Invited Commentary

Methodological challenges when performing a systematic review

Nina Kay Anderson* and Yasas S. N. Jayaratne**

*Developmental Biology, Harvard School of Dental Medicine, Boston, MA, USA, **Division of Orthodontics, Department of Craniofacial Sciences, University of Connecticut School of Dental Medicine, Farmington, CT, USA

Correspondence to: Nina Kay Anderson, Developmental Biology, Harvard School of Dental Medicine, 188 Longwood Avenue, Boston, MA 02115, USA. E-mail: nina_anderson@hsdm.harvard.edu

Summary

Well-conducted systematic reviews can provide a foundation upon which to base educational curricula, practice guidelines and healthcare policy. We aim to provide an overview of important methodological issues that need to be addressed during the five phases of performing a systematic review. 1. The specific problems to be addressed by the systematic review need to be formulated as clear, unambiguous and structured question/s. 2. Literature from multiple resources should be searched to avoid publication biases, as positive outcomes are more likely to be published than null/negative results. 3. The quality of selected articles should be assessed using a checklist, while the reliability of graders, as well as the quality of the checklist, needs to be established a priori. 4. Statistical heterogeneity needs to be assessed to determine if a meta-analysis is appropriate to pool the data. If not, a narrative synthesis of the evidence/overall findings needs to be performed within a clinical context.

The motivation for seeking orthodontic treatment is a multifac-torial experience involving objective need, subjective symptoms, and social sufficiency (1–4). Compared to adults, the impact of untreated malocclusions on oral health-related quality of life (OHRQOL) in young patients may fluctuate, possibly due to the physical and behavioral changes that occur from childhood through adolescence. OHRQOL has been defined as ‘a standard of health of oral and related tissues which enables an individual to eat, speak, and socialize without active disease, discomfort, or embarrassment’ (1) or ‘the absence of negative impacts of oral conditions on social life and a positive sense of dentofacial self-confidence’ (1).

Why do a systematic review?

The purpose of the systematic review by Dimberg et al. (5) was to evaluate the evidence regarding the influence of malocclusions on OHRQOL in children and adolescents. Research that is relevant to OHRQOL is scattered throughout a wide range of sources and literature. A well conducted systematic review can provide a critical appraisal and synthesis of the evidence from large numbers of individual studies on OHRQOL, in order to evaluate their conclusions and recommendations for application in a clinical setting. In general, systematic reviews can provide a foundation upon which to base educational curricula, practice guidelines and healthcare policy. Systematic reviews can also be used as teaching tools (e.g. identifying strengths and weaknesses in others’ work prior to initiating your own research), for recognizing research gaps, aid in developing research agendas and allocating resources (6). For all of their potential merit systematic reviews remain shrouded in mystery. How do you go from hundreds to less than a dozen studies? Why are so few studies selected and so many others rejected? How do you combine results and how did seemingly insignificant findings become significant? Finally, systematic reviews themselves vary greatly in quality and reporting characteristics. In this invited commentary we hope to provide an overview of some methodological issues that need to be addressed when performing a systematic review using the article by Dimberg et al. (5) as an example. It is beyond the scope and skill of the authors to provide readers with enlightenment, however, we hope to stimulate others to appreciate the difficulty of performing a systematic review, and their value
in an area such as orthodontics and quality of life containing so many studies.

There are five steps to conducting a systematic review (6):

Step 1: Framing questions for a review
Step 2: Identifying relevant literature
Step 3: Assessing the quality of the literature and evidence
Step 4: Summarizing the evidence
Step 5: Interpreting the findings

**Framing the question**

Framing the question to be answered is a rigorous multifactorial process and is important so as to avoid having to change the question later during the review. The specific problems to be addressed by the systematic review need to be formulated as a clear, unambiguous, and structured questions. The Population, Intervention, Comparison and Outcome (PICO) format can be used when framing questions.

Dimberg et al. (5) state that the question to be addressed was ‘whether malocclusions have an impact on OHQOL in children and adolescents’. To define their population, the search mesh included ‘child 6–12 years and adolescent 13–18 years’ with the final studies having ages ranging from 8 to 15 years. Further on in the results, one of the included studies refers to pre-or early adolescents or younger children (8–10 years). The terms children and adolescents can represent a range of different ages and stages of physical and cognitive development. For example, the Center for Disease Control (CDC) uses developmental categories of middle childhood (6–14years), young teens (12–14 years), and teenagers (15–17years) (6). The World Health Organization defines Adolescents as ‘young people between the ages of 10 and 19 years’ (8).

There may be large variation in study designs across the collected evidence, (randomized controlled, experimental study without randomization, observational (cohort, case–control), observational without control groups (cross-sectional, before/after, case series) and case reports. In this systematic review, all six of the studies included in the final analysis were cross sectional. However, assuming that there would be several types of study designs to be included, it would have been important to have an *a priori* plan for stratifying studies by design and quality, and to state procedures for modifying quality assessments and interpretation of findings. Furthermore, features related to study design such as the setting (school versus orthodontic clinic) may also have been associated with variable effects.

**Identifying relevant literature**

The authors appropriately used assistance of a specialist in informatics at the Medical Library to navigate through bibliographic databases and electronic information. On the other hand, only five databases were utilized. Examples of additional resources which could have been used include: Clinical Trial Registries (International Clinical Trials Registry Platform, ClinicalTrials.gov, European Union Clinical Trials Registry, Pan African Clinical Trials Registry) and Databases (Cochrane Library, Web of Science Scopus, Latin American and Caribbean Health Sciences Literature—LILACS).

In addition, other acceptable sources of obtaining literature should have included hand-searching journals and gray literature. ‘Gray Literature’ is the term for information obtained outside traditional publications such as books and journals. These include reports (e.g. from Centers for Disease Control, World Health Organization, etc.), dissertations, conference papers, governmental or private sector research, clinical trials—ongoing or unpublished, and personal communications with experts and researchers in the field (9).

The inclusion of literature from multiple resources helps to avoid publication bias. Positive findings are more likely to be published than null or negative results. The tendency to report outcomes which are favourable to the hypotheses or found to be effective can present an obstacle to obtaining relevant data/information (10, 11). Studies that support the current zeitgeist may also be more likely to be favourably reviewed and published. Rigorous systematic reviews need to attempt to identify relevant studies via gray literature and/or personal communications in order to minimize publication biases. The publication bias can be further explored with Funnel Plots where the effect estimates from individual studies are plotted against a measure of study size (6).

In this systematic review, only English and Scandinavian language articles were included in the search, which was a limitation. Overall, there appears to be no evidence that non-English language articles are of lower quality than English only (12), however, the costs associated with translations may be prohibitive for most teams embarking on a systematic review.

Spelling differences in the search terms may have also influenced the search results, for example a Pubmed search (18 August 2014) using for term ‘dental esthetics’ resulted in 10640 results, while ‘dental aesthetics’ returned 11338; a difference of 698 items. Similarly, using ‘well being’ returned 464085 items while ‘wellbeing’ returned only 6898, a difference of 457187 possibilities.

**Assessing the quality of the literature and evidence**

There is no universal agreement as to a specific system or approach for assessing and grading the quality of articles. Although all of the conventional grading systems share common features, different factors are used in assessing and establishing hierarchies. The authors used the Grades of Recommendation Assessment, Development and Evaluation (GRADE) system to assess the quality of the evidence (13). The GRADE working group reports that 50 organization endorse or use an adapted version of the system. The authors stated that ‘In case of inter-examiner conflicts each article was reread and discussed until consensus was reached’. Ideally there should be an initial training using a gold standard article with retest at time points for reliability of graders throughout to make check sure that they were still calibrated and that subjectivity did not evolve during the screening/grading process.

When published studies do not contain information needed for a systematic review such as dropout rate, it is appropriate to contact the authors to clarify or obtain missing data (14). It appears that additional studies could have potentially been included had the study authors been contacted in order to obtain the dropout rates.

**Summarizing the evidence**

Although the selected studies were subjected to a quality assessment using critical appraisal guides and design-based quality checklists, assessments should also have been used for exploring heterogeneity and determine if a meta-analysis was appropriate (6). Statistical heterogeneity relates to the variability in observed treatment effects across studies. Large amounts of variability may be an indication of a poorly formulated question, confounding factors, and reduce the confidence in reporting that effects are valid (6). A graphical representation such as a forest plot (5) to display the between study heterogeneity can provide a concise representation of the size and precision of individual study effects and the aggregated effects.
Interpreting the findings

From a search of 1142 titles, 22 publications met the authors’ inclusion criteria, a final 6 were utilized in this systematic review leading to the conclusion that anterior malocclusions have a negative impact on OHRQOL predominantly in the emotional and social dimensions. The authors provided a descriptive summary of the study results and some of their limitations. It would have been important to provide more narrative synthesis (15) of the body of evidence of OHRQOL within a clinical context. In other words synthesizing their overall findings to ‘tell the story’ of how malocclusion impacts OHRQOL, for whom and under what circumstances.

Role of systematic reviews in orthodontics

Systematic reviews are useful for establishing practice guidelines by exploring the clinical relevance of their results (e.g. describe how the review findings contrast with conventional wisdom). Besides summarizing the evidence, systematic reviews can be used for identifying research gaps and discussing recommendations for future research.

There is a vast amount of research in orthodontics. Thus, improving our skills on how to critically appraise a systematic review related to orthodontics will advance not only our knowledge on conducting a rigorous review, but also our ability of making meaningful clinical decisions with the synthesized evidence.

Given the complexity, time, expense, and resources required to perform a high quality systematic review, the authors deserve kudos for this effort.

References