

Reliability of Quality Assessments in Research Synthesis: Securing the Highest Quality Bioinformation for HIT

Francesco Chiappelli^{1, 2*}, André Barkhordarian, C Phil^{1, 2}, Rashi Arora^{1, 2}, Linda Phi^{1, 2}, Amy Giroux^{1, 2}, Molly Uyeda^{1, 2}, Jason Kung^{2, 3} & Manisha Ramchandani^{1, 2}

¹Division of Oral Biology & Medicine, UCLA School of Dentistry, Los Angeles, CA 90095-1668; ²Evidence-Based Study Group, EBD-Practice-Based Research Network; ³Zen Dental Center - Private Practice, Campbell, CA; Francesco Chiappelli - Email: fchiappelli@dentistry.ucla.edu; *Corresponding author

Received June 30, 2012; Accepted July 06, 2012; Published July 21, 2012

Abstract:

Current trends in bio-medicine include research synthesis and dissemination of bioinformation by means of health (bio) information technology (H[b] IT). Research must secure the validity and reliability of assessment tools to quantify research quality in the pursuit of the best available evidence. Our concerted work in this domain led to the revision of three instruments for that purpose, including the stringent characterization of inter-rater reliability and coefficient of agreement. It is timely and critical to advance the methodological development of the science of research synthesis by strengthening the reliability of existing measure of research quality in order to ensure H[b] IT efficacy and effectiveness.

Keywords: Health (bio) information technology (H[b] IT), research quality validation, reliability, acceptable sampling analysis, efficacy, effectiveness

Background:

Current trends in bio-medicine include timely and critical new developments in the assimilation, synthesis and dissemination of bioinformation. Health (bio) information technology (H[b] IT) has now gained universal recognition across all fields of healthcare worldwide. H[b] IT refers to the cluster of approaches that pertain to the management of health information across computerized systems, and its secure dissemination among stakeholders (i.e., patients, caregivers, clinicians, governmental and private healthcare entities, and healthcare insurance providers). In 2001, the US Institute of Medicine launched a call for establishing electronic documentation systems in all aspects of healthcare. Current trends establish H[b]IT as one among the most promising tools for improving the overall quality, safety as well as effectiveness and efficacy (i.e., efficiency) of the health delivery system [1]. H[b]IT is most simply defined as an application of information

processing that involves both computer hardware and software and that deals with the storage, retrieval, sharing, and use of health care information, data, and knowledge for communication and clinical decision-making. H[b]IT stands today at the forefront of development and innovation in new technologies in hardware (e.g., increased size capacity of servers to store bioinformation; faster and more reliable provider point of entry technologies), in software (e.g., improved electronic health record software for the collection, storage and retrieval of patient information, medical histories, laboratory and imaging data, clinical diagnoses and prognostic observations), and in improved advanced health (bio)information informatics (i.e., synthesis of the research evidence pertaining to the patient's condition into the consensus of the best available evidence). The best available evidence leads to new and improved clinical practice

guidelines, and is translated for dissemination, via H[b] IT, to the stakeholders [2, 3].

Three prominent current trends examples of the importance of H[b] IT in general and of advanced health (bio) information informatics are: (1) British National Health Service Program for IT, NPfIT, 2006; (2) rising budgets of the Agency for Health Research Quality (AHRQ) and the Patient-Centered Outcomes Research Institute (PCORI) since Pres. Obama signed the Patient Protection and Affordable Care Act (PPACA, 2010; upheld, US Supreme Court 06/29/12) [Case in point: the July 2012 National Workshop to Advance the Use of Electronic Data in Patient-Centered Outcomes Research held by PCORI, and the June 2012 PCORI White Paper by Gabriel et al. and the PCORI Methodology Committee, cf. www.pcori.org]; (3) Cochrane organization, now established world-wide and across all continents: the premier entity for generating and disseminating systematic reviews, and establishing the fundamental research synthesis methodology, including the risk of bias assessment tool. (2005-present, development and dissemination of the Risk of Bias tool by the Methods unit, Cochrane Statistical Methods Group, for examining flaws in the design, conduct, analysis, and reporting of research studies, in particular clinical trials, that might cause the effect of an intervention to be underestimated or overestimated. The content validity of the tool is ensured because it covers six domains of possible bias: selection bias, performance bias, detection bias, attrition bias, reporting bias, and other possible biases. Within each domain, qualitative assessments and ranking (i.e., high, moderate, low) are made for one or more items, which may cover different aspects of the domain, or different outcomes. To date, the reliability of the tool is inconclusive and has not been established [4]. Systematic reviews are research reports that are distinct from traditional biomedical research reports in that they consist of a synthesis of all of the available peer-reviewed and non-peer-reviewed research. The objective of systematic reviews is to identify the best research evidence for the clinical outcomes sought. Systematic reviews are also distinct from traditional literature reviews in that they follow a systematic protocol, which characteristically adheres stringently to a patient-centered and patient-tailored research question. Systematic reviews are scientific reports of research synthesis designs strictly driven by the patient's clinical problem, the potential interventions under considerations, and the desired clinical outcome (hence the acronym, P.I.C.O.) [5]. The P.I.C.O. statement yields the necessary keywords and medical subject headings to gain access to all of the pertinent available evidence – the bibliome. The bibliome must be accessed through at least three databases, thus generating redundancy and ensuring capture of all of the available evidence. The bibliome search (i.e., “bibliometric analysis”) is refined to retain only the research that is truly pertinent to the P.I.C.O. question. To ensure reliability, the bibliometric analysis is performed by two or more investigators carefully trained and standardized [6, 7].

Each report in the bibliome is examined for two fundamental questions: 1) “what type of study was performed to obtain the evidence?”, and 2) “how well was the study performed?”. Both questions are of highest relevance to the research synthesis process of the systematic review, because both questions have profound implications for the consensus of the best available evidence that will be obtained. The first question (“what type of

study was performed”) seeks to establish whether the design was an observational study or a clinical trial, an experimental animal or bench study or a diagnostic study, etc. It pertains to the level of the evidence – a higher level of evidence being given to clinical trials than observational studies, for example. The second question (“how well was the study performed”) speaks to whether the reported study followed, or strayed from the fundamental and widely accepted standards of research methodology, design and data analysis – the “quality of the evidence”. The first question addresses “what” was done, whereas the second question points to “how well” it was done [4-6], above and beyond the assessment of the risks of bias. These assessments yield data that are analyzed statistically for: 1) which studies in the research synthesis protocol most adhere to the standards of research methodology, design and data analysis, and hence are acceptable for putative utilization in a patient-centered intervention modality, vs. which studies deviate from said standards and are – based on acceptable sampling statistics [6, 7] – unacceptable for patient treatment, and inclusion in the consensus of the best available evidence; 2) meta-analysis of the acceptable reports to establish the statistically-speaking best available evidence. When appropriate the stated clinical relevance among all acceptable reports may also be extracted and analyzed statistically by means of thematic inference [8].

Having thusly obtained the consensus of the best available evidence, systematic review findings and conclusions are disseminated to providers, patients, caregivers and other stakeholders [2, 3].

Of all the steps outlined above, the greater challenge to H (b) IT presently remains establishing the reliability of assessment of the quality of the evidence. Most of the scales available for this purpose are derived, or expanded from the original JADAD scale [9], and are often limited to rating subject randomization, blinding, and drop-out. These domains are hardly representative of the vast number of criteria that establish the standards of research methodology, design and data analysis. Other available instruments often suffer from fundamental flaws of reliability, including most importantly standardization of the readers as noted by Hartling and collaborators in a 2012 AHRQ report (“Inter-rater variability resulted more often from different interpretation of the tool rather than different information identified in the study reports...”) [10]. The report further states in no uncertain terms the “need to determine inter-rater reliability and validity in order to support the uptake and use of individual tools that are recommended by the systematic review community...” (p. 2). In brief, current trends demand an articulated research program for validating these tools, because of the fundamental importance and relevance of sound assessments of research quality to the process of obtaining the best available evidence for H[b] IT.

With respect to the validity of such instruments, our EBD Study Research Group (EBD-PBRN) has considered the meta-construct of “research quality” as consisting of three constructs: methodology, design, and data analysis. We populated each domain with content specific to the fundamental standards of methodology, design, and analysis. Individual items in each group were crafted to incorporate criteria of excellence of sampling & allocation, instrumentation, and validity

(methodology), trials, observational and diagnostic designs (design), and descriptive and inferential statistics (data analysis) [11]. To ensure reliability, we preceded along the stringent protocol outlined in **Table 1 (see supplementary material)**. Stability (intra-rater coefficient) and homogeneity (Cronbach alpha internal consistency coefficient) proffered little or no added information to this process. Through classic test theory, we revised the scoring protocols of the AMSTAR [6] and the GRADE instruments [7], common tools for assessing quality of systematic reviews and of primary research. We did not alter the item content, thus preserving the instruments' content and construct validity. We augmented the scope of the GRADE instrument with a sub-scale targeted to assess clinical relevance [7], and validated that sub-scale following the protocol outlined above. By means of a process of qualitative factor and cluster analysis of the literature, we identified a set of four criteria for each item in each scale, such that one point was attributed for each criterion satisfied – thus each item had now a semi-continuous range of measurement from 1 to 4. By proceeding through the steps a-f outlined above, we verified the reliability of our revised AMSTAR (R-AMSTAR, 6), and expanded GRADE (Ex-GRADE, 7), independently from a similar line of work engaged by AHRQ, which however yielded less definitive outcomes than ours [12]. The scores for each item across the bibliome could be analyzed for acceptable sampling by means of the Freedman nonparametric test.

In brief, our work offers a research protocol to establish and improve the reliability of research quality assessment tools. Current trends in H[b]IT, clearly have the potential to enhance the efficacy and the effectiveness of patient-centered care, “despite marked heterogeneity in study characteristics and quality, substantial evidence exists confirming that health IT applications with [patient-centered care] PCC-related components have a positive effect on health care outcomes...” [13]. The available tools now demand further applications of the approach outlined in **Table 1 (see supplementary material)** for improving the fundamental methodology of research synthesis, as outlined in cf., “Our Questions, Our Decisions: Standards for Patient-centered Outcomes Research”, PCORI Methodology Committee, Spring 2012, and specifically in terms of the

validation and reliability characterization of study quality instruments e.g., AGREE-II, [14].

Acknowledgement:

The authors thank the Evidence-Based Decisions Active Groups of Stakeholders (EBD-AGS) and the EBD Study Group of the Evidence-Based Decisions in Dentistry-Practice-Based Research Network (EBD-PBRN.org) for their invaluable and timely contributions to our work. The authors in particular recognize the critical contributions of pre-dental students Laura Chiu, Muniza Siqqidi, Nora Godhousi, Raveena Mandawer, who acted as second readers in our reliability studies. The authors report no funding in support of the research presented here, and no conflicts of interest.

References:

- [1] Chaudhry B *et al.* *Ann Intern Med.* 2006 **144**: 742 [PMID: 16702590].
- [2] Barkhordarian A *et al.* *Bioinformation.* 2011 **7**: 315 [PMCID:PMC3280503]
- [3] Barkhordarian A *et al.* *Bioinformation.* 2012 **8**: 293 [PMCID:PMC3338971]
- [4] Higgins J *et al.* *BMJ.* 2011 **343**: d5928 [PMCID: PMC3196245]
- [5] Barkhordarian A *et al.* *Patholog Res Int.* 2011: 359242. [PMCID:PMC3108376]
- [6] Kung J *et al.* *Open Dent J.* 2010 **4**: 84 [PMCID: PMC2948145]
- [7] Phi L *et al.* *Open Dent J.* 2012 **6**: 31 [PMCID: PMC3269009]
- [8] Dousti M *et al.* *Dental Hypotheses.* 2011 **2**: 165 [ISSN 2155-8213]
- [9] Moher D *et al.* *Control Clin Trials.* 1995 **16**: 62 [PMID: 7743790]
- [10] Hartling L *et al.* *AHRQ Methods for Effective Health Care.* 2012 #12-EHC039-EF [PMID: 22536612]
- [11] Chiappelli F *et al.* *Evid Based Complement Alternat Med.* 2006 **3**: 411 [PMCID:PMC1697743]
- [12] Owens D *et al.* *J Clin Epidemiol.* 2010 **63**: 513 [PMID: 19595577]
- [13] Finkelstein J *et al.* *AHRQ 2012 Publication No. 12-E005-EF.*
- [14] Brouwers M *et al.* *CMAJ.* 2010 **182**: 1045 [PMCID: PMC2900328]

Edited by P Kanguane

Citation: Chiappelli, *Bioinformation* 8(14): 691-694 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: Protocol to establish the reliability of research quality assessment tools

A	train multiple readers in sharing a common view for the cognitive interpretation of each item
B	train the readers to read and evaluate independently and blindly the quality of a set of papers
C	refine the process until the inter-rater correlation coefficient and Cohen coefficient of agreement are 0.80-0.85 (over 65% shared variance)
D	obtain independent and blind reading assessments from readers on reports under study
E	compute means and standard deviation of scores for each question across the reports, repeat process if the coefficient of variations are greater than 5% (i.e., less than 5% error among the readers across each questions)
