Systematic reviews and meta-analyses of diagnostic studies: a practical guideline

Ramin Sadeghi1 · Giorgio Treglia2

Received: 3 October 2016 / Accepted: 19 December 2016 © Italian Association of Nuclear Medicine and Molecular Imaging 2016

Abstract
Background Actually, systematic reviews and meta-analyses are the cornerstone of evidence-based practice and the number of these evidence-based articles on diagnostic studies is increasing.
Objective The aim of this article is to provide a practical guideline for the researchers who intend to perform a systematic review or meta-analysis of diagnostic studies.
Methods A guideline was prepared according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and Cochrane handbook for systematic reviews of diagnostic test accuracy.
Results Several steps needed for systematic reviews or meta-analyses of diagnostic studies are briefly discussed (i.e. formulating the question of systematic reviews, search strategy, inclusion and exclusion criteria, quality assessment of the included studies, data extraction, pooling diagnostic indices across studies, reporting heterogeneity and discussion of main findings).
Conclusion To publish a high-quality systematic review or meta-analysis of diagnostic test accuracy, certain methodology should be followed. Only methodologically sound systematic reviews or meta-analyses can change or support the clinical use of a diagnostic test.

Introduction
Evidence based medicine is defined as using the best available evidence for everyday clinical practice [1–3]. Synthetic literature plays an important role in evidence based medicine. Actually systematic reviews and meta-analyses are the cornerstone of evidence based practice. The main difference between a systematic review and a narrative review is the clear method of the former including a clear search and predefined inclusion criteria. The methodology of systematic reviews makes them reproducible which is not the case in narrative reviews [1–3]. The number of systematic reviews and meta-analyses on nuclear medicine diagnostic studies is increasing [4, 5] (Fig. 1).

In the current manuscript a practical guideline has been prepared for the researchers who intend to perform a systematic review or meta-analysis of diagnostic studies.

A clear topic for systematic review: formulating the question
The most important step in preparing a synthetic study is to have a clear topic. The topic is usually divided into several aspects including: patients (the population of the study), intervention (the diagnostic test under study), comparison (the procedures comparative to the index test), outcome (the outcome which is going to be evaluated which are usually sensitivity and specificity). The above-mentioned method is called patients-intervention-comparison-
outcome (PICO) [6, 7]. The search strategy of systematic reviews is based on the PICO question. Here are two examples:

1. How does positron emission tomography (PET) [Intervention] work for detection of recurrence [Outcome] in endometrial carcinoma [Patients]?
2. How does sentinel node mapping [Intervention] work for lymph node staging [Outcome] of primary endometrial carcinoma [Patients] as compared to pelvic lymph node dissection [Comparison]?

Which articles should be included? Search strategy, inclusion and exclusion criteria

Search strategy is based on our PICO question. The key words and databases used for searching should be selected to minimize the chance of missing any relevant article. Using Boolean operators (i.e. AND, OR, NOT) is inevitable. This makes your search as sensitive as possible. For example for the above-mentioned PICO questions the following key words seem to be optimal:

1. PET AND (endometrial OR endometrium OR uterine) AND recurrence.
2. Sentinel AND (endometrial OR endometrium OR uterine).

At least two databases should be included in the search strategy. Pubmed/Medline and SCOPUS (or EMBASE) are two main important sources for any systematic review.

The inclusion and exclusion criteria should be as clear as possible too. The following factors should be considered to set useful inclusion criteria.

1. Standard of reference: included studies should describe the reference or gold standard with which the diagnostic test is compared.
2. Outcome data: enough information should be available to reconstruct a 2×2 diagnostic table of each study.
3. Language and time limit: preferably no language or time limit should be imposed.

For example for the above-mentioned PICO questions, the following inclusion criteria can be set:

1. All studies which compared PET with conventional imaging for detection of recurrence in endometrial cancer.
2. All studies which compared sentinel node mapping with pelvic lymph node dissection in endometrial cancer.

Full-texts of all relevant studies should be retrieved. The reference of primary studies and all relevant reviews should be checked to search for additional primary studies that could have been missed. Remember to keep the
records of all the searches, as well as included and excluded studies.

Quality assessment of the included studies

Not all included studies are of same quality. Quality of each study should be checked and reported. Several checklists are available for diagnostic studies [8, 9]. Two of the most commonly used checklists are:

2. Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) [10]. QUADAS-2 is the revised, 2011 version of the 2003 QUADAS and consists of 4 dimensions (patient selection, index test, reference standard, and finally, flow and timing), the first three of which requires an answer among the three available responses (yes/high, no/low, and unclear).

Data extraction

All relevant data should be extracted from the included studies. Detailed information regarding the study population, method of the diagnostic test, gold standard test, outcome variables such as false and true negative (FN, TN), false and true positive (FP, TP) cases should be extracted. Extraction of data should be as complete as possible to allow reconstruction of $2 \times 2$ diagnostic tables as well as sub-group analyses [11, 12].

Pooling diagnostic indices across studies and reporting heterogeneity

In this final step, the numerical results of the included studies would be pooled together. First of all, diagnostic indices of each included study should be presented. The following diagnostic indices should be reported:

- Sensitivity = $TP / TP + FN$
- Specificity = $TN / TN + FP$
- Positive likelihood ratio ($LR_+$) = sensitivity / (1 - specificity)
- Negative likelihood ratio ($LR_-$) = (1 - sensitivity) / specificity
- Diagnostic odds ratio (DOR) = $LR_+ / LR_-$

Meta-analysis is a special statistical method for pooling data across different studies and giving pooled diagnostic indices. For this purpose, a weight is attributed to each study and the weighted diagnostic indices are pooled together. Fortunately, special softwares are available for this purpose, including SAS, R and STATA. OpenMeta [Analyst] is another free software for meta-analysis of diagnostic studies. This software is available online at http://www.cebm.brown.edu/openmeta/downloads/open_meta_analyst_win8.zip [13].

The least required data to be provided in a meta-analysis are:

1. Pooled indices: they can be perfectly reported by forest plots which gives all included studies as well as the pooled data in one view (Fig. 2 is an example of a forest plot).
2. Pooling method: we recommend random effects model for pooling studies as fixed model would not account for heterogeneity among included studies [14].

Fig. 2 An example of a forest plot. The squares are sensitivity of individual studies. The size of each square is proportional to their sample size. The blue lines on the sides of the squares represent the confidence intervals of the effect size of each study (in this figure the effect size is sensitivity). The diamond represents the pooled sensitivity and its horizontal diameter represents its confidence interval.

<table>
<thead>
<tr>
<th>Sensitivity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>G</td>
</tr>
</tbody>
</table>

Pooled Sensitivity = 0.50 (0.46 to 0.54)
Chi-square = 31.71; df = 6 (p = 0.0000)
Inconsistency (I-square) = 81.1%
3. **Heterogeneity**: included studies of a systematic review are different from each other on several accounts such as studied population, methodology of the diagnostic test, etc. Several factors contribute to the heterogeneity among studies: sampling error of the individual studies including true differences between included studies and finally the threshold effect [15, 16]. Methods for undertaking analyses which account for both sensitivity and specificity, the relationship between them, and the heterogeneity in test accuracy, require fitting hierarchical random effects models [17].

4. **Threshold effect**: a unique source of heterogeneity in meta-analysis of diagnostic studies is the threshold effect. Not all studies use the same cut-off value for a positive result. This can be due to an explicit cut-off point value or explicit human or instrumental factors. This should be addressed in all diagnostic meta-analyses. Although the summary receiver operating characteristic curve (SROC) method and reporting $Q^*$ have been used traditionally for evaluating the threshold effect in diagnostic studies, the best way to report the possible effect of threshold effect is bivariate meta-analyses [18, 19]. In this method, correlation between specificity and sensitivity is used as a variable to correct the results of the meta-analyses for possible threshold effect. This method has been incorporated in the last version of OpenMeta [Analyst] and can be easily reported (Fig. 3). The traditional SROC method is no longer recommended.

5. **Publication bias**: although there is substantial literature relating to publication bias in systematic reviews and meta-analyses of randomized controlled trials [20], little research has been done in the context of systematic reviews and meta-analyses of diagnostic studies [17].

Discussion and conclusion of systematic reviews

The discussion and final conclusion of a systematic review and meta-analysis should be as objective as possible. The authors should discuss the main results of the systematic review and meta-analysis. Final conclusion should be based on the main results of the systematic review.

Any heterogeneity of the included studies should be explained and the possible reasons should be discussed.

Standard method of reporting systematic reviews and meta-analyses

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) provides a minimum requirement for reporting systematic reviews and meta-analyses.
[21]. Although it is originally prepared for systematic reviews of randomized clinical trials, systematic reviews of diagnostic accuracy studies can be reported using PRISMA too. PRISMA statement and checklist can be found in the following link: http://www.prisma-statement.org/.

Furthermore Cochrane handbook for systematic reviews of diagnostic test accuracy is another resource freely available at the following link: http://methods.cochrane.org/sdt/handbook-dta-reviews.

Final comment

To publish a high quality systematic review of diagnostic test accuracy, certain methodology should be followed. Only methodologically sound systematic reviews are worth publication and can change or support clinical use of a diagnostic test. Hopefully, the above-mentioned methodology could help the researchers through the process of systematic review preparation.

Compliance with ethical standards

Conflict of interest All authors (Ramin Sadeghi and Giorgio Treglia) declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References