to teeth?* *Is a fixed appliance more effective than a functional appliance for correcting a Class II malocclusion*
Split-mouth design—each patient is his/her own control. A pair of similar teeth, or groups of teeth (quadrants) may be selected and randomly allocated to different treatment groups	
Non-randomized controlled clinical trial—allocation of participants under the control of the investigator, but the method falls short of genuine randomization.	
Observational studies	
Cohort—a longitudinal study, identifying groups of participants according to their exposure/intervention status. Groups are followed forward in time to measure the development of different outcomes	Measuring the incidence of a disease; looking at the causes of disease; determining prognosis *Do teenagers with crossbites go on to develop temporomandibular joint disorders?*
Case-control—involves identifying two groups; those who have the (outcome of interestcases) and those who have not (controls). The investigator then looks back in time to see who had the exposure/intervention of interest	Looking at the causes of disease; identification of risk factors; suitable for examining rare diseases *What are the risk factors associated with root resorption in patients who have undergone orthodontic treatment?*
Cross-sectional survey—the observation of a defined population at a single point in time or time interval. The status of an individual with respect to the presence or absence of both exposure/intervention and outcome are determined at the same time	Measuring the prevalence of a disease; examining the association *What is the prevalence of children born with a cleft-lip/palate in North-West England?* *Is drinking carbonated drinks during orthodontic treatment associated with an increased risk of decalcification?*

The orthodontic literature is predominately made up of observational studies. A recent, systematic hand search and classification of 5 years of orthodontic literature (1994–1998) has shown that experimental studies account for around 15 per cent of the research published in three leading journals (*British Journal of Orthodontics*, *American Journal of Orthodontics and Dentofacial Orthopedics*, and the *European Journal of Orthodontics*).[6] The remainder of the articles were observational studies, literature reviews, author's opinion, or articles detailing current availability of products, materials, methods, or services. A similar study, examining articles on the methods of distalizing maxillary molars, classified 22 per cent of identified articles as experimental and 70 per cent as observational studies, although difficulties in classifying research articles from the methods presented were noted.[7] Although it is disappointing that so few studies are experimental in design, this may well be a reflection of the questions being asked within orthodontic research.

How well was the study conducted?

If it can be established that the research question has been addressed using an appropriate research design, we still need to consider the fact that within any particular design there is huge variability between studies with regard to how well they are executed. This may mean that in certain cases, for example, a poor RCT may be no more reliable than a well-conducted cohort study.

An important issue to consider at this stage is the study's validity, in particular internal validity.

Internal validity refers to the degree to which the results of a study are likely to approximate to the 'truth' for the circumstances being studied.[8] Has the study been conducted in such a way that systematic error (bias) has been minimized? External validity refers to the degree to which the effects observed in the study are applicable to the outside world; how generalisable are the results to other circumstances?[4] Obviously, if internal validity does not exist, there is little point in considering a study's external validity.

There are four main biases that can affect the internal validity.[9]

Selection bias

The term selection bias is used in different ways within the medical literature. It is often used in relation to bias occurring during the selection of representative subjects or to bias occurring during the selection of subjects to exposures.[10] The former of these is more to do with the study participants' characteristics and is linked to external validity. If bias occurs during the selection of subjects to exposures, however, systematic differences between comparison groups in prognosis or responsiveness to treatment may arise. If the groups under comparison are not similar at baseline, the differential effects of any intervention may be distorted due to confounding factors. Observational studies are particularly vulnerable to such selection bias. For example, in case-control studies it is imperative that the cases are as similar as possible to the controls, except for the presence or absence of the disease/outcome under study. Although it is feasible to ensure the groups are comparable with regard to known confounding factors, this is not the case for unknown confounding factors. Confounding factors also cause problems with cross-sectional surveys and cohort studies.

The process of random allocation to treatment groups, within experimental study designs, aims to produce groups that are comparable in terms of both known and unknown confounding factors, thus minimizing selection bias. Ideally, the generation of the random allocation sequence should be unpredictable (computer-generated random numbers, coin tossing, drawing lots, throwing dice, etc.). Allocation based on case record number, date of birth, date of admission or alternation are all open to manipulation, and therefore introduce a greater risk of selection bias into the study. True randomization not only requires the allocation sequence to be unpredictable, but also that the sequence is concealed from the investigators involved in the enrolling of patients in order to avoid the selective enrolment of patients based on prognostic factors.

A related form of bias is recall bias, relating to differences in the way exposure information is remembered or reported by participants who have experienced an adverse health outcome and by those who have not. For example, orthodontic appliance wearers who experience enamel demineralization may either over or under report the use of topical fluorides, frequency of tooth brushing or consumption of carbonated drinks, in comparison with those who do not experience demineralization. Study designs that select subjects at outcome, such as cross-sectional survey and case-control studies, are particularly susceptible to recall bias.

Performance bias

Performance bias refers to systematic differences in care provided to participants in a study, apart from the intervention being evaluated.[4] The knowledge of assignment to different treatment groups may affect a study participant's reporting of symptoms. In addition, an investigator may treat participants receiving one intervention differently from those receiving the comparison intervention. For example, in an RCT comparing the effectiveness of manual versus powered toothbrushes on oral hygiene in orthodontic patients, the study investigator may be tempted to provide brushing instructions to one group and not the other, depending on which intervention they favour (either consciously or subconsciously). Blinding of study participants and investigators to treatment allocation helps minimize performance bias. When a study is described as single blind only the participants are blind to their group allocation. When both participants and investigators are blind to group allocation the study is described as double blind. Blinding can be achieved through the use of placebo interventions, ensuring both groups receive interventions that appear identical in terms of taste, smell, mode of delivery, etc. However, as in the toothbrush example above, blinding to treatment group is not always feasible.

Measurement/detection bias

Even when blinding to treatment groups cannot be achieved, blinding to outcome assessment is usually possible. In orthodontics this can be achieved by blind assessment of study models, radiographs and/or photographs. This can help minimize systematic differences that may occur in how outcomes are ascertained from the groups under comparison (measurement or detection bias). Blind outcome assessment is of particular importance when the outcome being assessed is subjective in nature (e.g. dental or facial aesthetics; decalcification; degree of paraesthesia). Empirical evidence has shown that trials with open assessment of the outcome can over estimate the treatment effects by 35 per cent.[11]

Attrition bias

Attrition bias occurs when there are systematic differences between comparison groups in withdrawals or exclusions of participants from the results of a study.[9] For example, patients may drop out of a study because of side effects of the intervention or difficulty in wearing a particular appliance. Excluding these patients from the analysis could result in an over-estimate of the effectiveness of the intervention. Conversely, participants might drop out of a study due to an improvement in the symptoms or malocclusion, e.g. overjet or crowding, resulting in an underestimate of treatment effect if they are not included in the analysis. In order to minimize attrition bias, all study participants should be accounted for in the analysis and the analysis undertaken on an intention-to-treat basis (participants analysed according to the group to which they were initially allocated, regardless of whether they dropped-out, fully complied with the treatment or ended up crossing over to the other treatment group).

What are the results?

Assuming the methods of the study are such that systematic biases have been avoided, the results of the study then need to be considered. Care should be taken to ensure that the results fulfil the aims of the study.[12] Gaps in the results may be due to several reasons:

- lack of space to present all the results in the published article;
- an oversight on the author's part;
- data not collected;
- the finding was not acceptable to the author or those involved in the publication process.

This can lead to a form of publication bias where only the significant changes or results are published. Studies with multiple outcomes, e.g. cephalometric studies, are particularly prone to this kind of bias. Ideally, the results of the study should be presented in a clear, logical way, so that the reader can draw their own conclusions from them, rather than having to rely on the author's interpretation. The tables and figures presented in the paper should stand alone, and the numbers tally with those discussed in the text.

Consideration needs to be given to the sample size of the study. When assessing clinical trials, larger studies are deemed 'better' as they have greater statistical power and can produce a more precise estimate of effect. The presentation of an *a priori* calculation of sample size gives some indication that the authors have considered the statistical power of the study. Smaller studies are often under-powered and may therefore be unable to detect a statistically significant difference between comparison groups, even if one exists. However, with regard to observational studies, bigger is not always better. Larger studies may be less able to pay as much attention to characterizing the exposure or outcome of interest, and the confounding factors, than smaller studies.[13]

For prospective studies (cohorts, clinical trials) the duration of follow-up needs to be long enough for the effect of the intervention to be demonstrated by the outcome of interest. For example, there is little point in conducting a clinical trial examining the effectiveness of orthodontic appliances for correcting posterior crossbites if the duration of the study is too short to allow expansion of the upper jaw/teeth to occur and/or the permanent dentition to be established. In orthodontics, where relapse (post-treatment changes) can be a problem, sufficient follow-up of patients after active treatment and/or out of retention should also be included in the trial design. It is also necessary to ensure that as many people as possible are followed up for the full period of the study. This is of particular concern for cohort studies.

Attention should be given to the outcomes measured: are all the important outcomes considered? How were the outcomes assessed? Outcome assessment measures should be described and the issue of validity (the ability of the method of assessment to truly measure what it is supposed to) and reliability (the ability of the measure to achieve similar results when applied on more than one occasion) addressed.[12]

Appropriate statistical techniques should be used within the study and, ideally, be presented in the methods section of the paper. Although difficult to identify from a published article, the use of numerous statistical tests may be misleading. The more tests that are undertaken, the more likely a result of spurious significance will be identified.[12] The statistical significance of the main findings should be assessed. These are commonly presented as either *P*-values or confidence intervals. Typically a *P*-value of less than 0.05 is used to show that the result is unlikely to have occurred by chance. The smaller the *P*-value, the greater the confidence that the result was real. Confidence intervals, along with providing a test of statistical significance, provide a range within which the true value could lie. An important point to remember is that not all statistically significant differences in outcomes are necessarily clinically significant.

The statistical analysis of orthodontic studies is often poorly carried out. A recent systematic review looking at orthodontic adhesives excluded a large number of studies for various reasons, the most common reason being inappropriate or unclear statistical analysis.[14] There are two common problems with the statistical analysis of orthodontic studies, one is the inappropriate analysis of studies with a split-mouth design and the other is the analysis of teeth/sites as though they were independent of the patient. It is important that the data are analysed taking the clustering or pairing within the patient into account.

Are the results relevant to your clinical practice?

Not all valid research articles will be relevant to your clinical practice. Even if a study focuses on a particular intervention or outcome of interest, consideration needs to be given to the external validity of the piece of research. That is, how generalizable are the findings to other populations outside of the study? Are the participants included in the research article similar enough to the patients you deal with or are they so different that the results don't apply? Although a hugely subjective process, the inclusion/exclusion criteria for the study and details of the characteristics of the study participants can help to establish whether the results are likely to be relevant to your patient population or not.

The feasibility of implementing any research findings into practice also needs to be considered. The results of a study may be favourable and relevant to your patient population, but issues such as the cost of the intervention, training in new techniques, and additional monitoring of patient outcomes may render a change in practice unrealistic.

Discussion

The ever-increasing volume of research literature leaves clinicians exposed to information overload. Although the increasing accessibility of such literature is commendable, unless the available information is relevant to the clinician, and of good quality, it can, at the very least, be misleading. The use of systematic reviews and the introduction of 'secondary' journals over the last decade[15] has helped, to some extent, to alleviate this problem by identifying and appraising the validity of clinically useful articles for the clinician. Sources of such secondary information available to dentists include:

- The Cochrane Database of Systematic Reviews (CDSR) (for details of free access see www.cochrane.org).
- The Cochrane Oral Health Group's homepage (www.cochrane-oral.man.ac.uk) also provides free access to the abstracts of reviews completed by members of the group, and lists details of all protocols and titles currently registered with the group.
- The Database of Abstracts of Reviews of Effectiveness (DARE)—a database of quality assessed systematic reviews.*
- The NHS Economic Evaluation Database (NHS EED)—a database of quality assessed economic evaluations (including cost-benefit, cost-utility and cost-effectiveness analysis).*
- The Health Technology Assessment (HTA) Database—a database containing records of ongoing projects being conducted by members of the International Network of Agencies for Health Technology Assessment (INAHTA), as well as publications reporting completed technology assessments carried out by INAHTA members and other technology assessment organizations.*
- Clinical evidence (www.evidence.org) and in printed version—a compendium of evidence on the effects of clinical interventions, updated every six months.
- Bandolier (www.jr2.ox.ac.uk/Bandolier)—a healthcare journal, using evidence-based healthcare techniques to provide advice about particular treatments or diseases. Available in print and via the Internet.
- Evidence-Based Dentistry (www.nature.com/bdj)—a resource covering issues related to the evidence-based approach in dentistry today.

(*Available through http://agatha.york.ac.uk/welcome.htm and also on the Cochrane Library.)

Whilst all of these sources assimilate and appraise research relevant to dentistry, at present they provide relatively little information on questions relating to orthodontics. As the number of topics covered by such sources of information increases, so their usefulness to the busy clinician will also increase. However, in order for orthodontists to be sure that their clinical practice is based upon the best available research evidence, they cannot solely rely on secondary sources of information. An understanding of the different types of study designs and associated limitations can help in the identification of articles that are of a high quality and relevant to clinical practice.

Acknowledgments

The authors would like to thank Dr Helen Worthington for her comments on the statistical analysis of orthodontic literature.

References

1. Bickley S, Harrison JE. How to … find the evidence. *J Orthod* **30**(1): 72–78.
2. Greenhalgh T. How to read a paper: getting your bearings (deciding what a paper is about). *BMJ* 1997; **315:** 243–6.
3. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II How to use an article about therapy or prevention. *JAMA* 1993; **270:** 2598–601.
4. NHS Centre for Reviews and Dissemination. *Undertaking Systematic Reviews of Research on Effectiveness: CRD's guidance for carrying out or commissioning reviews*, 2nd edn. York: NHS Centre for Reviews and Dissemination, University of York, 2001.
5. Harrison JE. Evidence-based orthodontics—how do I assess the evidence? *J Orthod* 2000; **27:** 189–97.
6. Harrison JE. 2001. *Evidence based orthodontics: the way forward or an unrealistic dream?* PhD Thesis, University of Liverpool, 2002.
7. Atherton G, Glenny AM, O'Brien K. Methods of distalizing maxillary molars: a systematic review of the literature. *J Orthod* (in press) 211–216.
8. Juni P, Altman DG, Egger M. Assessing the quality of randomised controlled trials. In: M. Egger M, Davey Smith G, Altman D (Eds) *Systematic Reviews in Health Care. Meta-analysis in Context*. London: BMJ Publishing Group, 2001.**[Q3]**
9. Clarke M, Oxman A (Eds) *Cochrane Reviewers' Handbook 4.1* [updated June 2000]. In: *Review Manager* (*RevMan*) [Computer program], Version 4.1. Oxford: Cochrane Collaboration, 2000.
10. Mark DH. Interpreting the term selection bias in medical research. *Fam Med* 1997; **29:** 132–6.
11. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; **282:** 1054–60.
12. Crombie IK. *The Pocket Guide to Critical Appraisal: a handbook for health care professionals*. London: BMJ Publishing Group, 1996.
13. Egger M, Davey Smith G, Schneider M. Systematic reviews of observational studies. In: Egger M, Davey Smith G, Altman D (Eds) *Systematic Review in Health Care. Meta-analysis in Context*. London: BMJ Publishing Group, 2001. 211–227.
14. Mandall NA, Millett DT, Mattick CR, Hickman J, Worthington HV, Macfarlane TV. Orthodontic adhesives: a systematic review. *J Orthod* 2002; **29:** 205–10.
15. Harrison JE. Evidence-based orthodontics: where do I find the evidence? *J Orthod* 2000; **27:** 71–8.