**Exploring possible sources of bias in diagnostic test accuracy beyond the QUADAS (Quality Assessment of Diagnostic Accuracy Studies): a study protocol for a meta-epidemiological study**

**Authors and affiliation:**

Akira Onishi, MD, PhD, MPH,[1] Tomoko Fujii, MD,[2,3] Yusuke Tsutsumi, MD,[4] and Toshi A Furukawa, MD, PhD[5]

[1]Department of Rheumatology and Clinical Immunology, Kobe University Graduate School of Medicine, 7-5-1 Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan.

[2]Department of Epidemiology and Preventive Medicine, Graduate School of Medicine, Kyoto University, Yoshida Hon-machi, Sakyo-ku, Kyoto 606-8501, Japan

[3]Japan Society for the Promotion of Science, Kojimachi Business Center Building, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-0083, Japan

[4]Department of Healthcare Epidemiology, Graduate School of Medicine, Kyoto University, Yoshida Hon-machi, Sakyo-ku, Kyoto 606-8501, Japan

[5]Department of Health Promotion and Human Behavior, Graduate School of Medicine, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

**Correspondence to:**

Akira Onishi, MD, PhD, MPH. Department of Rheumatology and Clinical Immunology, Kobe University Graduate School of Medicine, 7-5-1 Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan.

Phone: 81-78-382-6207; Fax: 81-78-382-6208; E-mail: aonishi@med.kobe-u.ac.jp.

## ABSTRACT

### Introduction

Meta-epidemiological research to date of diagnostic test accuracy (DTA) studies have elucidated several characteristics included in the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool, such as patient selection, index test, reference standard or flow and timing of the study, which may pose important risks of bias. The influence of other potentially important methodologic and non-methodologic characteristics has not been examined, including sample size, number of study centers, funding source, publication year or publication status. The purpose of our study is to explore possible sources of bias in DTA beyond the QUADAS tool.

### Methods and analysis

We will include systematic reviews (SRs) of DTA with meta-analyses that include 10 or more primary studies published in the Cochrane Database of Systematic Reviews. We will independently screen, retrieve SRs and extract data. We will estimate the relative diagnostic odds ratio comparing each study characteristics in each meta-analysis. Study characteristics to be estimated will include publication year (before or as of 2004), impact factor of publication journal, country (high or middle/low income), type of language (English or not), sample size (<100 or ≥100), healthcare setting (primary healthcare or not), funding source (industrial sponsorship or not), publication status (full-publication or not), and number of study centers (single center or multicenter).

### Ethics and dissemination

This meta-epidemiological study does not require ethical approval because we use only secondary data from published reports. We will publish the findings of this study in a peer-reviewed scientific journal and may present them at scientific conferences.

Title registration number: UMIN000033079

## INTRODUCTION

Many diagnostic tests have been developed and the number of studies that examine diagnostic test accuracy (DTA) has been increasing. Because diagnostic strategies for patient management and policymaking needs reliable evidence, diagnostic tests should be evaluated thoroughly and DTA studies are essential in this process. However, only a few studies of DTA fulfill essential methodological quality and the reporting quality of DTA studies was moderate.[1]

Empirical evidence of study characteristics influencing the quality of the DTA studies is scanter than that for intervention studies. For treatment, meta-epidemiological studies have shown that lack of allocation concealment, lack of blinding, or excluding patients from the analysis might lead to underestimation or overestimation of the treatment effect.[2-4] In addition, empirical evidence suggested other study characteristics, such as the number of the study centers, trial sample size, and the status of the study outcome (surrogate or final relevant outcome), were associated with intervention estimate effect.[5-7] By contrast, meta-epidemiological studies of DTA are still limited. Meta-epidemiological research to date of DTA studies have elucidated several characteristics included in the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool,[8 9] such as patient selection, index test, reference standard or flow and timing of the study, which may pose important risks of bias.[10-12] To the best of our knowledge, however, the influence of other potentially important methodologic and non-methodologic characteristics has not been examined, such as sample size, number of study centers, funding source, publication year or publication status.

The purpose of our study is to explore possible sources of bias in DTA beyond the

QUADAS tool.

**Strengths and limitations of this study**

- This is the first study to explore possible sources of bias in diagnostic test accuracy studies beyond the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool.
- Our results will provide what type of studies should be interpreted with caution for researchers and clinical decision-makers.
- Our results will be based on only systematic reviews (SRs) published in the Cochrane Database of Systematic Reviews, although Cochrane SRs have generally higher methodological and reporting quality compared with non-Cochrane SRs.
- Our sensitivity analyses to control for methodological quality might have limitations because the risk of bias domains may be poorly reported and discrepancies might exist between the reported and actual quality.

## METHODS AND ANALYSIS

### Eligibility criteria for considering systematic reviews for this study

We will include SRs in which the authors evaluate the DTA of one or more tests against a reference standard. We will only include DTA reviews with meta-analyses that include 10 or more primary studies and provide data to reconstruct 2×2 tables. We will exclude narrative reviews, genomic reviews, animal reviews, reviews that apply a language or quality restriction, reviews that assess the analytical validity of tests, and reviews that only evaluate other measures of diagnostic performance such as reproducibility and reliability. We will place no restrictions on language of publication, test type, purpose of the test, setting, or disease area.

### Data Sources and Searches

We will search the Cochrane Database of Systematic Reviews (CDSR) because Cochrane reviews and their protocols are peer reviewed and adhere to a structured methodological approach and format. SRs coded as diagnostic in the CDSR will be identified to identify relevant SRs of DTA.

**Review Selection**

Three reviewers (AO, TF, YT) will independently screen titles and abstracts of all the potential SRs identified by the electronic search for relevance. We will retrieve the full-text SRs and three review authors (AO, TF, YT) will independently screen the full text and identify SRs for inclusion. We will resolve any disagreement through discussion or, if required, we will consult a third person (TAF)

**Data Extraction**

Three review authors (AO, TF, YT) will extract independently the following characteristics and outcome data of primary studies from the included SRs: the author, publication year, publication journal, impact factor of publication journal, country, type of language (English or not), test under evaluation, type of test evaluated (physical examination, laboratory tests, or imaging), role of the test in diagnostic pathway (replacement, triage, or add-on),[13] target condition, patient population, healthcare setting (primary healthcare or not), reference standard, data to populate two-by-two tables (the number of true positives, false positives, false negatives, and true negatives), risk of bias (patient selection, index test, reference standard, and flow and timing) using the QUADAS-2,[9] funding source (industrial sponsorship or not), publication status (full-publication or not), and number of study center (single center or multicenter). If QUASAS[8] or part of QUADAS is used in a SR, we will use part of QUADAS items and extract additional characteristics to judge QUADAS-2 items. When the abovementioned information is not reported in the SRs, we will extract data from the primary studies

ourselves. We will contact the authors of primary studies when the information is not reported in the primary studies. In case of no response, we will send two reminders requesting the missing data.

Where multiple tests for the same target condition are evaluated in an SR, we will choose the test with the largest number of primary studies, and if the tests include the same largest number of primary studies, we will choose the test described first. We will exclude DTA of combined multiple tests.

**Data Synthesis and Analysis**

The analyses will be undertaken in 2 parts. First, we will consider whether evidence shows differences between each characteristic of primary studies in comparisons of overall accuracy measured as diagnostic odds ratios (DORs). Second, we will consider the magnitude of the difference in terms of measures of sensitivity and specificity.

Hierarchical summary receiver-operating characteristic (HSROC) models will be used to estimate the relative DORs (RDORs) comparing each study characteristics in each meta-analysis.[14] Study characteristics to be estimated will include publication year (before or as of 2004),[15] impact factor of publication journal (lower or median and higher than median impact factor among primary studies from one SR), country (high income, middle/low income, or both),[16] type of language (English or not), sample size (<100 or ≥100), healthcare setting (primary healthcare or not), funding source (industrial sponsorship or not), publication status (full-publication or not), and number of study center (single center or multicenter). Definition of high-income countries are based on the list of advanced economies of the International Monetary Fund. Israel is also classified as a high-income country regardless of period because it has a longstanding tradition in clinical research.[16] All other countries are considered as middle/low-income countries. When a study is not fully published, the presentation year or the end of the study will be used instead of publication year and the impact factor of

publication journal will be regarded as zero. The HSROC model allows studies using different test thresholds to be included in the same meta-analysis by estimating summary ROC curves. We will estimate the RDORs of each study characteristics by including a variable indicating it as a covariate in the HSROC model. They indicate the diagnostic performance of a test in studies with each study characteristics, relative to its performance in studies failing to satisfy the corresponding feature. If the RDOR is larger than 1, studies with the characteristics yield larger estimates of the DOR than studies not satisfying this corresponding feature. The average bias in estimates of study characteristics across the topics will be computed by a second-level meta-analysis of RDORs by using a random-effects DerSimonian and Laird inverse variance weighted model.[17]

Finally, because meaningful estimates of differences in average sensitivity and specificity can be obtained only when studies share a common test threshold, we will identify the subset of studies in which common or consistent thresholds for each test are used. We will additionally estimate differences in sensitivity and specificity between each study characteristics in HSROC models.

To control for potential confounding by methodological quality, we will also conduct sensitivity analyses adjusting for the different domains of the risk of bias tool (patient selection, index test, reference standard, and flow and timing).[9]

We will conduct subgroup analyses based on the type of test evaluated (physical examination, laboratory tests, or imaging), the role of the test in the diagnostic pathway (replacement, triage, or add-on)[13] to investigate the existence of factors that potentially influence RDORs.

Analyses will be performed by using R 3.4.1 (R foundation for Statistical Computing) and Stata software, version 13.0 (Stata Corp., College Station, Texas).

## ETHICS AND DISSEMINATION

This protocol has been registered in the University Hospital Medical Information Network (UMIN) Clinical Trials Registry (Trial registration number: UMIN???). This meta-epidemiological study does not require ethical approval because we use only secondary data from published reports. We will publish the findings of this study in a peer-reviewed scientific journal and may present them at scientific conferences.

## DISCUSSION

To the best of our knowledge, this is the first study to assess empirically the impact of study characteristics beyond the QUADAS tool on the estimates of DTA. Our results will also be based on SRs representing various medical areas. Because it is controversial whether synthesizing all primary studies included in SRs provides the best estimate of DTA, or overestimate or underestimate DTA, our study will provide a clue to this question. In addition, our results will also suggest what study characteristics should be interpreted with caution for reviewers and readers.

There are some limitations for this study. First, the generalizability will be limited because our results will be based on only SRs published in the CDSR. Cochrane SRs, however, have generally higher methodological and reporting quality compared with non-Cochrane SRs.[18 19] Second, our sensitivity analyses to control for methodological quality might also have limitations because the risk of bias domains may be poorly reported[1] and discrepancies might exist between the reported and actual quality.

In conclusion, this study will provide the first investigation to assess possible sources of bias in DTA beyond the QUADAS tool. Our findings may have important implications for researchers and clinical decision-makers.

## Acknowledgement

We would like to thank the valuable comments from the members of Research Group on Meta-epidemiology at Kyoto University School of Public Health (Miho Kimachi

Yuki Kataoka, Yasushi Tsujimoto, Yan Luo, Aran Tajika, Kiyomi Shinohara and Kenji Omae)

**REFERENCES**


1. Korevaar DA, van Enst WA, Spijker R, et al. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med* 2014;19:47-54.
2. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Jama* 1995;273:408-12.
3. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol* 2005;34:79-87.
4. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *Bmj* 2008;336:601-5.
5. Bafeta A, Dechartres A, Trinquart L, et al. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. *Bmj* 2012;344:e813.
6. Ciani O, Buyse M, Garside R, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *Bmj* 2013;346:f457.
7. Dechartres A, Trinquart L, Boutron I, et al. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *Bmj* 2013;346:f2304.
8. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in

systematic reviews. *BMC Med Res Methodol* 2003;3:25.

9. 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.
10. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *Jama* 1999;282:1061-6.
11. Whiting P, Rutjes AW, Reitsma JB, et al. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189-202.
12. Rutjes AW, Reitsma JB, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies. *Cmaj* 2006;174:469-76.
13. Bossuyt PM, Irwig L, Craig J, et al. Comparative accuracy: assessing new tests against existing diagnostic pathways. *Bmj* 2006;332:1089-92.
14. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865-84.
15. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Bmj* 2003;326:41-4.
16. Panagiotou OA, Contopoulos-Ioannidis DG, Ioannidis JP. Comparative effect sizes in randomised trials from less developed and more developed countries: meta-epidemiological assessment. *Bmj* 2013;346:f707.
17. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
18. Moja LP, Telaro E, D'Amico R, et al. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional

study. *Bmj* 2005;330:1053.

19. Moseley AM, Elkins MR, Herbert RD, et al. Cochrane reviews used more rigorous methods than non-Cochrane reviews: survey of systematic reviews in physiotherapy. *J Clin Epidemiol* 2009;62:1021-30.