

EDITORIAL

Open Access



Classifying scientific evidence as the basis for evidence-based decision making: is strength of evidence absolute?

Eyal Rosen* and Igor Tsesis

Background

One of the most challenging aspects of evidence-based decision making is the appraisal of the available evidence in order to separate the wheat from the chaff. It has been stated that “*not all evidence is created equal*” (Cochrane Consumer Network - Levels of evidence), and it has been recommended to grade the available literature by the strength of evidence as determined by the methods used to minimize bias within a study design (Cochrane Consumer Network - Levels of evidence; Bossuyt & Leeftang 2008; Gutmann 2009; Mileman & van den Hout 2009; Reitsma et al. 2009; Rosenberg & Donald 1995; Suebnukarn et al. 2010; Sutherland & Matthews 2004; Zwahlen et al. 2008). A cornerstone of this appraisal process is the use of hierarchical systems for classifying the evidence. This hierarchy is known as the *Levels of evidence* (Burns et al. 2011).

Traditional hierarchical systems of levels of evidence

One of the earliest reports of a *Levels of evidence* (LOE) hierarchical system was published in 1979 by the Canadian Task Force on the Periodic Health Examination (Burns et al. 2011; Canadian Task Force on the Periodic Health 1979). LOE was traditionally defined as “*hierarchical grading systems for classifying study strength/quality*” (Burns et al. 2011; Wright et al. 2003; Sutherland 2001; Oxford Centre for Evidence-based Medicine – The Oxford 2011; Petrisor et al. 2006). Since the introduction of LOE, several other organizations have adopted various classification systems, most of which share a lot in common (Bossuyt & Leeftang 2008; Gutmann 2009; Mileman & van den

Hout 2009; Reitsma et al. 2009; Rosenberg & Donald 1995; Suebnukarn et al. 2010; Sutherland & Matthews 2004; Zwahlen et al. 2008; Burns et al. 2011; Canadian Task Force on the Periodic Health 1979). To-date, classification systems such as the system presented by the “Oxford Centre for Evidence-Based Medicine” (Oxford Centre for Evidence-based Medicine – The Oxford 2011) attempt to provide comprehensive hierarchical grading for classifying scientific evidence (Bossuyt & Leeftang 2008; Gutmann 2009; Mileman & van den Hout 2009; Reitsma et al. 2009; Rosenberg & Donald 1995; Suebnukarn et al. 2010; Sutherland & Matthews 2004; Zwahlen et al. 2008; Burns et al. 2011; Canadian Task Force on the Periodic Health 1979; Tsesis et al. 2009).

The traditional hierarchical systems of classifying the evidence primarily use the study design as the basis for the grading process (Bossuyt & Leeftang 2008; Gutmann 2009; Mileman & van den Hout 2009; Reitsma et al. 2009; Rosenberg & Donald 1995; Suebnukarn et al. 2010; Sutherland & Matthews 2004; Zwahlen et al. 2008; Burns et al. 2011; Canadian Task Force on the Periodic Health 1979; Tsesis et al. 2009). A clinical study may be experimental (interventional) or observational. In experimental studies, the intervention is under the control of the investigator, whereas in observational studies, the investigator observes patients at a point in time (cross-sectional studies) or over time (longitudinal studies). These observations are done either by looking forward and gathering new data (prospective), or by collecting already existing data (retrospective studies) (Sutherland 2001). As an example of the levels of evidence, in practically all LOE classification systems *Randomized controlled trials (RCTs)* are considered as a high LOE, as opposed to *Case reports* and *Case series* that are considered as a low LOE (Bossuyt & Leeftang 2008; Gutmann 2009; Mileman & van den Hout 2009; Reitsma et al. 2009; Rosenberg & Donald 1995; Suebnukarn et al. 2010; Sutherland & Matthews 2004; Zwahlen et al. 2008; Burns et al. 2011;

* Correspondence: dr.eyalrosen@gmail.com

Department of Endodontology, Maurice and Gabriela Goldschleger School of Dental Medicine, Tel Aviv University, P.O. Box 39040, Tel Aviv 6997801, Israel

Canadian Task Force on the Periodic Health 1979; Sutherland 2001; Tsesis et al. 2009).

Levels of evidence from the clinical decision making perspective

From the clinical perspective, the study design and the ensuing LOE classification may not always be the most important factor to consider while assessing the available evidence. Occasionally, even a well conducted RCT may not necessarily generate new knowledge superior to the knowledge gained from a case series study. Thus an RCT may be less significant to the clinical decision making than a case series study. For example, when a clinician assesses the benefits and risks of a possible treatment modality, the use of LOE alone can be misleading. An historical example is the use of Bisphosphonates for dental purposes:

Bisphosphonates (BPs) are a class of drugs that inhibit bone resorption and they are successfully used across a wide range of medical disciplines for bone diseases (Molvik & Khan 2015; Costa 2014; Anagha & Sen 2014; Giusti 2014; Bhatt et al. 2014; Eriksen et al. 2014). However, like any drug, BPs possess the risk of side effects. BPs-related osteonecrosis of the jaw (BRONJ) has been characterized as a major side effect of BPs therapy. This serious side effect may appear following a triggering event such as tooth extraction (Fliefel et al. 2015). The earliest descriptions of BRONJ were published in 2003 (Fliefel et al. 2015), including a letter to the editor by Marx (Marx 2003) who identified 36 cases of painful jaw bone exposures that were unresponsive to surgical or medical treatments. All patients were receiving BPs therapy. Marx stated that: *“it represents a heretofore unrecognized and unreported serious adverse affect; caution should be used when prescribing these drugs”* (Marx 2003).

This early warning of a potential serious side effect of BPs and other reports that emerged in the following months (Ruggiero et al. 2004) had to ring warning bells (European-Environment-Agency 2001; Kheifets et al. 2001). However, in the following years, placebo-controlled RCTs that randomized patients to receive BPs therapies or placebo for periodontal treatment proposes (Lane et al. 2005), (Rocha et al. 2004), resulted in a conclusion that *“bisphosphonate treatment may be an appropriate adjunctive treatment to preserve periodontal bone mass”* (Lane et al. 2005).

Since then, numerous reports on the development of BRONJ in patients treated with BPs have been published (Fliefel et al. 2015; Ruggiero et al. 2004; Ruggiero et al. 2014), and it was realized that the risk for BRONJ after some periodontal procedures may be comparable to the risk associated with tooth extraction (Ruggiero et al. 2014). Therefore, in retrospect, the administration of

BPs for periodontal purposes while exposing the patients to the risk of BRONJ is unthinkable.

At the same time period in which the first reports on BRONJ emerged in case series of relatively low LOE (Marx 2003; Ruggiero et al. 2004), other high LOE studies (RCTs) recommend the use of BPs for dental purposes (Lane et al. 2005; Rocha et al. 2004; Rocha et al. 2001; Tenenbaum et al. 2002). This fact highlights the conclusion that high LOE studies don't always contribute new knowledge to support the practitioner's clinical decision making that is superior to new knowledge from low LOE studies, and that the study design alone as a decisive factor for the evidence appraisal, can be misleading.

A comprehensive approach for assessing evidence as the basis for clinical decision making

Traditional study-design-based LOE grading systems may therefore only provide information on the credibility of the study's results, but do not provide any information regarding the relevance of the investigated clinical question to the practitioner's decision making. Thus, not only the strength of the evidence should be considered but also the *clinical significance* and *relevance* of the evidence, i.e. how appropriate the outcome measure is for assessing the benefits (or harms) of the treatment (Cochrane Consumer Network - Levels of evidence) in the relevant patients.

In addition, these traditional hierarchal systems of classifying evidence, with RCTs at their top end, were developed to a large extent for questions related to interventional studies. For questions related to diagnosis, prognosis or causation, other study designs such as cohort studies may often be more appropriate, and for these types of studies, it is useful to think of the various study designs not as a hierarchy, but as categories of evidence, where the strongest design which is possible, practical and ethical should be used (Sutherland 2001). Therefore, different types of questions require different types of evidence (Richards 2009).

Thus, classifying the evidence should not be done by using the traditional hierarchical systems of LOE alone (Richards 2009). A comprehensive appraisal of the evidence regarding a specific clinical question should combine an assessment of the *strength* of the evidence with other dimensions such as the *significance* and *relevance* of the evidence (Sutherland 2001).

The significance and relevance of the evidence

Unlike the well-established hierarchical systems of classifying the *strength* of evidence (Cochrane Consumer Network - Levels of evidence); Burns et al. 2011; Sutherland 2001; Oxford Centre for Evidence-based Medicine – The Oxford 2011; Petrisor et al. 2006; Richards 2009), to-date, determining the *significance*

and *relevance* of the evidence seems to be more intuitive and subjective.

Clinical significance may be defined as “*the practical or applied value or importance of the effect of an intervention*” (Kazdin 1999). Treatments that produce reliable effects may be quite different in their impact on patients’ function, and clinical significance brings this issue to light (Kazdin 1999). The assessment of *clinical significance* represents an important advancement in the evaluation of effects of interventions, including treatments, but also extending to prevention, education, and rehabilitation (Kazdin 1999). This assessment of the importance of the change and the impact on patient function adds critical dimensions to the overall evaluation of the evidence.

The *relevance* of the evidence may be defined as “*the appropriateness of the outcomes measured including any outcomes that are likely to be important to patients*” (Medical Research Council 2000). It is an important dimension of the evidence appraisal that includes the assessment of several aspects regarding the appropriateness of the outcomes, such as: their potential short and the long-term effects; their causal relation to outcomes of importance to the patient; and the extent to which the intervention can be replicated in other settings and patient groups unlike those in which its efficacy has been tested (Medical Research Council 2000).

In conclusion

The goal of decision making in healthcare is to choose the interventions that are most likely to deliver the outcomes that are of most interest to patients, and to prevent possible harmful outcomes. In this context, it is important to review the information about the *strength* of the evidence (LOE), together with the clinical *significance* and *relevance* of the evidence.

It will be useful for the medical community to develop hierarchical systems for classifying the *significance* and *relevance* of the evidence to enable a more objective evaluation process.

Published online: 03 August 2016

References

- Anagha PP, Sen S. The efficacy of bisphosphonates in preventing aromatase inhibitor induced bone loss for postmenopausal women with early breast cancer: a systematic review and meta-analysis. *J Oncol*. 2014;2014:625060.
- Bhatt RN, Hibbert SA, Munns CF. The use of bisphosphonates in children: review of the literature and guidelines for dental management. *Aust Dent J*. 2014;59(1):9–19.
- Bossuyt PM, Leeflang MM. Chapter 6: Developing Criteria for Including Studies. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008.
- Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128(1):305–10.
- Canadian Task Force on the Periodic Health. The periodic health examination. *Can Med Assoc J*. 1979;121(9):1193–254.
- Cochrane Consumer Network - Levels of evidence; Available from: <http://consumers.cochrane.org/levels-evidence>. Accessed 21 June 2016.
- Costa L. Bisphosphonates in adjuvant setting for breast cancer: a review of the meta-analysis of bisphosphonates’ effects on breast cancer recurrence presented in December 2013 at San Antonio breast conference. *Curr Opin Support Palliat Care*. 2014;8(4):414–9.
- Eriksen EF, Diez-Perez A, Boonen S. Update on long-term treatment with bisphosphonates for postmenopausal osteoporosis: a systematic review. *Bone*. 2014;58:126–35.
- European-Environment-Agency, editor. Late lessons from early warnings: the precautionary principle 1896–2000. Environmental issue report No 22; 2001. Copenhagen, Denmark: Luxembourg: Office for Official Publications of the European Communities; 2001.
- Fliefel R, Troltzsch M, Kuhnisch J, Ehrenfeld M, Otto S. Treatment strategies and outcomes of bisphosphonate-related osteonecrosis of the jaw (BRONJ) with characterization of patients: a systematic review. *Int J Oral Maxillofac Surg*. 2015;44(5):568–85.
- Giusti A. Bisphosphonates in the management of thalassemia-associated osteoporosis: a systematic review of randomised controlled trials. *J Bone Miner Metab*. 2014;32(6):606–15.
- Gutmann JL. Evidence-based/guest editorial. *J Endod*. 2009;35:1093.
- Kazdin AE. The meanings and measurement of clinical significance. *J Consult Clin Psychol*. 1999;67(3):332–9.
- Kheifets LI, Hester GL, Banerjee GL. The precautionary principle and EMF: implementation and evaluation. *J Risk Res*. 2001;4(2):113–25.
- Lane N, Armitage GC, Loomer P, Hsieh S, Majumdar S, Wang HY, et al. Bisphosphonate therapy improves the outcome of conventional periodontal treatment: results of a 12-month, randomized, placebo-controlled study. *J Periodontol*. 2005;76(7):1113–22.
- Marx RE. Pamidronate (aredia) and zoledronate (zometa) induced avascular necrosis of the jaws: a growing epidemic. *J Oral Maxillofac Surg*. 2003;61(9):1115–7.
- National Health and Medical Research Council, Commonwealth of Australia (2000); How to use the evidence: assessment and application of scientific evidence; Handbook series on preparing clinical practice guidelines; 2000. Available from: <https://www.nhmrc.gov.au/guidelines-publications/cp69/>. Accessed 21 June 2016.
- Mileman PA, van den Hout WB. Evidence-based diagnosis and clinical decision making. *Dentomaxillofac Radiol*. 2009;38(1):1–10.
- Molvik H, Khan W. Bisphosphonates and their influence on fracture healing: a systematic review. *Osteoporos Int*. 2015;26(4):1251–60.
- Oxford Centre for Evidence-based Medicine – The Oxford 2011 Levels of Evidence (2011); Available from: <http://www.cebm.net/ocbm-levels-of-evidence/>. Accessed 21 June 2016.
- Petrisor BA, Keating J, Schemitsch E. Grading the evidence: levels of evidence and grades of recommendation. *Injury*. 2006;37(4):321–7.
- Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0.0. The Cochrane Collaboration, 2009. Available from: <http://srdta.cochrane.org/>.
- Richards D. GRADING—levels of evidence. *Evid Based Dent*. 2009;10(1):24–5.
- Rocha M, Nava LE, Vazquez de la Torre C, Sanchez-Marin F, Garay-Sevilla ME, Malacara JM. Clinical and radiological improvement of periodontal disease in patients with type 2 diabetes mellitus treated with alendronate: a randomized, placebo-controlled trial. *J Periodontol*. 2001;72(2):204–9.
- Rocha ML, Malacara JM, Sanchez-Marin FJ, Vazquez de la Torre CJ, Fajardo ME. Effect of alendronate on periodontal disease in postmenopausal women: a randomized placebo-controlled trial. *J Periodontol*. 2004;75(12):1579–85.
- Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ*. 1995;310(6987):1122–6.
- Ruggiero SL, Mehrotra B, Rosenberg TJ, Engroff SL. Osteonecrosis of the jaws associated with the use of bisphosphonates: a review of 63 cases. *J Oral Maxillofac Surg*. 2004;62(5):527–34.
- Ruggiero SL, Dodson TB, Fantasia J, Goodday R, Aghaloo T, Mehrotra B, et al. American association of oral and maxillofacial surgeons position paper on medication-related osteonecrosis of the jaw—2014 update. *J Oral Maxillofac Surg*. 2014;72(10):1938–56.
- Suebunukarn S, Ngamboonsirisingh S, Rattanabanlang A. A systematic evaluation of the quality of meta-analyses in endodontics. *J Endod*. 2010;36(4):602–8.
- Sutherland SE. Evidence-based dentistry: part IV. Research design and levels of evidence. *J Can Dent Assoc*. 2001;67(7):375–8.
- Sutherland SE, Matthews DC. Conducting systematic reviews and creating clinical practice guidelines in dentistry: lessons learned. *J Am Dent Assoc*. 2004;135(6):747–53.

- Tenenbaum HC, Shelemay A, Girard B, Zohar R, Fritz PC. Bisphosphonates and periodontics: potential applications for regulation of bone mass in the periodontium and other therapeutic/diagnostic uses. *J Periodontol.* 2002;73(7):813–22.
- Tsesis I, Faivishevsky V, Kfir A, Rosen E. Outcome of surgical endodontic treatment performed by a modern technique: a meta-analysis of literature. *J Endod.* 2009;35(11):1505–11.
- Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am.* 2003;85-A(1):1–3.
- Zwahlen M, Renehan A, Egger M. Meta-analysis in medical research: potentials and limitations. *Urol Oncol.* 2008;26(3):320–9.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
