



# A Practical Guideline on Diagnostic and Prognostic Meta-Analyses

# 2

Ramin Sadeghi and Giorgio Treglia

## 2.1 Introduction

Evidence based medicine is defined as using the best available evidence for everyday clinical practice [1–3]. Synthetic literature including systematic reviews and meta-analyses plays an important role in evidence based medicine. Essentially 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 and prognostic studies is increasing [4,

5]. In the current chapter, a practical guideline has been prepared for the researchers who intend to perform a systematic review or meta-analysis of diagnostic and prognostic studies.

## 2.2 A Clear Topic for Systematic Review: Formulating the Question

The single most important step in preparing a systematic review 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 or a prognostic factor which is being evaluated), comparison (the procedures comparative to the index test), outcome (the outcome which is going to be evaluated which are usually sensitivity and specificity for diagnostic studies and overall survival (OR) and progression free survival (PFS) in prognostic ones).

The abovementioned method is called patients-intervention-comparison-outcome (PICO) [6, 7]. The search strategy for 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]?

R. Sadeghi (✉)

Nuclear Medicine Research Center, Mashhad  
University of Medical Sciences, Mashhad, Iran  
e-mail: [sadeghir@mums.ac.ir](mailto:sadeghir@mums.ac.ir)

G. Treglia

Clinic of Nuclear Medicine and Molecular Imaging,  
Imaging Institute of Southern Switzerland, Ente  
Ospedaliero Cantonale,  
Bellinzona and Lugano, Switzerland

Department of Nuclear Medicine and Molecular  
Imaging, Lausanne University Hospital and  
University of Lausanne, Lausanne, Switzerland

Health Technology Assessment Unit, Academic  
Education, Research and Innovation Area, Ente  
Ospedaliero Cantonale, Bellinzona, Switzerland  
e-mail: [giorgio.treglia@eoc.ch](mailto:giorgio.treglia@eoc.ch)

2. What is the prognostic significance of PET [Intervention] for predicting survival [Outcome] in renal cell carcinoma [Patients]?

### 2.3 Which Articles Should Be Included? Search Strategy, Inclusion and Exclusion Criteria

Search strategy is based on our PICO question. The keywords and databases which are used for searching should minimize the chance of missing any relevant article. Using Boolean operators (i.e., AND, OR, NOT) is highly recommended. This makes your search as sensitive as possible.

For example for the abovementioned PICO questions the following keywords seem to be optimal:

1. (PET OR “Positron Emission Tomography”) AND (endometrial OR endometrium OR uterine) AND recurrence.
2. (PET OR “Positron Emission Tomography”) AND (RCC OR “renal cell carcinoma” OR kidney).

At least two databases should be included in the search strategy. PubMed/Medline and SCOPUS (or EMBASE) are two main 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:

- (a) Standard of reference: Included studies should describe the reference or gold standard with which the diagnostic test is compared.
- (b) Outcome data: Enough information should be available to reconstruct a  $2 \times 2$  diagnostic table or prognostic factors (such as hazard ratio (HR)) of each study.
- (c) Language and time limit: Preferably no language or time limit should be imposed.

For example for the abovementioned 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 evaluated the prognostic significance of PET factors (SUVmax, SUVmean, etc.) in survival (OS or PFS) of renal cell carcinoma patients.

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 (backward searching of the citations). In addition, articles citing the relevant included articles can be used to find any other missing articles (forward searching of the citations). The citing articles can be found easily using Google Scholar (<https://scholar.google.com/>), SCOPUS, or Dimensions (a free newly launched application with many useful options: <https://app.dimensions.ai/discover/publication>).

Remember to keep the records of all the searches, as well as included and excluded studies.

### 2.4 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:

1. Oxford Center for Evidence Based Medicine worksheet for diagnostic studies (available at <https://www.cebm.net/wp-content/uploads/2018/11/Diagnostic-Accuracy-Studies.pdf>).
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 four dimensions (patient selection, index test, reference standard, and finally, flow and timing), the first three of which require an answer among the three available responses (yes/high, no/low, and unclear).

Several checklists are also available for prognostic studies [11].

Two of the most commonly used checklists are:

1. Oxford Center for Evidence Based Medicine worksheet for prognostic studies (available at <https://www.cebm.net/wp-content/uploads/2018/11/Prognosis.pdf>).
2. QUIPS tool (quality in prognostic factor studies) [12]. QUIPS has several domains (study participation, study attrition, prognostic factor measurement, outcome measurement, adjustment for other prognostic factors, and statistical analysis). Risk of bias in each domain should be rated as high or low.

Checklists are usually equivalent to each other; however, each domain or dimension for all included studies should be explained in detail to give the reader of systematic reviews a clue regarding the quality of the included studies. Reporting only based on numbers (quality scores) should be discouraged.

---

## 2.5 Data Extraction

All relevant data should be extracted from the included studies. Detailed information regarding the study population, method of the diagnostic or prognostic test, gold standard test, follow-up times, methods of ascertaining outcomes, outcome variables such as false and true negative (FN, TN), false and true positive (FP, TP) cases for diagnostic studies, and hazard ratios (HR) for OS and PFS for prognostic studies should be extracted. Extraction of data should be as complete as possible to allow reconstruction of  $2 \times 2$  diagnostic tables or HR of prognostic factors as well as sub-group analyses [11, 13, 14].

Extraction of data in prognostic studies can be very tricky: not all studies reported HR, and only Kaplan Meier (KM) curves and associated log rank tests are usually reported. HR can be extracted from KM curves. Usually the survival data can be extracted manually from KM curves using special software such as getdata graph digitizer (available at <http://getdata-graph-digitizer.com/download.php>). Finally the extracted survival data can be converted to HR by Parmar method using a special Excel file provided by Parmar et al. [15].

Another important aspect of extraction data in prognostic systematic reviews is type of prognostic factor (quantitative vs. qualitative factors) and evaluation of other prognostic factors (multivariate vs. univariate analysis). HR of quantitative variables (such as SUVmax) can be provided in two ways: first, the prognostic factor can be used as a quantitative variable and a HR using Cox regression is provided. The second type of HR can be calculated by categorizing a quantitative variable into two ranks (for example, SUVmax  $>7$  and  $\leq 7$ ). These two types of HR cannot be pooled with each other even for a same prognostic factor. In addition, only univariate or multivariate HR should be used for pooling data across studies. Pooling univariate HR with a multivariate HR is discouraged as the latter (but not the former) takes into account other potential prognostic factors.

---

## 2.6 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 or prognostic 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+) =  $\text{sensitivity}/(1 - \text{specificity})$

- Negative likelihood ratio (LR-) =  $(1 - \text{sensitivity})/\text{specificity}$
- Diagnostic odds ratio (DOR) =  $\text{LR-}/\text{LR+}$

The following prognostic indices should be reported:

- Log rank test and  $p$ -value of each prognostic factor (only available for categorical variables).
- Hazard ratios (HR) for prognostic factors.
- Univariate and multivariate HR (if available) should be reported. In addition, for quantitative variables HR for the quantitative variable as well as HR for categorized variable (if available) should be reported (see the previous section for more information).

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. Special software are available for this purpose, including SAS, R, and STATA.

For diagnostic studies, two free software are available:

1. OpenMeta [Analyst] is a free software for meta-analysis of diagnostic studies. This software is available online at [http://www.cebm.brown.edu/openmeta/downloads/OpenMeta-analyst\\_Windows.zip](http://www.cebm.brown.edu/openmeta/downloads/OpenMeta-analyst_Windows.zip) [16].
2. Meta-Disc is another free software which has been specially designed for diagnostic studies. This software is available online at <https://download.freownloadmanager.org/Windows-PC/Meta-DiSc/FREE-1.4.html> [17].

For prognostic studies, usually hazard ratios should be pooled across included studies. Several software are available in this regard, such as R, SAS, and Comprehensive Meta-Analysis (CMA).

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

1. Pooled indices: They can be perfectly reported by forest plots which give all included studies as well as the pooled data in one view.

2. Pooling method: We recommend random effects model for pooling studies as fixed model would not account for heterogeneity among included studies [18].
3. Heterogeneity: Included studies of a systematic review are different from each other on several accounts such as studied population, methodology of the diagnostic tests or prognostic factors, 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 [19, 20]. 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 [21]. To report heterogeneity for each meta-analysis, at least Cochrane  $Q$  value and its associated  $p$ -value and  $I$  squared should be reported. Several methods are available in order to address the heterogeneity across included studies such as sub-group analysis, meta-regression, and sensitivity analysis. The authors should use these methods to explain the underlying reasons of heterogeneity across included studies.
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 [22, 23]. 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. 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, little research has been done in the context of systematic reviews and meta-analyses of diagnostic studies. However, publication bias can be visually presented by funnel plots and can be quantified by several methods such as Egger's regression intercept or trim and fill method [24, 25].

could help the researchers through the process of systematic review and meta-analysis preparation.

## 2.7 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, 26]. 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/>.

## 2.8 Final Comment

To publish a high quality systematic review or meta-analysis of diagnostic test accuracy or prognostic studies, certain methodology should be followed. Only methodologically sound systematic reviews and meta-analyses are worth publication and can change or support clinical use of a diagnostic test or a prognostic factor. Hopefully, the abovementioned methodology

## References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71–2.
2. Sadeghi R, Zakavi R, Kakhki VRD. How to apply the evidence-based medicine concept to nuclear medicine diagnostic studies—a review. *Nucl Med Rev*. 2009;12:59–64.
3. Kianifar H-R, Akhondian J, Najafi-Sani M, Sadeghi R. Evidence based medicine in pediatric practice: brief review. *Iran J Pediatr*. 2010;20:261.
4. Treglia G, Sadeghi R. Meta-analyses and systematic reviews on PET and PET/CT in oncology: the state of the art. *Clin Transl Imaging*. 2013;1:73–5.
5. Treglia G, Sadeghi R, Del Sole A, Giovannella L. Diagnostic performance of PET/CT with tracers other than F-18-FDG in oncology: an evidence-based review. *Clin Transl Oncol*. 2014;16:770–5.
6. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol*. 2011;11:15.
7. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *J R Soc Med*. 2003;96:118–21.
8. Sadeghi R. Evidence based medicine in nuclear medicine practice; part I: introduction, asking answerable questions and searching for the best evidence. *Iran J Nucl Med*. 2009;17(1):41–8.
9. Sadeghi R. Evidence based medicine in nuclear medicine practice; part II: appraising and applying the evidence. *Iran J Nucl Med*. 2009;17:49–56.
10. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–36.
11. Riley RD, Moons KG, Snell KI, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ*. 2019;364:k4597.
12. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158:280–6.
13. Devillé WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol*. 2002;2:9.
14. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120:667–76.
15. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*. 1998;17:2815–34.

16. Wallace BC, Schmid CH, Lau J, Trikalinos TA. Meta-analyst: software for meta-analysis of binary, continuous and diagnostic data. *BMC Med Res Methodol.* 2009;9:80.
17. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol.* 2006;6:31.
18. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc A Stat Soc.* 2009;172:137–59.
19. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21:1539–58.
20. Patsopoulos NA, Evangelou E, Ioannidis JP. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. *Int J Epidemiol.* 2008;37:1148–57.
21. Bae J-M. An overview of systematic reviews of diagnostic tests accuracy. *Epidemiol Health.* 2014;36:e2014016.
22. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005;58:982–90.
23. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med.* 1993;12:1293–316.
24. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics.* 2000;56:455–63.
25. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315:629–34.
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med.* 2009;151:264–9.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

