Journal of Clinical Epidemiology 101 (2018) 73−78

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

High and unclear risk of bias assessments are predominant in diagnostic accuracy studies included in Cochrane reviews

Nicola Di Girolamo[a,*], Alexandra Winter[b], Reint Meursinge Reynders[c,d]

[a]*EBMVet, Via Sigismondo Trecchi 20, Cremona, Italy*
[b]*Private Practice of Surgery, Chicago, IL 60654, USA*
[c]*Private Practice of Orthodontics, Via Matteo Bandello 15, 20123 Milan, Italy*
[d]*Department of Oral and Maxillofacial Surgery, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands*

Accepted 9 May 2018; Published online 16 May 2018

Abstract

Objectives: The validity of outcomes of systematic reviews is highly dependent on the extent of bias in the included primary studies. This study reports the risk of bias (ROB) of primary studies included in systematic reviews of diagnostic accuracy.

Study Design and Setting: All systematic reviews of diagnostic accuracy studies published in the Cochrane database in 2015 and 2016 that used the Quality Assessment of Diagnostic Accuracy Studies−2 ROB tool and reported results with the ROB figure were eligible. The primary outcome was the prevalence of ''high'' or ''unclear'' ROB scores for the four Quality Assessment of Diagnostic Accuracy Studies−2 domains: ''patient selection,'' ''index test,'' ''reference standard,'' and ''flow and timing''.

Results: Of 46 eligible reviews, 35 fulfilled the inclusion criteria. A total of 1045 primary studies with 4133 bias assessments were identified. Of those, 56% (2319/4133) were assessed to be at ''high'' or ''unclear'' ROB and 44% (1814/4133) at low ROB. For all domains except ''flow and timing,'' most outcomes were scored as ''high'' or ''unclear'' ROB. A total of 47 (47/1045; 4.5%, 3.4 to 5.9%) primary studies were scored at low ROB for all domains. Older article age was significantly associated with likelihood of ''high'' or ''unclear'' ROB (odds ratio: 1.02; 95% confidence interval: 1.01 to 1.03; $P < 0.001$).

Conclusion: Systematic reviews of diagnostic accuracy are based on studies with a majority of ''high'' or ''unclear'' bias assessments. The age of the articles explained only a small part of the variability of the score assessments, therefore not justifying an a priori exclusion of older articles in systematic reviews. There is an urgent need to improve the quality of design, conduct, and reporting of diagnostic accuracy studies.

Keywords: Risk of bias; Systematic reviews; Diagnostic accuracy; Reproducibility; Study design

1. Introduction

Diagnostic accuracy is a fundamental aspect of medicine. A lack of knowledge of the accuracy of a test may produce serious diagnostic errors, which may affect treatment decisions and patient outcomes. In this cross-sectional study, we explored the quality of diagnostic accuracy studies included in 2 years of Cochrane systematic reviews.

Typically, primary diagnostic test accuracy studies include a series of patients in which categorical results of a test under evaluation (index test) are compared with the results for the current reference standard. A 2×2 table cross-tabulates index and reference test results, providing a summary description of study participants. If a study has limitations in design or conduct, estimates of diagnostic accuracy can differ systematically from the true accuracy, leading to bias [1]. Established sources of bias in diagnostic accuracy studies include case−control design, observer variability, availability of clinical information, reference standard, partial and differential verification bias, demographic features, and disease prevalence and severity [1]. Recently, Korevaar et al. [2] investigated the completeness of reporting of diagnostic accuracy studies using the Standards for Reporting of Diagnostic Accuracy checklist [2]. That study found some improvement in the quality of reporting of diagnostic accuracy studies over the last decade. However, there is evidence suggesting that improved quality of reporting does not necessarily decrease risk of bias

Conflict of interest/Financial disclosure: Nothing to disclose.

* Corresponding author. Tai Wai Small Animal & Exotic Hospital, G/F, Shop C-D, Lap Wo Building, 69-75 Chik Shun Street, Tai Wai, Shatin, Sha Tin, Hong Kong. Tel.: +393292003570.

E-mail address: nicoladiggi@gmail.com (N. Di Girolamo).

https://doi.org/10.1016/j.jclinepi.2018.05.001
0895-4356/

What is new?

Key findings

- Of 4133 risk of bias (ROB) assessments identified in systematic reviews of diagnostic accuracy, 56% (2319/4133) were assessed to be at "high" or "unclear" ROB.
- For all domains except "flow and timing," most outcomes were scored as "high" or "unclear" ROB.
- Forty-seven (4.5%) of 1045 primary studies were scored at low ROB for all domains.

What this adds to what was known?

- Older article age was associated with likelihood of "high" or "unclear" ROB, but explained only a small part of the variability of the score assessments, therefore not justifying an a priori exclusion of older articles in systematic reviews.

What is the impication and what should change now?

- Inclusion of studies with high or unclear risk of bias in systematic reviews and meta-analyses may jeopardize final results.
- There is an urgent need to improve the quality of conducting and reporting of diagnostic accuracy studies.

(ROB) [3]. In the last 10 years, these findings have encouraged a shift toward evaluating the overall quality and transparency of research reports separately from assessments of ROB [4]. This change has been accompanied by a gradual move from checklists and numeric scores toward domain-based assessment tools.

The Cochrane ROB tool for randomized trials, and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool for systematic reviews of diagnostic test accuracy studies are current examples of such domain-based assessment tools [5,6]. The QUADAS tool was originally published in 2003 and refined in 2011 as the QUADAS-2 tool [5,7]. Four key ROB domains are assessed with this tool: "patient selection", "index test", "reference standard", and "flow and timing". For each of these four domains, various signaling questions are used to evaluate "high", "low" or "unclear" ROB, respectively [5]. Answering "yes" to all signaling questions indicates "low" ROB. Answering "no" to one of these questions indicates "potential bias", which should then be further explored for a possible "high" ROB score. A score of "unclear" ROB is assigned when insufficient data are reported to allow for a judgment [5].

It has been suggested that systematic reviews of diagnostic accuracy studies are often characterized by highly heterogeneous results because of differences in the quality of research design and conduct of the included studies [8,9]. We explored this issue in the present study by undertaking a comprehensive evaluation of the current quality of primary diagnostic accuracy studies in Cochrane systematic reviews by quantifying bias scores with the QUADAS-2 tool. Our specific hypothesis was that a large proportion of bias domains in these primary studies would score as "high" or "unclear" ROB.

2. Materials and methods

2.1. Design and outcomes

The present study was a cross-sectional evaluation of diagnostic accuracy reviews published in the Cochrane database of Systematic Reviews. The primary outcome was the prevalence of "high" and "unclear" ROB scores for the four QUADAS-2 domains: "patient selection," "index test," "reference standard," and "flow and timing" among the primary studies included in these diagnostic accuracy reviews.

2.2. Inclusion criteria

Systematic reviews of diagnostic accuracy studies published in the Cochrane database of Systematic Reviews from January 2015 to December 2016 were eligible. We included 2 years of systematic reviews of diagnostic accuracy to provide an updated evaluation of the status of systematic reviews in this field. Systematic reviews were eligible if they (1) included at least one primary study; (2) used the QUADAS-2 ROB tool; and (3) reported ROB scores for at least one primary study with an ROB figure.

2.3. Data extraction

Two investigators (N.D.G., R.M.R.) independently selected eligible reviews and independently extracted data. For each eligible systematic review, the following data were extracted: (1) title; (2) editorial (authorship) group; (3) year of publication; (4) last name of the first author of each primary study included; (5) year of publication of each primary study included; and (6) results of the assessment for the QUADAS-2 domains for each primary study. All the score assessments were extracted, even duplicates (i.e., originated from the same primary study), but the number of studies that were reported multiple time in the same or different systematic review was counted and reported in the results. Data regarding individual articles were only extracted as part of evaluation using the QUADAS-2 ROB figure. All extracted data were imported into a spreadsheet software program. After data extraction, the spreadsheets of