

[Review Article]

診断精度研究のバイアスリスク評価ツールQUADAS-2 : a Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies 2 の活用

Utilization of QUADAS-2 ; a Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies 2 as a Risk of Bias Evaluation Tool



小島原典子^{*1,2} 森實 敏夫^{*1,3} 中山 健夫^{*1,4} 福岡 敏雄^{*1,5}
河合富士美^{*1,6} 佐藤 康仁^{*1,2} 吉田 雅博^{*1,7} 山口 直人^{*1,2}

ABSTRACT

The QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool, a quality assessment tool of diagnostic studies was developed in 2003, and thereafter the Cochrane Collaboration recommended QUADAS for use in systematic reviews of diagnostic accuracy. Following further revision, mainly by the GRADE (Grading of Recommendations Assessment, Development and Evaluation) working group, QUADAS-2 was released in 2011 (<http://www.bris.ac.uk/quadas/>). With the permission of the original authors, Japanese language localization was undertaken and released on the MINDS website in April 2014 (<http://minds.jcqhc.or.jp/>), and release on the QUADAS website is scheduled for May 2014.

The original QUADAS tool includes 14 items assessing bias risk, applicability, and for reporting quality. The QUADAS-2 tool consists of 4 phases : 1) summarization of the review question, 2) tailoring of the tool and the production of review-specific guidance, 3) construction of a flow diagram for the primary study, 4) assessment of bias and applicability. The last phase is applicable to 4 domains : patient selection, index test, reference standards, and flow and timing. Each domain is assessed according to bias risk using signalling questions of “yes,” “no,” or “unclear.” The first 3 domains are also assessed in terms of applicability. This tool will facilitate increased transparency in the rating of bias and improve applicability in the systematic review of diagnostic accuracy studies.

(Jpn Pharmacol Ther 2014 ; 42 suppl 2 : s127-34)

^{*1}公益財団法人日本医療機能評価機構

^{*2}東京女子医科大学衛生学公衆衛生学第二

^{*3}東邦大学医学部

^{*4}京都大学大学院医学研究科社会健康医学系専攻健康情報学分野

^{*5}公益財団法人大原記念倉敷中央医療機構倉敷中央病院総合診療科

^{*6}聖路加国際大学学術情報センター図書館

^{*7}国際医療福祉大学臨床医学研究センター

Noriko Kojimahara^{*1,2}, Toshio Morizane^{*1,3}, Takeo Nakayama^{*1,4}, Toshio Fukuoka^{*1,5}, Fujimi Kawai^{*1,6}, Yasuto Sato^{*1,2}, Masahiro Yoshida^{*1,7}, Naohito Yamaguchi^{*1,2}

^{*1}Japan Council for Quality Health Care, ^{*2}Department of Public Health, Tokyo Women's Medical University,

^{*3}Faculty of Medicine, Toho University, ^{*4}Department of Health Informatics, Kyoto University School of Public Health,

^{*5}Department of General Medicine, Kurashiki Central Hospital, Ohara HealthCare Foundation, ^{*6}St. Luke's International University Library,

^{*7}Center for Clinical Science, International University of Health and Welfare

KEY WORDS QUADAS-2 : a revised tool for the quality assessment of diagnostic accuracy studies 2, signaling questions, diagnostic test accuracy (DTA), index test, reference standard

背 景

診療ガイドライン作成などを目的としてシステムティックレビューの必要性が高まっているが、診断のシステムティックレビューは治療と比較して方法論の確立が遅れており、透明性の高い作成方法を適用している論文は少ない^{1,2)}。診断のシステムティックレビューの実行手順は、①重要臨床課題に基づくPICOを設定、レビュークエスチョンごとに②網羅的にエビデンスを収集、③エビデンスの質を評価、④アウトカムごとにエビデンスの強さを評価する、⑤エビデンスの統合であり、治療のシステムティックレビュー³⁾とアウトラインは同じだが、研究対象の選定、文献検索方法、効果指標、メタアナリシスの手法など、治療のシステムティックレビューとは異なる点が多く注意が必要である。

現在の診断研究の質のレベルはさまざまで、システムティックレビューを行う際には、レビューに組み入れるかの判断のために質の評価を慎重に行うステップが重要である。診断のシステムティックレビューのための診断研究の質評価ツールとして、診断研究の論文執筆ガイダンスである Standards for Reporting of Diagnostic Accuracy (STARD)⁴⁾などが参考になるが、本稿では、コクランでも採用されている QUADAS-2 (a revised tool for the quality assessment of diagnostic accuracy studies 2)⁵⁾を紹介する。

2003年に、診断研究のバイアスリスクの測定ツールとして開発された QUADAS (Quality Assessment of Diagnostic Accuracy Studies)⁶⁾は、14の質問に「はい」、「いいえ」、「不明」で答える形で評価するツールで、エビデンスに基づく診断精度 (diagnostic test accuracy : DTA) 研究、診断のシステムティックレビューの発展に貢献した。コクランでは、11の質問にまとめられて採用されていたが⁷⁾、対象範囲の代表性、解釈不能な検査結果や脱落例の報告の評価が難しい、参照基準の検証が部分的でなかったかという項目と脱落についての項目はバイアスリスクとし

て重複する可能性がある、参照基準がフォローアップ検査となる状況では使いにくいなどの問題点があげられていた⁵⁾。2011年、これらを改良した QUADAS-2 が発表され、われわれは、著者らの許可を得てチェックリスト⁸⁾と背景文書⁹⁾を翻訳し、2014年4月 Minds ウェブサイトに公開した¹⁰⁾。2014年5月現在、University of BRISTOL の QUADAS サイト (<http://www.bris.ac.uk/quadas/quadas-2/>) にも公開準備中である。

DTA 研究に基づく診断のシステムティックレビューの特徴

一般に DTA 研究は、新たに関心のある検査 (インデックス検査 : index test) が、現在 gold standard として用いられている検査 (参照基準 : reference standard) と比較してどの程度よいかを検討する。参照基準となる検査は、いわゆる臨床検査だけでなく、臨床面接から得られる症状、身体所見を含めた臨床診断、画像診断、質問票など幅広い。単一の検査でも一連の検査でもよいが、現在の医療状況で最も妥当な検査である必要がある。

対象は、診断標的 (target condition) として、特定の疾患、または健康状態を明確に定義することが重要である。診断標的の検出に対するインデックス検査の有効性は、感度 (sensitivity)、特異度 (specificity) などを用いて評価する。感度・特異度は診断能の指標であるが、診断閾値によって変動し、疾患スペクトルの影響を受ける。陽性および陰性的中度 (positive and negative predictive values) や正診率は事前確率によって変動するため、スクリーニング・サーベイランス・診断などの用途で異なる。感度・特異度から計算される、陽性および陰性尤度比 (positive and negative likelihood ratios)、診断オッズ (diagnostics odds ratios)、ROC 曲線 (receiver operating characteristic curves) などが用いられることもある。

治療介入研究の個別研究デザインではランダム化

比較試験 (RCT) がエビデンスレベルの最も高い研究デザインとされるが、DTA 研究では、インデックス検査が実際の臨床の状況で使用されると想定される対象者と同じ対象者で行われる横断研究またはコホート研究が理想的である。対照群を疾患群とは別の臨床的状況で集めた、あるいは、疾患群と対照群をあいまいな基準で集めた症例対照デザインは、バイアスが大きいため診断能評価の研究デザインとして避けるべきである。

QUADAS-2 の概要

QUADAS-2 を用いて診断精度に関する一次研究のバイアスリスクを評価する手順と留意点を解説する。網羅的文献検索にて集められた一次研究の論文をスクリーニング後、システマティックレビューのデータ抽出プロセスとしての主要データ (例: 研究デザイン, 結果等) について抽出作業と並行して QUADAS-2 による研究の質の検討を行う。バイアスリスクについて問題が大きい研究はシステマティックレビューから除外することも検討する。本ツールは、フェーズ 1: レビュークエスチョンの記述 (表 1), フェーズ 2: 各レビューに合わせたシグナリングクエスチョンの調整と評価法の記述, フェーズ 3: フローダイアグラム作成, フェーズ 4: バイアスおよび適用可能性の判定の 4 つのフェーズからなる。フェーズ 4 は、表 2 に示すように、患者選択, インデックス検査, 参照基準, フローとタイミングの 4 つの重要なドメインで構成される。ドメイン 1, 2, 3 では、バイアスリスクの判定のポイントとなるシグナリングクエスチョンに、「はい」、「いいえ」、「不明」で回答することで最終的なバイアスリスクの評価を簡便に行うことができる仕組みとなっている。診断のシステマティックレビューのために QUADAS-2 を利用する場合は、2 名のレビュー著者が独立して評価を行い、それぞれ評点のポイントを明記し、相違点は議論することにより評価の透明性を確保することが重要となる¹¹⁾。

表 1 フェーズ 1: レビュークエスチョンの記述

フェーズ 1: レビュークエスチョンの 記述	患者 (セッティング, インデックス 検査の使用目的, 症状, 事前検査):
	インデックス検査:
	参照基準および診断標的:

本表の掲載にあたり、QUADAS-2 グループおよび Minds の許可を得た。

レビュークエスチョンの記述とバイアスの 評価の準備

DTA 研究のレビュークエスチョンは、いずれの診断法の精度が高いかの疑問文の形で表される。検査の精度は診断手順のどこで用いられるかによって変わるため、レビュー著者は、セッティング, インデックス検査の使用目的, 患者の症状, および事前検査について表 1 のようにフェーズ 1 で明確に定義する必要がある^{12,13)}。その成分は PICOTS からなり、以下に例をあげる。

- P : patients 結核が疑われる成人患者において
- I : intervention γ-インターフェロン検査 (インデックス検査) は
- C : comparison ツベルクリン検査と比較して (参照基準)
- O : outcome 診断の正確度が高い (有用である)。
- T : time (診断標的としての結核確定診断のための初期診断検査)
- S : study design (コホート研究)

フェーズ 2: 各レビューに合わせたシグナリングクエスチョンの調整と評価法の記述, フェーズ 3: フローダイアグラム作成の詳細については、Minds サイトに掲載されている和訳版背景文書¹⁰⁾を参照されたい。

シグナリングクエスチョンによる バイアスリスク評価と、適用可能性の判定

フェーズ 4 におけるバイアスの評価は、レビューごとに改訂されたシグナリングクエスチョンに答えることによって評価する。表 2 に、デフォルトの判定の根拠となるシグナリングクエスチョンによるバ

表 2 フェーズ 4: バイアスリスクと適用可能性

ドメイン (1~4)	1. 患者選択	2. インデックス検査	3. 参照基準	4. フローとタイミング
バイアスリスクに関する記述	患者選択方法: 組み入れられた患者 (事前検査, 症状, インデックス検査の使用目的およびセッティング):	インデックス検査と、それがどのように実施され解釈されたか [#] :	参照基準と、それがどのように実施され解釈されたか:	インデックス検査および/または参照基準を受けなかった患者、あるいは 2×2 分割表から除外された患者 (フローダイアグラム参照): インデックス検査から参照基準までの期間とその期間に行われた介入:
シグナリングクエスチョン (はい/いいえ/不明)	a) 連続した患者がランダムサンプルを組み入れたか。 b) 症例対照デザインではないか*。 c) その研究は不適切な除外を行っていないか*。	a) インデックス検査の結果の解釈は、参照基準の結果が分らない状態で行われたか。 b) 閾値が用いられた場合、閾値は事前に定義されたものであったか。	a) 参照基準は診断標的を正しく分類していると考えられるか。 b) 参照基準結果の解釈は、インデックス検査結果が分らない状態で行われたか。	a) インデックス検査と参照基準の間に適切な期間があったか。 b) すべての患者が参照基準を受けたか。 c) 患者は同一の参照基準を受けたか。 d) すべての患者が解析に含まれていたか。
A. バイアスリスク (低/高/不明)	患者選択はバイアスを生じた可能性があるか。	インデックス検査の実施または解釈はバイアスを生じる可能性があるか。	参照基準の実施や解釈はバイアスを生じる可能性があるか。	患者のフローはバイアスを生じた可能性があるか。
B. 適用可能性に関する懸念 (低/高/不明)	組み入れられた患者はレビュークエスチョンに合致していない懸念があるか。	インデックス検査の実施や解釈はレビュークエスチョンと異なる懸念があるか。	参照基準により定義された診断標的はレビュークエスチョンに合致しない懸念があるか。	

[#]複数のインデックス検査が用いられている場合、検査ごとに作成のこと。

*訳者注:「症例対照デザインでない」,「不適切な除外を行っていない」など、好ましい回答が「はい」となる。

本表の掲載にあたり、QUADAS-2 グループおよび Minds の許可を得た。

バイアスリスク評価と、適用可能性の判定方法のまとめを示すが、実際評価する場合は、Minds サイトに掲載されている和訳版 QUADAS-2 シート¹⁰⁾を使用していただきたい。各ドメインについて、すべてのシグナリングクエスチョンに対する回答が「はい」の場合はバイアスリスクが低いことを示し、いずれかのシグナリングクエスチョンが「いいえ」の場合、バイアスの可能性がある。「不明」は、報告されたデータが判定に不十分な場合にのみ使用し、最終的なバイアスリスクは「低」、「高」、「不明」に判定される。適用可能性の判定にはシグナリングクエスチョンは含まれず、研究がレビュークエスチョンに合致しない懸念について「低」、「高」、「不明」のいずれかに

評価する。

1. ドメイン 1: 患者選択

a 連続した患者がランダムサンプルを組み入れたか。

b 症例対照デザインではないか。

c その研究は不適切な除外を行っていないか。

疑われる疾患を有するすべての連続患者、またはランダムサンプルを組み入れる研究デザインの場合は、選択バイアスの可能性が低く、シグナリングクエスチョン a は「はい」と回答される。シグナリングクエスチョン b では、診断の確定した疾患群と対照群を比較する症例対照デザインでは、診断精度が誇張される^{1,14)}ため、避けられるべきで「はい」と答えられるのが理想的である。同様に、「診断が難し

い」患者を除外するなど不適切な除外が行われた研究は、診断精度の推定値が過度に高くなる可能性があるため、避けられるべきである（シグナリングクエスチョン c）。このように、好ましいシグナリングクエスチョンに「はい」と答えられる場合、バイアスリスクは「低」と判定される。

試験に含まれた患者が、診断標的の重症度、人口統計学的特性、鑑別診断や併存疾患の有無、研究のセッティング、事前検査について、レビュークエスチョンと異なる場合、適用可能性に関する懸念が「高」と判定される。

2. ドメイン 2：インデックス検査

a インデックス検査の結果の解釈は、参照基準の結果がわからない状態で行われたか。

b 閾値が用いられた場合、閾値は事前に定義されたものか。

介入研究における盲検化同様、インデックス検査結果の解釈は、参照基準結果を知っていることにより影響を受ける可能性がある¹⁾。インデックス検査がつねに参照基準の前に実施され解釈される場合、シグナリングクエスチョン a は「はい」と評価されることができる。感度および/または特異度を最適化する閾値を用いると、同一の閾値が用いられた別の患者集団においては診断能が劣る可能性があるため¹⁵⁾、閾値は研究計画時にあらかじめ定義されるべきである。これらの質問にいずれも「はい」と回答される場合、バイアスの可能性は「低」と判定される。適用可能性は、検査の技術、実施や解釈におけるばらつきが診断精度の推定値に影響を及ぼすため、インデックス検査の方法がレビュークエスチョンで定義されているものと異なる場合、懸念が「高」と判定される。

3. ドメイン 3：参照基準

a 参照基準は診断標的を正しく分類していると考えられるか。

b 参照基準結果の解釈は、インデックス検査結果が分からない状態で行われたか。

参照基準に関するシグナリングクエスチョンは、ドメイン 2：インデックス検査のものとよく似ている。参照基準の診断標的に対する感度・特異度は 100%と仮定されるため、参照基準とインデックス検査間の不一致はインデックス検査による誤分類に

よるものと解釈される^{16,17)}。バイアスの可能性は、参照基準の解釈がインデックス検査の結果を知った状態で行われると高くなる¹⁾。

参照基準はバイアスの影響が少なくても、参照基準が定義する診断標的がレビュークエスチョンで定義された診断標的とは異なる場合、適用可能性は「低」と判定される。たとえば、活動性肺結核に関する研究では、参照基準となる結核培養検査にはバイアスの影響は少ないと考えられるが、診断標的としての活動性肺結核の定義は研究によって異なる可能性がある。

4. ドメイン 4：フローとタイミング

a インデックス検査と参照基準の間に適切な期間があったか。

b すべての患者が参照基準を受けたか。

c 患者は同一の参照基準を受けたか。

d すべての患者が解析に含まれていたか。

インデックス検査と参照基準の結果は同時期に同一患者で収集されることが理想であるが、遅延がある場合や、治療がインデックス検査と参照基準の間に開始される場合には、状態の回復や悪化により誤分類が生ずる可能性がある（シグナリングクエスチョン a）。慢性疾患では数日の遅延は問題にならないことが多いが、急性の感染症では短い遅れが重要となる可能性があり、バイアスリスクが高いと評価される期間は状況によって異なることに注意が必要である。参照基準にフォローアップ検査が含まれるデザインのように、診断標的の有無を評価するには最小限のフォローアップ期間が必要となる場合もある。

シグナリングクエスチョン b, c では、対象群の全員が同一の参照基準による診断確認を受けない場合の検証バイアスを確認している。さらに、参照基準の検査の実施、あるいはどの参照基準を用いるかの判断にインデックス検査結果が影響を及ぼす場合、推定診断精度にバイアスが生ずる可能性がある¹⁴⁾。その研究に組み入れられたすべての患者が解析に含まれ、シグナリングクエスチョン d は、「はい」と回答されるのが理想的である¹⁸⁾。追跡不能例が研究に残っている患者と系統的に異なっていたり、組み入れ患者数が結果の 2×2 分割表に含まれる患者数と異なったりする場合、バイアスの可能性がある。

レビュー著者は、すべてあるいは特定のドメイン

表 3 結果の表形式での提示例

研 究	バイアスリスク				適用可能性の懸念		
	患者選択	インデックス 検査	参照基準	フローと タイミング	患者選択	インデックス 検査	参照基準
研究 1	☺	☺	☺	☺	☹	☺	☺
研究 2	☺	☺	☺	☺	☹	☺	☺
研究 3	☹	☹	☺	☺	☹	☺	☺
研究 4	☹	☹	☺	☺	☹	☺	☺
研究 5	☹	?	☺	☺	☹	☺	☺
研究 6	☹	?	☺	☺	☹	?	☺
研究 7	☹	?	☺	☺	☹	☺	☺
研究 8	☹	?	☺	☺	☹	?	☺
研究 9	☹	?	☺	☺	☹	☺	☺
研究 10	☹	?	☺	☹	☹	☺	☺
研究 11	☺	?	☺	☹	☺	☺	☺

☺ 低リスク ☹ 高リスク ? リスク不明

本表の掲載にあたり、QUADAS-2 グループおよび Minds の許可を得た。

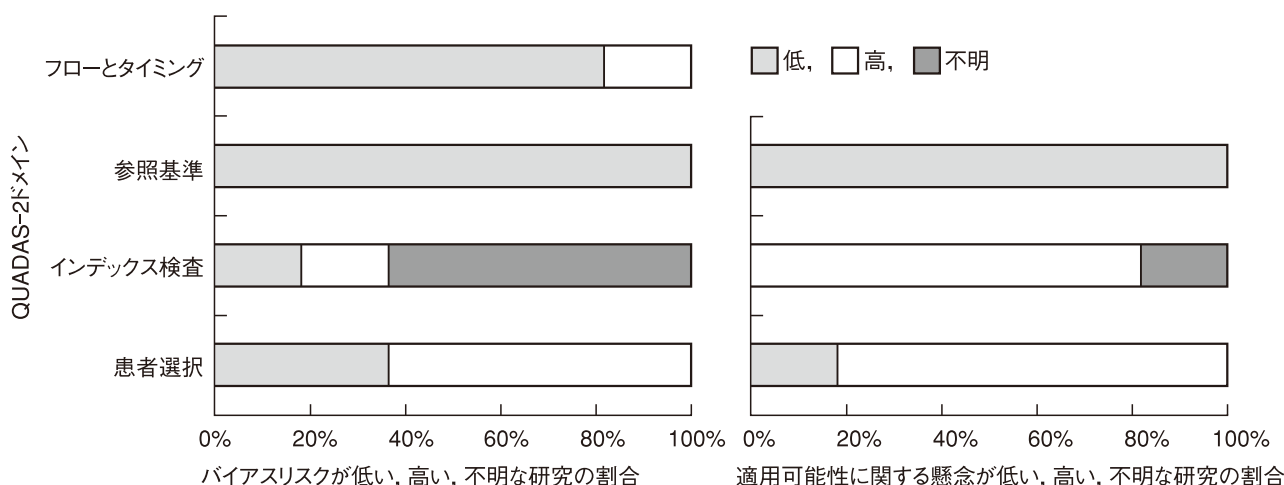


図 1 結果のグラフ表示例

本図の掲載にあたり、QUADAS-2 グループおよび Minds の許可を得た。

でバイアスリスクおよび/または適用可能性に関する懸念が低い研究だけを解析してもよいが、多くの場合、関連したすべてのエビデンスをレビューし、その後で異質性の可能性を検討するほうが望ましい^{19,20)}。レビューに含まれるすべての研究の QUADAS-2 評価結果の概要は、The Grading of Recommendations Assessment, Development and Evaluation (GRADE) が無償で公開している GRADEPro (GRADEprofiler)²¹⁾ などを用いて表形式 (表 3) やグラフ形式 (図 1) で表示する。QUADAS のウェブサイト (www.quadas.org) には、QUADAS-2 の英語版

のほか、トレーニングに関する情報、一連の追加シグナリングクエスチョン、各ドメインのより詳しい手引き、QUADAS-2 評価の例が掲載されている。データ抽出用の Microsoft AccessTM、結果をグラフ表示するための ExcelTM スプレッドシート、結果を要約するための WordTM 表のテンプレートなどのリソースもダウンロード可能となっているので、是非ご参照いただきたい。

考 察

診断は、臨床的枠組みにおいて治療の上流に位置することから、診断研究においても診断と治療を切り離さず患者アウトカムを考えるべきである²²⁾。しかしながら、診断に関する個別研究は診断精度を検討する論文が多く、GRADE ワーキンググループにおける論文の質の評価を含めた方法論の研究では、診断精度に関するものが多い^{23,24)}。Tatsioni A ら²²⁾による診断研究のレベルの診断確立の効果による6段階分類では、レベル1は安定した結果を目的とした技術の確立と最適化の研究、レベル2は感度・特異度の評価を目的とした診断能の研究で、現状ではほとんどこのレベルの研究である。治療選択への影響、患者アウトカムへの影響を効果指標としたさらに高いレベルの診断研究がコクランレビューなどで少しずつではあるが発表され始めている。

一方、診断過程における役割から診断検査分類についても、いずれに相当するかをレビュー研究の最初の段階に検討する必要がある。具体的には、評価対象となっているインデックス検査が現在の診断過程において、侵襲性の高い検査や高価な検査の実施を最小限にする「選別」、既存の検査よりも検査性能が劣る検査、負担や侵襲性やコストが高い検査を排除する「置換」、既存の検査による診断の精度をさらに向上させる「追加」の3種類である^{25,26)}。つまり、感度・特異度がより優れているというだけで、推奨を決められるわけではないことを認識する必要がある。また、複数の検査を組み合わせる施行した場合の、感度・特異度についてのデータは限られており、個別の検査の感度・特異度に基づく判断とは異なる可能性もある。特に診療ガイドライン策定で推奨作成をするときには、そのインデックス検査を用いて診断を確立することにより患者集団に対する益と不利益が総体的にどう見積もられるのか、関連する資源に見合ったものなのかを総合的に考量する方法論²⁷⁾が提言されているが、実際の臨床ではひとつの検査だけが用いられることはむしろまれであり、適用される状況は限定される。

しかしながら、本ツールで診断精度研究を評価し、適用可能性の高い研究から信頼しうる感度・特異度の値が明らかにされることは有用と考えられる。さ

らに、本ツールの和訳版を紹介することで、QUADAS-2 でバイアスリスクの可能性、適用可能性の懸念の両方が低いと評価される診断精度研究が増え、将来的には、治療選択、患者アウトカムへの影響を検討する診断研究へと発展していくことを期待している。

【謝 辞】

QUADAS-2 の和訳を許可してくださった、Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM ; QUADAS-2 Group に感謝します。また、翻訳権の取得などにご協力いただいた、公益財団法人日本医療機能評価機構 EBM 医療情報部の奥村晃子氏、畠山洋輔氏他、事務局の皆さまの支援に感謝します。

文 献

- 1) Whiting PF, 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.
- 2) Leeflang MG, Deeks JJ, Takwoingi Y, et al. Cochrane diagnostic test accuracy reviews. *Syst Rev* 2013 ; 2 : 82.
- 3) 福井次矢ほか監修, 山口直人ほか編集. 診療ガイドライン作成の手引き 2014. 医学書院, 2014.
- 4) Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy : The STARD Initiative. *Ann Intern Med* 2003 ; 138 : 40-4.
- 5) Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2 Group. QUADAS-2 : a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011 ; 155 : 529-36.
- 6) Whiting PF, 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 Medical Research Methodology* 2003 ; 3 : 25.
- 7) Reitsma JB, Rutjes AWS, Whiting P, et al. Chapter 9 : Assessing methodological quality. In : Deeks JJ, Bossuyt PM, Gatsonis C (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. The Cochrane Collaboration, 2009.
- 8) <http://www.bris.ac.uk/quadas/resources/quadas2.pdf> (2014.5.15 参照)
- 9) <http://www.bris.ac.uk/quadas/resources/background-doc.pdf> (2014.5.15 参照)
- 10) http://minds4.jcqh.or.jp/minds/guideline/pdf/QUADAS-2_JPN.pdf (2014.5.15 参照)

- 11) Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. *BMJ* 2011 ; 343 : d5928.
- 12) Leeflang MM, Deeks JJ, Gatsonis C, et al. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008 ; 149 : 889-97.
- 13) Bossuyt PM, Leeflang MM. Chapter 6 : Developing Criteria for Including Studies. In : *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. The Cochrane Collaboration, 2009.
- 14) Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999 ; 282 : 1061-6.
- 15) Leeflang MM, Moons KG, Reitsma JB, et al. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values : mechanisms, magnitude, and solutions. *Clinical Chemistry* 2008 ; 54 : 729-37.
- 16) Biesheuvel C, Irwig L, Bossuyt P. Observed differences in diagnostic test accuracy between patient subgroups : is it real or due to reference standard misclassification. *Clin Chem* 2007 ; 53 : 1725-9.
- 17) van Rijkom HM, Verdonchot EH. Factors involved in validity measurements of diagnostic tests for approximal caries—a meta-analysis. *Caries Research* 1995 ; 29 : 364-70.
- 18) Macaskill P, Gatsonis C, Deeks JJ, et al. Chapter 10 : Analysing and presenting results. In : Deeks JJ, Bossuyt PM, Gatsonis C (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. The Cochrane Collaboration, 2010.
- 19) Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Medical Research Methodology* 2006 ; 6 : 9.
- 20) Whiting P, Harbord R, Main C, et al. Accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis : systematic review. *BMJ* 2006 ; 332 : 875-84.
- 21) <http://tech.cochrane.org/revman/gradepro> (2014.5.15 参照)
- 22) Tatsioni A, Zarin DA, Aronson N, et al. Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med* 2005 ; 142 (12 Pt 2) : 1048-55.
- 23) Gopalakrishna G, Mustafa RA, Davenport C, et al. Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) to diagnostic tests was challenging but doable. *J Clin Epidemiol* 2014 ; 67 : 760-8.
- 24) Schünemann HJ, Oxman AD, Brozek J, et al. GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008 ; 336 : 1106-10.
- 25) Bossuyt PM, Irwig L, Craig J, et al. Comparative accuracy : assessing new tests against existing diagnostic pathways. *BMJ* 2006 ; 332 : 1089-92.
- 26) 相原守夫 訳. GRADE の適用 : 診療ガイドラインにおいて診断検査に関する根拠に基づく推奨を作成する. *Clin Eval* 2012 ; 40 : 199-212.
- 27) Hsu J, Broz JL, Terracciano L, et al. Application of GRADE : Making evidencebased recommendations about diagnostic tests in clinical practice guidelines. *Implementation Science* 2011 ; 6 : 62.

* * *