

QUADAS-2 背景文書

QUADAS-2

QUADAS-2 は診断精度に関する一次研究の質を評価するためにデザインされている。システマティックレビューのデータ抽出プロセスに代わるようデザインされたものではなく、レビューで使用するための主要データ（例：研究デザイン、結果等）を抽出に加えて QUADAS-2 を用いるべきである。QUADAS-2 は 4 つの重要なドメイン、すなわち患者選択、インデックス検査、参照基準、並びに研究を通しての患者フローおよびインデックス検査と参照基準のタイミング（「フローとタイミング」）から構成されている（表 1）。本ツールは 1) レビュークエスションの記述, 2) レビュー用手引きの作成, 3) 公表されている一次研究のフローダイアグラムのレビュー, あるいは論文中に報告がない場合はフローダイアグラムの作成, 4) バイアスおよび適用可能性の判定の 4 つのフェーズを経て完了する。各ドメインではバイアスリスクが評価され、最初の 3 つのドメインは適用可能性についても評価される。バイアスリスクの判定に役立つシグナリングクエスション（signalling questions, シグナルとなるクエスション）が含まれている。これら研究デザインの重要な面はバイアスの可能性に関連しており、レビューアがバイアスリスクを判定する手助けとなるものである。

フェーズ 1：レビュークエスション

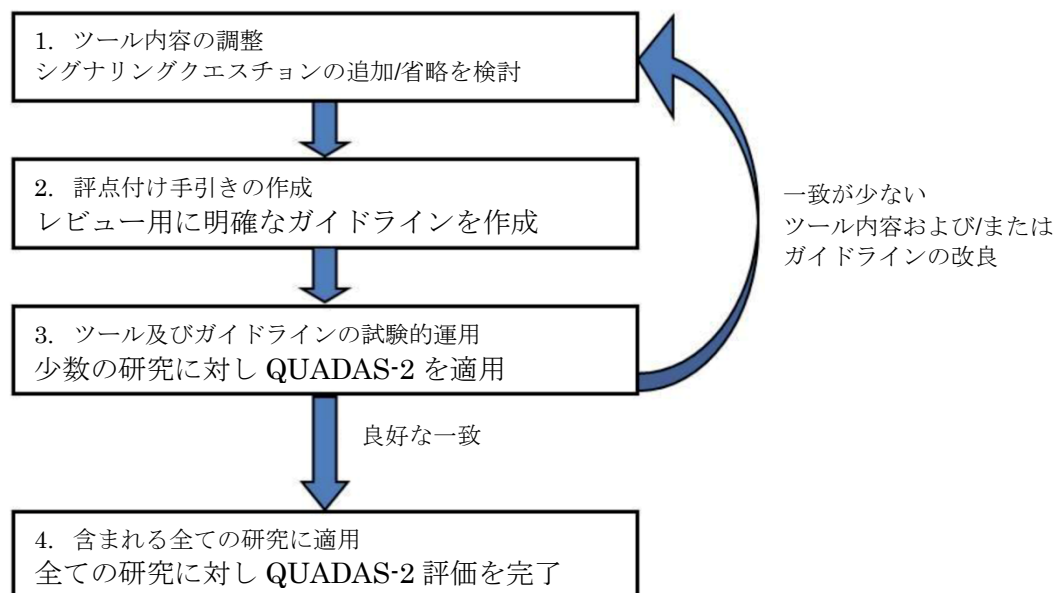
レビューアは最初に、患者、インデックス検査、並びに参照基準と診断標的に関するシステマティックレビューのクエスションを記述することが求められる。検査の精度は診断手順のどこで用いられるかによって変わるため、レビューアは、セッティング、インデックス検査の使用目的、患者の症状、および事前検査について記述するよう求められる（1;2）。

フェーズ 2：レビュー用調整（図 1）

レビューごとに QUADAS-2 を調整することは重要である。シグナリングクエスションを QUADAS-2 に追加あるいは省略したり、各シグナリングクエスションの評価方法ならびにバイアスリスク判定のためのこの情報の使い方に関するレビュー用手引きを作成する。最初のステップは、レビューに用いないシグナリングクエスションがないか、あるいは基本的なシグナルクエスションにより適切にカバーされないレビューに関する特異的な問題がないかどうかの検討である。例えば、客観的なインデックス検査のレビューでは、検査結果を解釈する者に対し参照基準が盲検化されていたかどうかに関するシグナリングクエスションは省略してもよいかもしれない。レビューアは、シグナリングクエスションを追加し過ぎてこのツールを複雑にすることは避けるべきである。ツール内容の合意が得られたら、レビュー用にあわせた評点付けの手引きを作成すべきである。ツールは少なくとも 2 人が独立して試験運用を行うべきである。良好な一致が得られたら、全ての対象研究の評

価にツールを用いることができる。一致が悪い場合、更なる改良が必要となる可能性がある。

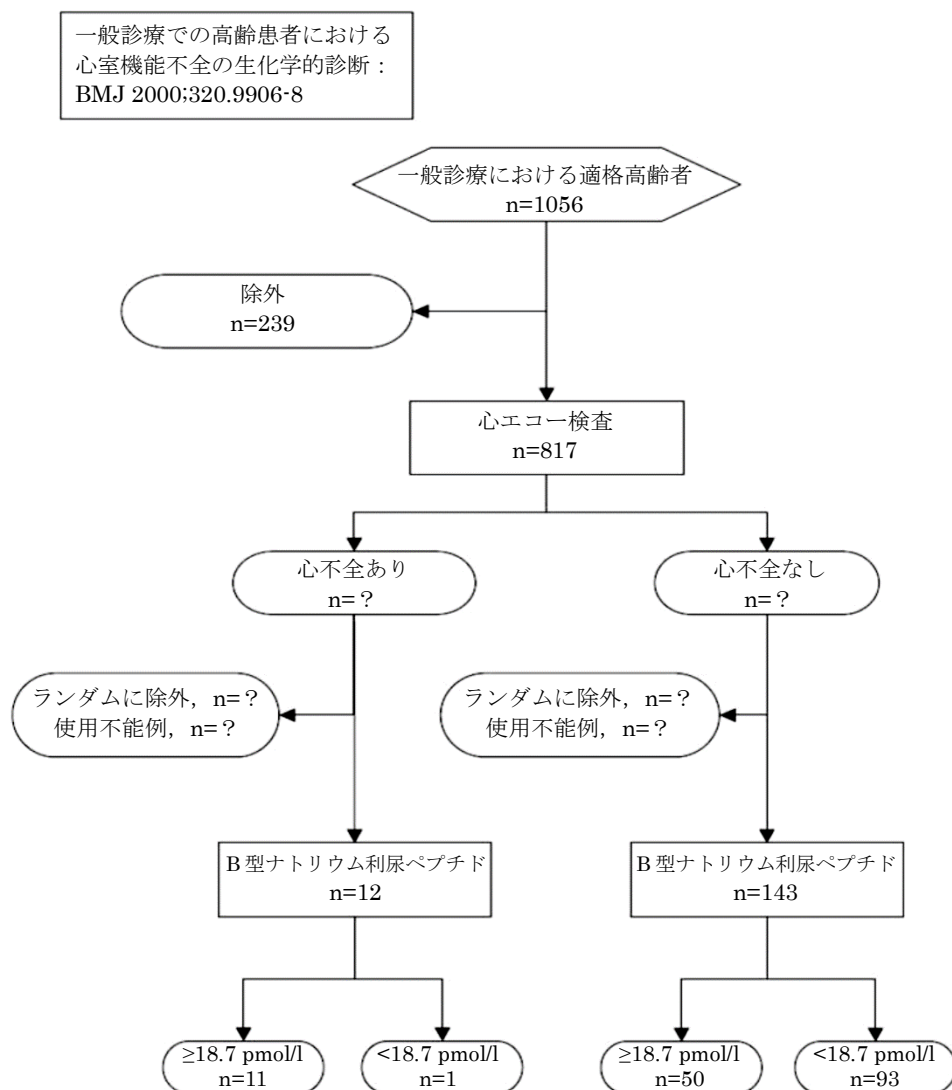
図 1 : 各システマティックレビュー用に QUADAS-2 を調整するプロセス



フェーズ 3 : フローダイアグラム

次の段階では、一次研究に関し発表されたフローダイアグラムのレビューを行うが、発表されたフローダイアグラムがない、もしくは適切でない場合はレビュー著者がフローダイアグラムを作成する。フローダイアグラムはバイアスリスクの判定を容易にするものであり、患者の募集方法（例：診断標的が疑われる特定の症状を有する連続した患者、または連続した症例及び対照に基づく）、検査実施順序、ならびにインデックス検査および参照基準を受けた患者数についての情報を提供すべきである。この段階は QUADAS-2 評価の一部として報告する必要があるため、手書きのチャートで十分である。図 2 に心不全診断のための B 型ナトリウム利尿ペプチドの一次研究に基づく例を示す。

図 2 : 心不全診断のための BNP の診断コホート研究に基づくフローダイアグラム



フェーズ 4：バイアスおよび適用可能性の判定

バイアスリスク

各ドメインの最初の部分はバイアスに関するもので、1) バイアスリスク判定の裏付けに用いる情報、2) シグナリングクエスチョン、および 3) バイアスリスクの判定の 3 つのセクションから構成される。判定に至る（「判定を裏付ける」）のに用いた情報を記録することにより、評定に透明性をもたせ、独立して評価を行うレビュー著者間での議論を促すことを目的としている(3)。追加シグナリングクエスチョンは判定に役立てるために含まれる。クエスチョンは「はい」、「いいえ」または「不明」で回答され、「はい」はバイアスリスクが低いことを示す。

バイアスリスクは「低」、「高」、「不明」に判定される。あるドメインに対する全てのシグナリングクエスチョンが「はい」と回答された場合、バイアスリスクは「低」と判定できる。いずれかのシグナリングクエスチョンが「いいえ」と回答された場合、バイアスの可能性を警告している。次に、レビュー著者はバイアスリスクを判定するためにフェーズ 2 で作成した手引きを使用する必要がある。「不明」の分類は、報告されたデータが判定に不十分な場合にのみ使用すべきである。

適用可能性

適用可能性のセクションはバイアスのセクションとよく似た構成となっているが、シグナリングクエスチョンは含まない。レビュー著者は、適用可能性の判定の基になる情報を記録し、次に、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判定するよう求められる。適用可能性に関する懸念は「低」、「高」、「不明」のいずれかに評価される。適用可能性の判定は、レビュークエスチョンが記録された最初のフェーズを参照すべきである。ここでも「不明」は報告されたデータが不十分な場合にのみ使用すべきである。

以下のセクションで、各ドメインでのシグナリングクエスチョンとバイアスリスク/適用可能性に関する懸念に関するクエスチョンについて簡単に説明する。

ドメイン 1：患者選択

バイアスリスク：患者の選択はバイアスを生じた可能性があるか。

シグナリングクエスチョン 1：連続した患者サンプルかランダムサンプルを組み入れたか。

シグナリングクエスチョン 2：症例対照デザインでないか。

シグナリングクエスチョン 3：その研究は不適切な除外を行っていないか。

理想としては、疑われる疾患を有する適格患者の全ての連続サンプルまたはランダムサンプルを研究に組み入れるべきである。そうでないとバイアスが生じる可能性がある。例えば「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では、診断精度の推定値が過度に楽観的となる可能性がある。関節リウマチ診断のための抗 CCP 抗体のレビューにおいて、一部の研究では連続した確定診断患者を組み入れていた。これらの研究では、疑われる疾患を有するが診断が確定していない患者、つまり「診断の難しい」患者を含めた試験よりも抗 CCP 検査の感度が高かった(4)。同様に、疾患を有することが既知の患者と、当該状態のない対照群を組み入れた研究では、診断精度が誇張される恐れがある(5;6)。診断標的を示す「明らかな徴候」がある患者は、診断が容易であると思われるが、こうした患者を除外することは診断精度の過小評価につながる恐れがある。

適用可能性：組み入れられた患者とセッティングはレビュークエスチョンに合致しない懸念があるか。

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、および以前の試験プロトコールに関し、レビュークエスチョンが対象とするものと比較して異なる場合、適用可能性に関する懸念がある可能性がある。例えば、大きい腫瘍は小さい腫瘍よりも画像検査で見つけやすく、大きな心筋梗塞は小さな梗塞よりも心筋酵素値がより高くなり、これにより検出が容易になるため、感度の推定値が上昇する(7)。

ドメイン 2：インデックス検査

バイアスリスク：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。

シグナリングクエスチョン 1：インデックス検査の結果の解釈は、参照基準の結果が分からない状態で行われたか。

この項目は介入研究における「盲検化」とよく似ている。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(6)。バイアスの可能性は、インデックス検査解釈の客観性と検査の順序に関連している。インデックス検査が常に参照基準検査の前に実施され解釈されるならば、この項目は「はい」と評価することができる。

シグナリングクエスチョン 2：閾値が用いられていた場合、それは事前に定義されたか。

感度および/または特異度を最適化する検査閾値を選択することは、診断能の楽観的な推定値につながる可能性があり、同一の閾値が用いられた独立した患者標本において診断能が劣る可能性が高い(8)。

適用可能性：インデックス検査、あるいはそ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。

検査の技術、実施や解釈におけるばらつきは、診断精度の推定値に影響を及ぼす可能性がある。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、適用可能性に懸念がある可能性がある。例えば、より高い超音波振動子周波数により、腹部外傷患者の評価の感度が向上することが示されている(9)。

ドメイン 3：参照基準

バイアスリスク：参照基準、あるいはその実施や解釈はバイアスを生じる可能性があるか。

シグナリングクエスチョン 1：参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。

検査精度の推定値は参照基準の感度・特異度が 100%であるとの仮定に基づくものであり、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不正確な分類に起因すると仮定される(10;11)。

シグナリングクエスチョン 2：参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。

この項目はインデックス検査の解釈に関連したシグナリングクエスチョンとよく似ている。バイアスの可能性は、参照基準解釈に対する事前の知識の潜在的影響に関連する(6)。

適用可能性：参照基準により定義された診断標的はクエスチョンに合致しない懸念があるか。

参照基準はバイアスの影響を受けないかもしれないが、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かもしれない。例えば、尿路感染の定義の際、参照基準は一般に検体培養に基づくが、結果が陽性と判断される基準となる閾値は異なる可能性がある(12)。

ドメイン 4：フローとタイミング

バイアスリスク：患者フローはバイアスを生じた可能性があるか。

シグナリングクエスチョン 1：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。理想的には、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。遅延がある場合、あるいは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、状態の回復や悪化により誤分類が生ずる可能性がある。高いバイアスリスクの原因となる期間の長さは状況によって異なる。慢性的な病態では数日の遅延は問題にならないかもしれないが、急性の感染症では短い遅れが重要となる可能性がある。これに対し、参照基準にフォローアップ検査が含まれる場合、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。例えば、多発性硬化症の早期診断のための磁気共鳴画像検査の評価では、今後多発性硬化症の診断基準を満たすとされる全ての患者が実際にそうなるだろうと確信するには最低でも約 10 年のフォローアップ期間を要する(13)。

シグナリングクエスチョン 2：全ての患者が同一の参照基準検査を受けたか。(訳者注：シートでは 2 つのシグナルクエスチョンに分かれている)

検証バイアスは、対象群の全員が同一の参照基準による診断確認を受けたわけではない場合に発生する。参照基準の検査を実施するかしないか、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(5;14)。例えば、肺塞栓症診断のための D-ダイマー検査の精度を評価する研究で、検査結果が陽性であった患者には換気血流スキャン(参照基準 1)を実施し、検査結果陰性患者には肺塞栓症の有無を判断するのに臨床経過観察を用いた場合(参照基準 2)、肺塞栓症を有するがインデックス検査で陰性だった一部の患者が臨床経過観察で見過ごされ、このため肺塞栓症なしと分類されてしまうように、偽陰性の一部を真陰性と誤って分類してしまう可能性がある。こうした誤分類は感度と特異度の過大評価につながる。

シグナリングクエスチョン 3：全ての患者が解析に含まれていたか。

研究に組み入れられた全ての患者が解析に含まれるべきである(15)。追跡不能例が研究に残っている患者と系統的に異なるなどして、組み入れ患者数が、結果の 2x2 分割表に含まれる患者数と異なる場合、バイアスの可能性がある。

診断精度レビューへの QUADAS-2 評価の組み込み

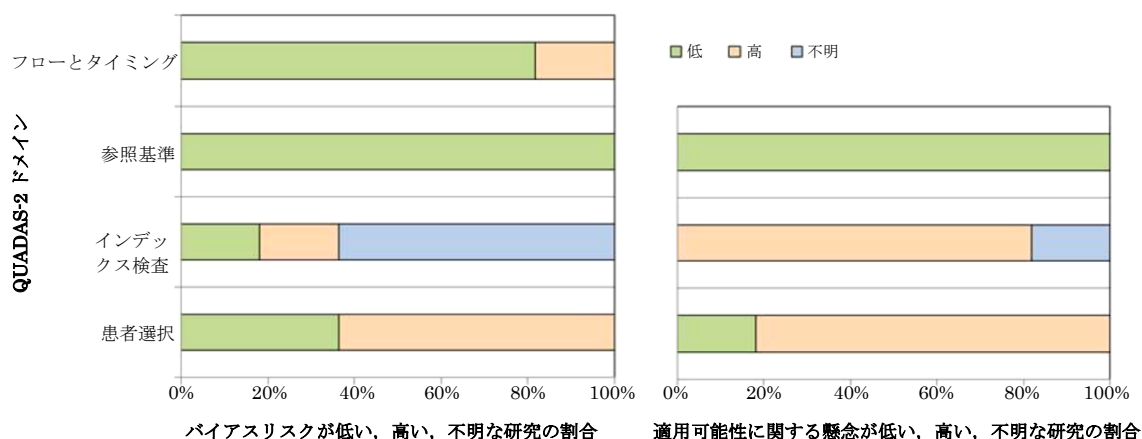
「質スコア」には既知の問題があることから、QUADAS-2 を「質スコア」の統合値の作成に使うべきでないということを強調したい(16;17)。ある研究がバイアスあるいは適用可能性に関する全てのドメインで「低」と判定される場合、この研究の全体の判定は「バイアスリスクが低い」あるいは「適用可能性に関する懸念が低い」とするのが適切である。ある研究が 1 つ以上のドメインで「高」または「不明」と判定された場合、「バイアスリスクがある」あるいは「適用可能性に関する懸念がある」と判定される可能性がある。

最低でも、含まれる全ての研究についての QUADAS-2 評価結果の概要をレビューで提示すべきである。これには各ドメインのバイアスリスク/適用可能性に関する懸念が低い、高い、または不明であることが判明した研究数のまとめを含むこともある。研究が特定のシグナリングクエスチョンに関し一貫して良好な評価あるいは悪い評価を得ている場合、レビューはこれを強調することもできる。表形式(表)とグラフ形式(図 3)の表示は、QUADAS-2 評価を要約するのに役立つ。

表：QUADAS-2 結果の表形式での提示の見本

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				

図 3 : QUADAS-2 結果のグラフ表示の見本



レビュー著者は、全てあるいは特定のドメイン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含むよう主要解析を限定することもできる。同様の基準に基づきレビューに含めるものを限定することが適切な場合もあるが、多くの場合、関連した全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の方が望ましい (13:18)。サブグループ解析や感度分析は、インデックス検査の精度の推定値が、全てあるいは特定のドメインに関し高い、低い、または不明と評価された試験間でどのようにばらついているかを調べることにより行うことができる。ドメインやシグナリングクエスチョンは、推定精度との関連を調べるためにメタ回帰分析の項目として含めることが可能である。

ウェブサイト

QUADAS のウェブサイト (www.quadas.org) には、QUADAS-2、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、完了した QUADAS-2 評価の例、データ抽出用の Microsoft Access™ を含むダウンロード可能なリソース、結果をグラフ表示するための Excel™ スプレッドシート、結果を要約するための Word™ 表のテンプレートが掲載されている。

参考文献

- (1) Bossuyt PM, Leflang MMG. Chapter 6: Developing Criteria for Including Studies. 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.
- (2) Leflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic

test accuracy. *Ann Intern Med* 2008; 149(12):889-897.

(3) Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD et al. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. *BMJ*. In press 2011.

(4) Whiting PF, Smidt N, Sterne JA, Harbord R, Burton A, Burke M et al. Systematic review: accuracy of anti-citrullinated Peptide antibodies for diagnosing rheumatoid arthritis. *Ann Intern Med* 2010; 152(7):456-464.

(5) Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282(11):1061-1066.

(6) Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004; 140(3):189-202.

(7) Reitsma J, Rutjes A, WP, Vlassov V, Leeflang M, Deeks J. 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) Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clinical Chemistry* 2008; 54(4):729-737.

(9) Stengel D, Bauwens K, Rademacher G, Mutze S, Ekkernkamp A. Association between compliance with methodological standards of diagnostic research and reported test accuracy: meta-analysis of focused assessment of US for trauma. *Radiology* 2005; 236(1):102-111.

(10) 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(10):1725-1729.

(11) van Rijkom HM, Verdonschot EH. Factors involved in validity measurements of diagnostic tests for approximal caries--a meta-analysis. *Caries Research* 1995; 29(5):364-70.

(12) Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J et al. Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model. *Health Technol Assess* 2006; 10(36):iii-xiii, 1.

(13) Whiting P, Harbord R, Main C, Deeks JJ, Filippini G, Egger M et al. Accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis: systematic review. *BMJ* 2006; 332(7546):875-884.

(14) Rutjes A, Reitsma J, Di NM, Smidt N, Zwinderman A, Van RJ et al. Bias in estimates of diagnostic accuracy due to shortcomings in design and conduct: empirical evidence [abstract]. XI Cochrane Colloquium: Evidence, Health Care and Culture; 2003 Oct 26 31; Barcelona, Spain 2003;45.

(15) Macaskill P, Gatsonis C, Deeks JJ, Harbord R, Takwoingi Y. 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.

(16) Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282(11):1054-1060.

(17) Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol* 2005; 5:19.

(18) Whiting PF, Westwood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Medical Research Methodology* 2006; 6:9.

出典: Whiting PF et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011 Oct 18;155(8):529-36.

この文書は <http://www.bris.ac.uk/quadas/resources/background-doc.pdf> を許可を得て和訳したものです (2014.1.18)

翻訳を許可していただいた Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeftang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group に感謝します。

監訳 小島原典子
森實 敏夫