

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*1

Application of GRADE: Making evidence-based recommendations
about diagnostic tests in clinical practice guidelines

Jonathan Hsu¹, Jan L Brozek^{1,2}, Luigi Terracciano³, Julia Kreis⁴, Enrico Compalati⁵, Airton Tetelbom Stein⁶,
Alessandro Fiocchi³ and Holger J Schünemann^{1,2,*2}

訳 相原 守夫¹⁾

抄録

背景：適切な医療には、正確な診断が欠かせない。しかし、医療現場にはさまざまな検査が存在し、資源の制約があることから、臨床医は診断検査の実行に際し、指針を必要とする。診断検査の使用に関する推奨を含め、ガイドラインに示される推奨は実行しなければならないものだと感じている医師が多い。しかし、診断検査に対するガイドラインパネルの理解や推奨作成のための方法は十分に吟味されているとはいえない状況である。そのため、われわれは診断検査に関する実行可能な推奨を作成するためにガイドライン作成者や利用者が考慮する必要のある要因について評価した。

方法：われわれはプロセスを批判的分析しながら、世界アレルギー機構 (World Allergy Organization) による牛乳アレルギー診断のための診療ガイドライン作成にあたってGRADE (Grading of Recommendations Applicability, Development and Evaluation) アプローチを使用したケーススタディの結果を示す。

結果：ガイドラインパネルが診断検査に関して十分な情報に基づいた推奨を作成できることを確実にするためには、疑問の定式化、診断検査における患者にとって重要なアウトカムの定義、エビデンスの要約を含むグループプロセスの重要性に対する認識を高める必要があるように見える。ガイドラインパネルによる作業を円滑化し、実行可能な推奨を作成するためには、検査前確率や治療閾値のようなエビデンスに基づく医療から導かれる診断の概念について明確に検討する必要がある。

考察：このケーススタディは、診断検査に関する推奨の実行の円滑化や推奨に対する信頼強化のためには診療ガイドラインに何が求められるのかについて、ガイドライン作成者や臨床医に有用な指針を提供する。推奨に影響するエビデンスや要因についての説明に透明性を求めるGRADEアプローチのような構造化された枠組みを適用することで、診断検査に関する推奨の作成・解釈・実行に必要なプロセスや決定因子を容易に整理できる。

1) 相原内科医院 (Aihara Clinic)

*1 本記事は、以下の記事の翻訳であり、著者である Holger J Schünemann の許諾を得て、BioMed Central の Editorial Office にも確認の上、掲載する。Jonathan Hsu, Jan L Brozek, Luigi Terracciano, Julia Kreis, Enrico Compalati, Airton Tetelbom Stein, Alessandro Fiocchi, Holger J Schünemann. Application of GRADE: Making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implementation Science* [Internet]. 2011 Jun 10; 6:62. Available from: <http://www.implementationscience.com/content/6/1/62>
著作権は原著者に帰属し、Creative Commons による原本の出版と他者による翻訳の承認によっても、原著者の著作権が変化することはない。本記事は、原著者の承諾を得て、本誌ホームページに公表される。 http://homepage3.nifty.com/cont/40_1/p199-212.pdf

*2 連絡先: schuneh@mcmaster.ca

¹ Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
著者情報の全リストはこの論文の最後に記載されている。

背景

実行可能な推奨を提供する質の高い診療ガイドラインは、医療における患者アウトカムを改善する理想的ツールである。したがって、ガイドラインは実行に役立つポイントが併記された透明性の高い明示的な推奨を提供しなければならない。たとえば、診断検査に関する推奨では、各検査によってもたらされる帰結が考慮されるべきである。つまり、正確な診断は有効な治療の必要条件ではあるが、正確な診断のみを切り離して考えるべきではない。診断を確立しても、ある患者または患者集団がその診断により利益を享受するかどうかの情報を提供することにはならない。疾患関連アウトカム（例：死亡率の減少）、検査がもたらす精神的帰結および資源活用アウトカムを含め、患者にとって重要なアウトカムについて、診断がもたらす利益が測定されるべきである。診断検査に関する推奨では、これらのアウトカムが総合的に正味の利益を達成するのか、またその正味の利益が関連する資源に見合ったものかが考慮されるべきである。しかし、診断検査の研究で患者にとって重要なアウトカムが注目されることは稀である¹⁾。さらに、診断精度 [diagnostic accuracy] に関するデータを集約するために使用される統計的手法の概念は複雑で結果の解釈が難しいことから、診断検査に関するエビデンスを統合するのは特に困難である²⁾。ガイドライン作成者は、これらの困難に直面しながら診断検査の使用に関する推奨を作成するが、その際に患者にとって重要なアウトカムにおける診断検査の適用の帰結が考慮されていないことが多いのではないかと考えられる³⁾。その原因の1つとして、診断検査や診断戦略に関する推奨作成のための適切な指針がないことがあげられる。

推奨を作成するにあたって関連する全ての局面を確認しなかった場合は深刻な事態になりかねない。たとえば、ある診断検査の使用に関する推奨を作成するには、患者にとって重要なアウトカム

に焦点をあてた複数の診断戦略と治療戦略を直接比較したエビデンスを必要とするか、あるいは有病率、診断検査精度、介入の有効性、患者の予後についての仮説を設定しなければならない。GRADE Working Groupによる過去の取り組みでは、診断検査に関する推奨の作成に関わる原則や課題の洗い出しを行ったが、診断検査に関する推奨作成においてGRADEあるいは他の明示的で透明性の高い枠組みを適用している例はほとんどない⁴⁾。

診断検査に関する推奨の作成にあたって明確な枠組みが適用されていない状況で、医療提供者はこれらの限界を認識することのないまま、診断検査の使用に関する推奨を含め、各推奨の根底にある仮説が明確にされていない状態でガイドラインの推奨を実行しているのではないかと考えられる。このように、ガイドライン提供組織は、推奨の実行を円滑にするために推奨作成者が診断検査実施の帰結をよく理解できるようにする方法を必要としている。これらの方法としてあげられるのが、どのようにガイドライン作成者や医療提供者にエビデンスを提示し、エビデンスから推奨を導き、実行を円滑にする推奨を定式化するかを示した指針である。

そのため、われわれは、ガイドラインパネルとの連携による診断検査に関する推奨作成に関わる課題と解決策について説明する。われわれのケーススタディは、乳児の1.9%～4.9%に発症する牛乳アレルギー（CMA：cow's milk allergy）という臨床領域における世界アレルギー機構（WAO：World Allergy Organization）のガイドライン⁵⁾へのGRADEアプローチの使用に基づくものである^{6～11)}。本稿では、ガイドラインパネル向けに患者にとって重要なアウトカムを明示すること、エビデンスを要約することについて、包括的かつ構造的に説明する。さらに、本ガイドラインパネルが透明性の高いエビデンスに基づく推奨を確実にするために使用したグループプロセスや合意プロセスについて説明する。

このアプローチは、診断検査に関する推奨を作

成するにあたって、すでに50以上の組織に採用されている方法¹²⁾であるGRADEアプローチ、またはその他の同様のアプローチを実行したいパネルにとっての指針となるだろう。また、ガイドライン利用者による解釈を容易にし推奨への確信を強めるには、診断に関する推奨に何が求められるのかについての認識を向上させるものと考えられる。

方法

一般的方法

われわれは、書面の記録や議事録、さらにはWAO CMAガイドラインの作成に際して使用されたプロセスの批判的分析に基づくケーススタディを実施した。本稿の執筆貢献者のうち3人(HJS, JLB, JK)は、GRADE Working Groupメンバーであり、GRADEアプローチの開発にさまざまな程度で寄与してきている。

パネルの選出と構成

本ガイドラインのパネルには、アレルギー専門医、小児科医、胃腸科専門医、皮膚科専門医、家庭医、疫学者、ガイドライン作成者、食品化学者、患者組織からの代表者など、世界各国の22名のメンバーが含まれていた。エビデンスの統合と診療に関わる推奨の作成は、GRADEアプローチの適用に豊富な実績を持つ方法論学者2人(HJSとJLB)の指導のもとで行われた。

利益相反

世界保健機関(World Health Organization)¹³⁾と米国胸部疾患学会(American Thoracic Society)¹⁴⁾の推奨に従い、会議に先立ち、パネルメンバーには利益相反申告書を記入してもらった。パネルは、特定の推奨に関する議論や投票については、メンバー各自が必要に応じて辞退すること、または議長により除外されることに合意した。

グループプロセス

ガイドライン作成時は、コアグループがエビデンス統合を進めるために定期的に会合した。まずパネル全体からの情報提供が必要な場合は、電子メールや電話会議を使って情報提供を求めることで、経済的かつ合理的プロセスを実現した。2009年12月には、パネルメンバー全員による対面会議を開き、系統的に収集されたエビデンスをレビューし、推奨について議論し、推奨の表現や強さについて合意した。さらなる明確化や議論を必要とする推奨については、フォローアップ電話会議で最終決定を行った。

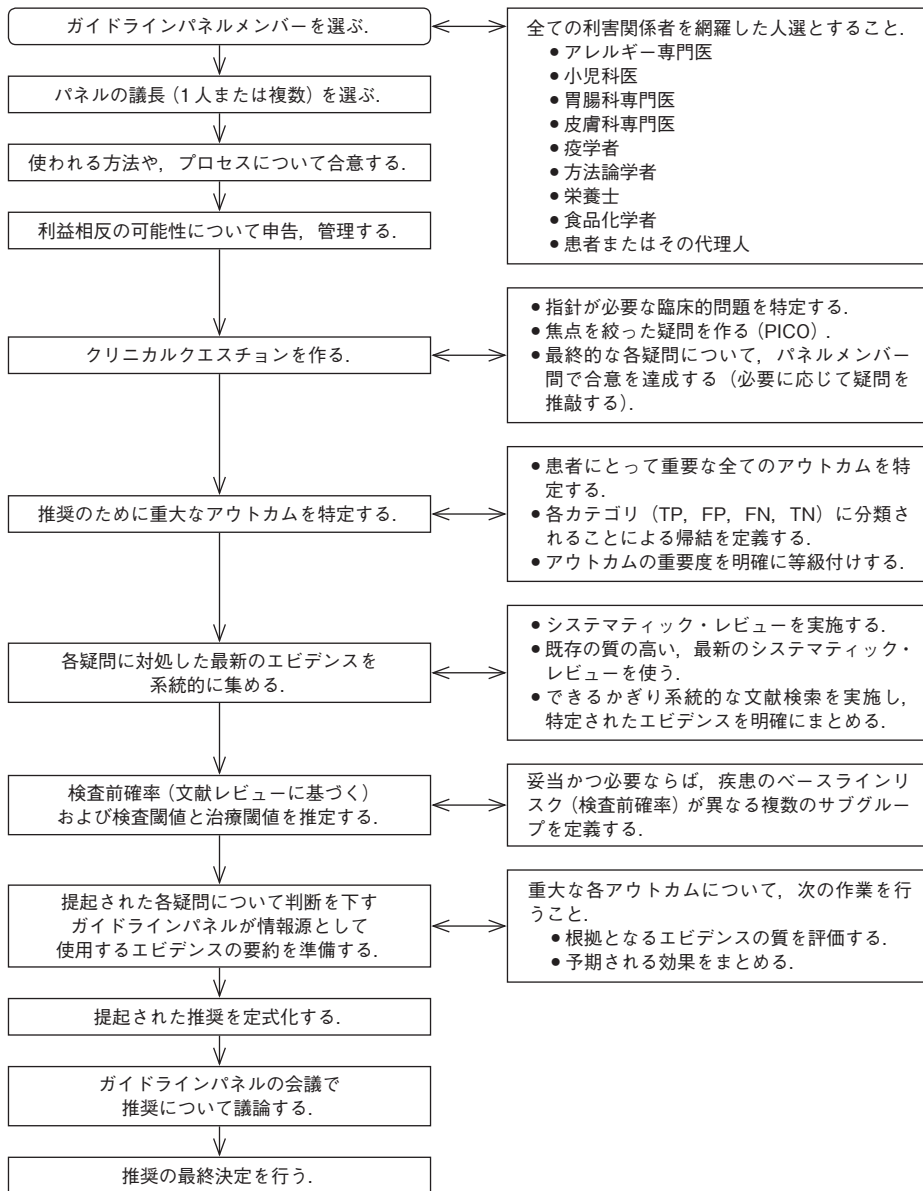
通常、グループプロセスでは、会議前のプロセスでは修正デルファイ法を採用し(電子メールによる質問で各個人の判断を募り、正式かつ明確な方法により回答やフィードバックを集計)、会議中のプロセスでは構造化議論法(structured discussion method)を採用した¹⁵⁾。この後者の方法は、特に診断に関する推奨の作成における複雑な方法論的問題に対する基本的理解を深め、推奨への合意を形成するのに有効であった。プロセス全般をFig. 1に示す。

疑問を定式化しアウトカムの重要度を決定する

懸案のヘルスケアクエスションの作成を含めガイドラインで網羅すべき事項についての審議に1回以上の会議が開かれることはめずらしくない。GRADEの適用は、まずPICOまたはその他の構造化された形式を使って適切なクリニカルクエスションを定式化することに始まる¹⁶⁾。この段階を踏むことで、所定の患者アウトカム(O: outcome)に従って、比較戦略(C: comparison)に照らし合わせて考慮された当該診断戦略または介入(I: intervention)の対象となる所定の集団(P: population)に焦点を合わせたクリニカルクエスションが定式化される。

CMAパネルは、IgE依存性CMAが疑われる(すなわち、診断が確定していない)患者(成人と小児)を対象集団とすることを決定した。どの検査を参照基準[reference standard]とするかについ

Fig. 1 診断検査に関する診療ガイドライン作成のための一般的な過程



ては、パネルはCMA診断には盲検化による食物経口負荷試験 (OFC : oral food challenge) が参照検査 (ゴールドスタンダード) として適切で、これを基準に全ての検査の評価を行うべきであると合意した。

指標検査 [index test] (すなわち、評価対象と

なっている検査、または「新しい」検査) は、既存の診断過程における3つの役割、(i)選別 [triage] (侵襲性の高い検査や高価な検査の実施を最小限にする)、(ii)既存検査の置換 [replacement] (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)、(iii)追加 [add-

on] (既存の検査による診断の精度をさらに向上させる) のいずれかの役割を果たすものと考えられる¹⁷⁾。疑問の定式化の段階において、パネルメンバーの間では、OFC実施に関連するリスク、資源活用、負担の理由から、参照基準に取って代わる検査としての指標検査に関心が示された。

診断に関する疑問を取り上げた推奨の作成における大きな課題は、パネルが診断検査の影響や、診断検査精度のデータから得られる定量的情報を理解することである⁴⁾。診断検査に関する推奨の作成において、GRADEは、次のようなアウトカムについて検討するための構造化された枠組みを提供する。すなわち、真陽性 (TP: true positive)、真偽性 (TN: true negative)、偽陽性 (FP: false positive)、偽陰性 (FN: false negative) として分類されることによる患者にとって重要な帰結、不確かな結果による帰結、新しい検査と参照基準による合併症、資源活用 (コスト) である。たとえば、どのような検査でも必ず、正しく分類される患者 (それらはさらに TP と TN に分類できる) と間違っ て分類される患者 (FP と FN) が出てくる。通常、正しい分類は利益あるいは有害アウトカムの減少と関連し、間違っ た分類は疾患の治療や場合によっては疾患の負担減少を妨げるなど、帰結の悪化 (害) と関連している。ガイドラインパネルは、正しい分類 (TP と TN) による利益が、間違っ た分類 (FP と FN) により起こりうる害を上回るかどうかに ついて評価する必要がある。しかしながら、利益と害は、分類後の措置によって生じるものであり、アウトカムの発生確率と患者にとっての各アウトカム (例: 死亡、疾患、症状など) の重要度によって決まるものである。検査により正しく分類される (TP または TN として) ことの利益が、間違っ て分類される (FP または FN として) ことによる害を十分に上回っ ていれば、当該診断検査の使用を推奨するガイドラインパネルは、その検査精度が比較的低い場合でも容認したいと思うかもしれない。

これらの推奨については、CMA 患者の治療経験が豊かな 2 人の臨床医 (AF と LT) により、ま

ず TP, TN, FP, FN と分類されることによる患者にとっての具体的帰結 (すなわち、検査後治療の有無の結果として、通常、当該患者らにとって重要などんなアウトカムが発生するのか。これらはその後の治療の有効性に関する仮説に基づいている) が提起された。次に、ガイドラインパネルはデルファイ法 (Table 1 参照) を使用し、患者にとって重要な帰結を明確に絞り込み、これらの帰結を使用することで、推奨作成時に各検討事項を客観的に比較検討した。正しい診断や間違っ た分類によるこれらの臨床的帰結 (介入の有効性に関する仮説に基づく) は各パネルメンバーに提供され、パネルメンバーらは推奨の作成や各アウトカムの重要度の判断を行う際に、これらの検討事項の比較検討を行った。

ガイドラインパネルメンバーは、各アウトカムによりもたらされる帰結に基づき、全てのアウトカムの相対的重要度について、9 段階 (1: 有用な情報だが意思決定には重要ではない、9: 意思決定に重大である) による事前 [*a priori*] (エビデンスの要約に目を通す前の段階) 評価を行った。

標的状態のリスク (検査前確率) が異なる明確な患者サブグループを特定する

診断を確立する過程で、臨床医は可能性のある標的状態のリストを検討し、それぞれについて確率 (すなわち、検査前確率) を推定する。この確率に応じて、診断検査は標的状態を確定 [rule in] または除外 [rule out] するために使用される。しかし、検査精度や検査の実施に伴う合併症によっては、検査がいずれか一方の目的のためには使用できても、もう一方の目的のためには使用できない場合がある。したがって、診断検査の使用に関する推奨を作成する際には、当該標的状態を有する初期 (検査前) 確率が異なる患者集団について考慮する必要がある。

CMA ガイドラインにおいては、パネルは CMA の検査前確率が低い患者 (例: 非特異的な胃腸症状を持つ患者)、検査前確率が中等度 (すなわち、われわれのシステムティック・レビューに組み入

Table 1 TP, TN, FP, FNのカテゴリに分類されることの患者にとって重要な帰結の例

| 疑問1：CMAの疑いのある患者において、IgE依存性CMAの診断のための皮膚プリックテストを使用すべきか。 | |
|---|--|
| 集団 | 牛乳アレルギー（CMA）の疑いのある患者 |
| 介入 | 皮膚プリックテスト（SPT：skin prick test） |
| 対照 | 食物経口負荷試験（OFC：oral food challenge） |
| アウトカム | |
| TP | TPの小児に対してはOFCが実施され、陽性結果となることから、管理された環境下ではあるが、アナフィラキシーのリスクを伴う。家族には時間や不安の面で負担がかかる。牛乳が除去され、アレルギー用ミルクが使用される。疾患の検査前確率“高”、または負荷試験中にアナフィラキシーショックを起こすリスクの高い小児については負荷試験を省略して治療を行う。その場合の治療の帰結は食物負荷試験を受けた小児と同様である。 |
| TN | TNの小児は家庭で牛乳を摂取してもアレルギー反応はなく、牛乳の除去は行われず、家族に時間的負担はなく、資源の利用（負荷試験の実施やアレルギー用ミルクの使用）を削減できる。小児や家族が抱える不安は家庭によって異なるだろう。症状の他の原因の究明が行われる。 |
| FP | FPの小児にはOFCが実施され、陰性となる。家族には時間や不安の面で不要な負担がかかる。経口負荷試験のために不要な時間と資源が費やされる。CMAの検査前確率が高い小児では負荷試験が省略されることから、牛乳を除去した食事やアレルギー用ミルクによる不要な治療が行われ、栄養不足に陥る可能性がある（例：発育不全、くる病、ビタミンDまたはカルシウム欠乏症）。また家族はストレスを抱え、高額な負担となりうる自己注射が可能なエピネフリン製剤を不要に携帯することとなり、症状の本来の原因の診断が遅れる。 |
| FN | FNの小児は帰宅を許され、家庭で牛乳へのアレルギー反応（場合によってはアナフィラキシー）を起こすことになる。両親は大きな不安を抱え、未知の食材を取り入れることに消極的になる。いくつもの食品目が除去された食事を摂取することになりかねない。症状の本来の原因（すなわち、CMA）が見逃されるため、不要な検査や治療につながる。 |
| 不確かな結果 | 陰性の陽性対照、または陽性の陰性対照のいずれかに相当する。小児には再度SPTが実施されるため、子供と親には精神的苦痛を伴うかもしれない。看護師が費やす時間や外来予約の反復から、資源への影響も考えられる。また、小児に対してはsIgE測定または食物経口負荷試験が実施される。 |
| 検査の合併症 | SPTは不快感や湿疹の悪化を引き起こし、それが原因で精神的苦痛を伴う場合や、両親の不安をあおる場合がある。食物負荷試験はアナフィラキシーや他の症状の悪化を引き起こす場合がある。 |
| 資源の活用（コスト） | SPTでは外来に費やされる時間が増えるが、OFCの方が資源への影響ははるかに大きい。 |

TP：真陽性（CMAであると正しく分類されること）、TN：真偽性（CMAではないと正しく分類されること）、FP：偽陽性（CMAであると間違えて分類されること）、FN：偽陰性（CMAではないと間違えて分類されること）。これらのアウトカムは、常に参照基準（すなわち、牛乳経口負荷試験）と比較して判断される。

れた全ての研究におけるCMAの平均有病率）の患者、検査前確率が高い患者（例：牛乳が原因と考えられるアナフィラキシーの既往がある患者）を対象とする推奨を作成する決定を下した。

われわれは、検査前確率（高、中、低）の概算値を出すために、ガイドラインの情報源として使用したわれわれのシステムティック・レビューに含まれた研究から、該当集団におけるCMAの有病率を抽出した。検査前確率“高”の推定値は、アナフィラキシーの既往があるCMA疑いの患者集団から算出した。この高リスク集団の中で実際にCMAと診断された（参照基準により確認された）患者の割合を、検査前確率“高”の推定値とした。検査前確率“低”の推定値は、非特異的な

胃腸症状を伴う状態が疑われる患者におけるCMAの有病率から算出した。第3の分類である検査前確率“中”は、われわれのシステムティック・レビューに組み込まれた全研究におけるCMAの平均有病率に基づいて推定した。推奨の理解と実行を円滑化するために、われわれはパネルメンバーとの連絡時やガイドライン利用者のために、特定の患者個人の初期CMAリスクが高、中、低のいずれかを推定するために臨床医が使用できる一般的な臨床症状の例を提示した。

このアプローチを使って、検査前確率“高”は約80%、検査前確率“低”は約10%、検査前確率“中”は約40%と推定された。これらの値は、システムティック・レビューからの診断精度と組み

合わせながら、検査前確率に応じた各指標検査の TP, TN, FP, FN に分類されることになる患者 1,000 人あたりの数を算出するために使われた。

検査閾値と治療閾値の推定

診断検査の使用に関する推奨を作成するガイドラインパネルメンバーは、ある診断検査から得られる新しい情報により、標的状態の確率が上昇または低下しうることを確認しなければならない。有用な診断検査は、医師が確信を持って治療を開始できるある特定の閾値（すなわち、治療閾値）をまたいで標的状態の確率を上げうる。あるいは有用な診断検査は、医師が確信を持って検査を中止し、疾患を除外できるある特定の閾値（すなわち、検査閾値）以下に標的状態の確率を下げうる。言い換えると、臨床医にとって価値のある診断検査とは、標的状態に関する不確実性を十分に減らし、その確定または除外を可能にする検査である。

考えられる帰結を考慮しながら CMA を確定または除外する際に臨床医が許容可能とする不確実性の程度を把握するために、われわれはパネルメンバーに CMA の治療閾値と検査閾値の提示を求めた。この情報により、ある診断検査によって不確実性が治療閾値よりも上、あるいは検査閾値よりも下に移動されるにはどの程度の検査精度が必要かが決まる。われわれは、病歴、臨床症状、ならびに指標検査のみ（つまり、参照検査である OFC の実施なし）の結果に基づく臨床セッティングを設定した上でパネルメンバーに治療閾値と検査閾値を推定してもらった。適用したプロセスの詳細は次のとおりである。アウトカムの重要度を決定するためのワークショップと並行して、メール経由でパネルメンバーに「4つの異なるシナリオを使用し、臨床医が CMA 診断のための検査を中止して治療を開始する（CMA の確率が非常に高い）閾値、または当該症状が CMA によるものではないことを患者もしくは両親に知らせる（CMA の確率が非常に低い）閾値を推定するためのワークショップ（ワークショップの詳細は追加

資料 1 の付録 1 に記載）」への参加を求めた。パネルメンバーには、推定してもらいたいこれらの閾値は主観的なものであり、個人の価値観と好みによって左右されるものであることを知らせた。また、提示した 4 つのシナリオは実際の状況を単純化したものではあるが、包括性と簡潔性のある程度兼ね備えた容認可能なものであると言えることを説明した。検査閾値と治療閾値の概念を詳細に説明し、CMA の状況設定を行い、アウトカムの発生確率とコスト推定値を提示した上で、詳細に説明した 4 つの各シナリオ（追加資料 1 の付録 1 を参照）について、検査閾値と治療閾値を参加者に判断してもらった。この調査結果をもとに、検査推奨の閾値が 0%～100% のどこにあるのかを調べた。

エビデンスプロファイルの作成

各疑問について、関連するアウトカムについての情報をまとめたエビデンスプロファイルを 1 つ以上作成した。エビデンスプロファイルは、パネルメンバーが十分な情報に基づく意思決定を行えるよう支援するために、効果推定値や根拠となるエビデンスの質評価を簡潔に要約したものである。理想的にはエビデンスプロファイルはシステマティック・レビューに基づいて作成されるべきである。CMA 診断のための検査の使用に関する既存のシステマティック・レビューを同定できなかったため、われわれはシステマティック・レビューを実施した。われわれは、関連する研究を特定するために MEDLINE, EMBASE, コクランライブラリ (Cochrane Central Register of Controlled Trials, DARE, NHS EED を含む) を検索した。2009 年 9 月までに出版された研究がシステマティック・レビューに組み込まれた。システマティック・レビューの結果を使って、各検査の総合精度を推定した。これは検査を受けた患者 1,000 人あたりの TP, TN, FP, FN に分類される患者数の推定値に相当する。入手可能なエビデンスの質を評価するために、われわれは、GRADE⁴⁾ や QUADAS ツール¹⁸⁾ により示される

分類 (バイアスのリスク [risk of bias], エビデンスの直接性 [directness], 結果の一貫性 [consistency] と精確さ [precision], 出版バイアス [publication bias] の可能性に関わる) を使用した。CMAの異なる初期確率および評価されている各検査の推定精度に基づいて, 検査を受けた患者

1,000人あたりのTP, TN, FP, FNに分類される患者の割合を計算した (Table 2 参照)。検査の精度は, レビュー結果のメタアナリシスに基づいて推定した。ガイドラインにおいて提起されたある1つの疑問のために作成したエビデンスプロファイル (Table 3) に示す。

Table 2 検査前確率20% (標的集団におけるCMAの有病率20%) に基づいて1,000人あたりのTP, TN, FP, FNに分類される患者数を決定するための計算例

| 新しい検査 | 参照基準 | | | 有病率: 20% |
|-------|------|---------------------|----------------------|----------|
| | 疾患あり | | 疾患なし | |
| | 陽性 | TP = 感度 × 200 | FP = (1 - 特異度) × 800 | |
| | 陰性 | FN = (1 - 感度) × 200 | TN = 特異度 × 800 | |
| | | 200 | 800 | 1000 |

Table 3 これらのガイドラインのために実施されたシステマティック・レビューに基づいて作成されたエビデンスプロファイルの例

疑問1, プロファイル1: CMAの疑い患者において, IgE依存性CMAの診断のために皮膚ブリックテストを使うべきか
カットオフ値 ≥ 3mm (全ての集団)

| アウトカム | 研究数 | 研究デザイン | エビデンスの質を下げるかもしれない要因 | | | | | 最終の質 (Final quality) | 1,000人あたりの効果 ¹ | 重要性 |
|--|-----------------|-----------------|---------------------|------------------------|-------------------------|-----------------------|--|-------------------------|--|-------|
| | | | 限界 (limitations) | 非直接性 (Indirectness) | 非一貫性 (Inconsistency) | 不精確さ (Imprecision) | 報告バイアス (Reporting bias) ^{*3} | | | |
| 真陽性 (CMAの患者) | 23件 (2,302人) | 連続的または非連続的なシリーズ | 深刻 ² | なし | 深刻 ³ | なし | なさそう | ⊕○○ 低 | 有病率 80% : 536人 有病率 40% : 268人 有病率 10% : 67人 | 重大 |
| 真陰性 ^{*4} (CMAのない患者) | 23件 (2,302人) | 連続的または非連続的なシリーズ | 深刻 ² | なし | 深刻 ³ | なし | なさそう | ⊕⊕○○ 低 | 有病率 80% : 148人 有病率 40% : 424人 有病率 10% : 666人 | 重大 |
| 偽陽性 ^{*4} (CMAがあると誤って分類される患者) | 23件 (2,302人) | 連続的または非連続的なシリーズ | 深刻 ² | 深刻 ⁴ | 深刻 ³ | なし | なさそう | ⊕○○○ 非常に低 | 有病率 80% : 52人 有病率 40% : 156人 有病率 10% : 234人 | 重大 |
| 偽陰性 (CMAがないと誤って分類される患者) | 23件 (2,302人) | 連続的または非連続的なシリーズ | 深刻 ² | なし | 深刻 ³ | なし | なさそう | ⊕⊕○○ 低 | 有病率 80% : 264人 有病率 40% : 132人 有病率 10% : 33人 | 重大 |
| 不確かな結果 ⁵ | 1件 (310人) | 非連続的なシリーズ | — | — | — | — | — | — | — | 重要 |
| 合併症 | 報告なし | — | — | — | — | — | — | — | — | 重要でない |
| コスト | 報告なし | — | — | — | — | — | — | — | — | 重要でない |

¹ 統合された感度67% (95%CI : 64 ~ 70) と特異度74% (95%CI : 72 ~ 77) に基づく。
² ほとんどの研究において, アトピー性湿疹または消化管症状のある厳選された患者が組み込まれ, 指標検査または標準検査の解釈は他の検査結果がわからない状態で行われていたかどうかについて報告した研究はなく, ある検査結果を解釈する人がもう一方の検査結果を知っていた可能性が非常に高い。1件の研究を除いて, 患者の脱落理由について説明した研究はなかった。
³ 感度の推定値は10% ~ 100%で, 特異度の推定値は14% ~ 100%である。研究の質, 使われた検査, または組み込まれた集団によってこの異質性を説明できなかった。
⁴ これらの患者の帰結に関しては不確実性がある。ある患者では, 他の潜在的な重症状態の診断が遅れるかもしれない。
⁵ 12ヶ月未満の小児を対象とした1つの研究において, 8%の不確かなチャレンジテスト結果が報告されていたが, 不確かな皮膚ブリックテストの件数については報告がなかった。

^{*3} 訳者注: 原文に従ってreporting bias (報告バイアス) としたが, 厳密にはoutcome reporting bias (アウトカム報告バイアス) と区別するために, publication bias (出版バイアス) が適切である。
^{*4} 訳者注: 論文にて記載されている1,000人あたりの効果の数値 (真陰性, 偽陽性) には誤植があり, TN, FPの正しい数値を本表では記載した。著者らに確認済みである。

結果

エビデンスプロファイルには、数値結果の解釈を提示する記述的エビデンス要約を添付した。これらの要約では、文献検索結果を明確に示し、システマティック・レビューに含まれた研究、組み込まれた患者、使用された検査に関する追加情報を提供した。さらに、参照基準と比較した場合に指標検査の使用に予期される利益と不利益や、臨床使用に役立つと考えられる追加情報についても説明し、新しい検査の使用に関する最終的結論を提示した。

全パネルが参加した会議では、メンバーがエビデンス要約をレビューし、ガイドラインを起草し、推奨について議論し、必要に応じて推奨を訂正した。根拠となるエビデンスの質、新しい検査の使用による望ましい帰結と望ましくない帰結のバランス、コスト、患者の価値観と好みを考慮した上で、全ての推奨および推奨の強さについて合意を達成した。投票が必要となった推奨は1つもなかった。検査閾値と治療閾値の推定を経てガイドラインパネルの出した結果にはばらつきがあった。一部のパネルメンバーは、CMAの存在に関する説明のつかない不確実性は一切容認しかねるという見解を示した。このように不確実性の余地を一切認めないとする態度は、指標検査のみの使用に際して非常に高い治療閾値が設定されることから見て取れる。このような高い治療閾値の設定から、常に参照検査（OFC）を実施するタイプの臨床医が特定される。その結果、OFCが必ず実施されるセッティングにおける指標検査の併用はその精度の限界からして無意味であり、併用すべきでないという主旨の推奨を作成した。

推奨の透明性を高め、解釈を促すために、新しい検査の使用による望ましい帰結と望ましくない帰結のバランスについて判断する際にパネルメンバーが想定した価値観と好みについて、推奨ごとに明確に示した。推奨の理解と実行に役立つかもしれないと思われるその他の追加的情報は、「備

考」に提供されている⁵⁾。

GRADEアプローチに従い、推奨を強い、または条件付き（「弱い」とも言う）に分類した。全体で、診断検査の使用に関する15の推奨を作成した。

考察

このアプローチの強みの1つは、診断検査に関する推奨の作成における特有の課題について検討できることにある。治療の有効性についての仮説設定に依存するものではあるが、われわれは検査精度のみに着目するのではなく、GRADEアプローチに従ってTP、TN、FP、FNに分類される患者のリスクと利益の要点を明確に示し、さらには診断検査に関するガイドラインの作成実績の少ないパネルをこのプロセスに組み込むことができた⁴⁾。このような検討事項は、ガイドライン作成時に検討されていない場合は治療を行う臨床医にゆだねられることになる。このワークショップの結果や、診断検査の精度データの理解は一筋縄ではいかないことを踏まえると、診断検査に関する推奨を作成・使用する人々に対してはさらなるサポートの提供が必要である。われわれは、検査前確率と臨床医らの検査閾値と治療閾値を把握し、診療のためのエビデンスに基づく推奨をガイドラインに盛り込むことができた。

このようなガイドライン作成から得られる重要な教訓の1つとして、エビデンスに基づく医療における診断に関わる概念の重要性があげられる。パネルメンバーからは、感度や特異度から診断精度を解釈するのは認知的に困難であるという見解が示された。これらの値を、検査を受けた患者1,000人あたりのTP、TN、FP、FNに分類される患者数に変換し、さらに、間違っ分類された場合に生じる負担を明確に判断することで、パネルメンバーは審査対象となっている診断検査の臨床的影響をより容易に把握することができた。

このことは、認知バイアスを最小限にし患者にとって重要なアウトカムを重視する標準化された

形式でガイドライン作成のためのエビデンスを提示することの重要性を強調している。特に多様な背景を持つ多分野のメンバーが名を連ねるパネルの場合、パネルメンバーがエビデンスに基づく推奨作成で使用する方法やプロセスを熟知していないことがあることがガイドライン作成に関する文献で指摘されているが¹⁹⁾、われわれの結果もこれに合致していた。したがって、公式あるいは非公式なトレーニングやサポートの提供が、理解や積極的参加を促す鍵となる。

われわれは、検査前確率を推定することで、感度と特異度から、検査を受けた患者1,000人あたりのTP、TN、FP、FNに分類される患者数として診断検査の臨床的影響を示すことができた。また、検査前確率の高、中、低を区別することで、臨床医はガイドラインの活用の際に一連の患者を分類することが可能となる。さらに、このようなサブグループへの分類により、新しい検査（指標検査）によりもたらされる各サブグループ特有の利益とリスクについて検討することが可能となる（例：OFCによるアナフィラキシーのような参照基準による有害作用の可能性と照らし合わせて考えた場合、検査前確率“高”の患者では、指標検査の精度が低めでも容認可能かもしれない）。

パネルメンバーからの治療閾値と検査閾値の提示も、推奨に大きな影響を与えた。提示された閾値のばらつきは、CMAの診断において容認可能な不確実性の程度がパネルメンバー間で異なることを浮き彫りにした。これにより、OFCに合併症やリスクが伴うとしても、それ以外の検査で生じる不確実性は容認できないという立場から常に参照検査を実施する臨床医グループの存在が明らかになった。したがって推奨では、このタイプの臨床医に対しては、指標検査の感度と特異度には限界があるためOFCとの併用で指標検査を使用するのは無意味であるという観点から、OFCとの併用を勧めない指針を示している。

結論

本稿では、診断に関する推奨作成へのGRADE適用について説明している。このケーススタディは、診断検査に関する推奨の実行の円滑化や推奨における確信強化のためには診療ガイドラインに何が求められるのかについて、ガイドライン作成者や臨床医に有用な指針を提供する。診断に関する推奨作成に特有の課題については、TP、TN、FP、FNの各カテゴリに合致する患者数を決定する各診断検査の感度と特異度の提示に加え、各カテゴリに分類されることに関連する患者にとって重要なアウトカムを明確に定義したエビデンスプロファイルを作成することにより対処した。さらに、検査前確率や検査閾値・治療閾値に関する検討のような、エビデンスに基づく医療における診断の概念の重要性が示された。ガイドライン全体をとおして、多くの専門分野からのパネルによる構造化されたプロセスにより、方法論的健全性や推奨の臨床的妥当性が確保された。まとめると、推奨に影響するエビデンスや要因について透明性の高い記述を求めるGRADEアプローチのような構造化された枠組みを適用することで、診断検査に関する推奨の作成・解釈・実行に必要なプロセスや決定因子を容易に整理することができる。

追加資料

追加資料1：付録1. 検査閾値と治療閾値の決定

謝辞

本ケーススタディで紹介したガイドラインパネルの取り組みへのメンバー各位のご尽力に謝意を表したい。ガイドライン作成作業は世界アレルギー機構 (World Allergy Organization) の資金援助の下で実施したが、本書の執筆については特定の資金援助を受けていない。

著者の詳細

¹Department of Clinical Epidemiology and Biostatistics,

McMaster University, Hamilton, Ontario, Canada. ² Department of Medicine, McMaster University, Hamilton, Ontario, Canada. ³ Department of Child Medicine, Fatebenefratelli/Melloni Hospital, Milan, Italy. ⁴ Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. ⁵ Allergy and Respiratory Disease Clinic, Department of Internal Medicine, University of Genoa, Genoa, Italy. ⁶ Department of Public Health UFCSPA, Ulbra and Conceicao Hospital, Porto Alegre, Brazil.

著者の貢献

JHはデータの収集、分析、解釈、ならびに原稿の執筆と批判的改訂に貢献した。JLBはデータの収集、分析、解釈、ならびに原稿の執筆と批判的改訂、統計学的分析、研究デザイン、研究の監修に貢献した。LT, JK, EC, AFはデータの収集、分析、解釈、ならびに原稿の批判的改訂に貢献した。ATSはデータの分析と解釈、ならびに原稿の批判的改訂に貢献した。HJSは本研究を考案し、データの収集、分析、解釈や原稿の執筆と批判的改訂、統計学的分析や研究デザインに貢献し、管理および技術的サポートを提供し、研究を監修し、原稿の内容に責任を持つ。著者全員が最終原稿を閲読および承認した。

利益相反

著者全員が、http://www.icmje.org/coi_disclosure.pdfに掲載されている統一利益相反申告書式を記入し（必要であれば責任著者から入手可能）、過去3年間にいずれの著者も当該投稿論文に利害を有すると考えられる者と関係を持たないこと、配偶者・パートナー・子供が当該投稿論文に関わる金銭的関係を持たないこと、JLBとJKはGRADE Working Groupのメンバーであること、HJSは現在GRADE Working Groupの共同議長であることを申告した。JLBとHJSは、マクマスター大学 (McMaster University) が著作権を有するGRADEproの共同作成者である。両者とも、本稿で説明した取り組みに関係する営利目的組織からの金銭の授与は受けていない。

GRADE Working Groupの取り組みには報酬は支払われない。JLBとHJSは世界各国でのGRADEアプローチの普及を支援してきた。当該ガイドライン作成作業は世界アレルギー機構の資金援助の下で実施したが、本稿の執筆については特定の資金援助を受けていない。JH, LT, EC, AFは、利益相反のないことを申告する。

Received: 18 February 2011 Accepted: 10 June 2011
Published: 10 June 2011

参考文献

- 1) Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, Lau J. Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med.* 2005; 142(12 Pt 2): 1048-55.
- 2) Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med.* 2008; 149(12): 889-97.
- 3) Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, Muti P, Jaeschke R, Guyatt GH. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evid Based Med.* 2008; 13(6): 162-3.
- 4) Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams JW Jr, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt GH; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ.* 2008; 336(7653): 1106-10.
- 5) Fiocchi A, Brozek J, Schünemann H, Bahna SL, von Berg A, Beyer K, Bozzola M, Bradsher J, Compalati E, Ebisawa M, Guzmán MA, Li H, Heine RG, Keith P, Lack G, Landi M, Martelli A, Rancé F, Sampson H, Stein A, Terracciano L, Vieths S; World Allergy Organization (WAO) Special Committee on Food Allergy. World Allergy Organization (WAO) Diagnosis and Rationale for Action against Cow's Milk Allergy (DRACMA) Guidelines. *World Allergy Organization Journal.* 2010; 3(4): 57-161.
- 6) Schrandt JJ, van den Bogart JP, Forget PP, Schrandt-Stumpel CT, Kuijten RH, Kester AD. Cow's milk protein intolerance in infants under 1 year of age: a prospective epidemiological study. *Eur J Pediatr.* 1993; 152(8): 640-4.
- 7) Høst A, Halken S, Jacobsen HP, Christensen AE, Herskind AM, Plesner K. Clinical course of cow's milk protein allergy/intolerance and atopic diseases in childhood. *Pediatr Allergy Immunol.* 2002; 13 Suppl 15: 23-8.
- 8) Saarinen KM, Juntunen-Backman K, Järvenpää AL, Kuitunen P, Lope L, Renlund M, Siivola M, Savilahti E. Supplementary feeding in maternity hospitals and the risk of cow's milk allergy: A prospective study of

- 6209 infants. *J Allergy Clin Immunol*. 1999; 104(2 Pt 1): 457-61.
- 9) Kvenshagen B, Halvorsen R, Jacobsen M. Adverse reactions to milk in infants. *Acta Paediatr*. 2008; 97(2): 196-200.
- 10) Venter C, Pereira B, Grundy J, Clayton CB, Roberts G, Higgins B, Dean T. Incidence of parentally reported and clinically diagnosed food hypersensitivity in the first year of life. *J Allergy Clin Immunol*. 2006; 117(5): 1118-24.
- 11) Rona RJ, Keil T, Summers C, Gislason D, Zuidmeer L, Sodergren E, Sigurdardottir ST, Lindner T, Goldhahn K, Dahlstrom J, McBride D, Madsen C. The prevalence of food allergy: a meta-analysis. *J Allergy Clin Immunol*. 2007; 120(3): 638-46.
- 12) Organizations that have endorsed or that are using GRADE [Internet]. Available from: <http://www.gradeworkinggroup.org/society/index.htm>
- 13) World Health Organization. Global Programme on Evidence for Health Policy. Guidelines for WHO Guidelines. EIP/GPE/EQC/2003.1. Geneva; 2003.
- 14) Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, Guyatt GH. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. *J Clin Epidemiol*. 2009; 62(4): 374-9.
- 15) Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess*. 1998; 2(3): i-iv, 1-88.
- 16) Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *CMAJ*. 1988; 138(8): 697-703.
- 17) Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*. 2006; 332(7549): 1089-92.
- 18) Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. 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.
- 19) Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. *Health Res Policy Syst*. 2006; 4: 15.

doi:10.1186/1748-5908-6-62

原本の書誌情報 : Hsu et al.: Application of GRADE: Making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implementation Science* 2011 6:62.

付録1. 検査閾値と治療閾値の決定

パネルメンバー各位

診断に関する疑問に対するガイドラインの推奨は、診断検査の実施が患者にもたらす望ましい帰結と望ましくない帰結のバランスによって決まる。正しい診断を下すこと、すなわち、患者がある状態を持つ、または持たないと正確に分類することは、必ずしも患者にとって重要なアウトカム（例：症状の減少、QOL向上）に影響を及ぼすものではない。

ある疾患の存在を100%確実に断定できる診断検査はほぼ皆無といつてよい。診断の正確さには、わずかではあっても常に不確実性が存在する。一方、診断検査の過程で臨床医が当該疾患の確率が十分に高くさらなる検査を中止し治療を開始できると確信する時点がある。たとえば、牛乳アレルギー（CMA：cow's milk allergy）の疑いが高い（例：最近アナフィラキシー反応を起こした）患者で、皮膚プリックテスト（SPT：skin prick test）で強い膨疹がある陽性結果となった場合は、食物負荷試験を実施せず、臨床症状とSPT結果に基づいて治療を開始することが多い。この決断の背景には、間違った診断をして実際には牛乳アレルギーではない人に治療を開始してしまうことによるリスクよりも、負荷試験実施によるリスクの方がはるかに高いという想定がある。

同様に、臨床医が当該疾患の確率が十分に低く、いかなる治療も行わずにさらなる検査を中止する時点がある。たとえば、CMAの疑いが低い（例：軽度の皮膚病変があるがアナフィラキシー反応を起こしたことの無い）患者でSPTが陰性結果となった場合、食物負荷試験を実施せず、臨床症状やSPTのみの結果から症状の原因がCMAである可能性は低いと決断し、CMA診断検査を中止する可能性がある。

CMA疑いの患者に対し、食物経口負荷試験（OFC：oral food challenge）ではなくSPTや血清IgE（sIgE）の使用を推奨可能にするには、CMAの確率として次の2つの閾値を設定する必要がある。

1. これを下回った時点で臨床医が病歴、臨床症状、SPTまたはsIgEの結果のみ（OFCは実施しない）に基づいて検査を中止しCMA治療は不要であると確信できる確率閾値。言い換えると、診断がCMAである可能性がどの程度低ければ、CMA診断のためのさらなる検査を中止し患者または家族にCMAではないことを告げることができるか。
2. これを上回った時点で臨床医が病歴、臨床症状、SPTまたはsIgEの結果のみ（OFCは実施しない）に基づいてOFCを実施せずに治療を開始する確率閾値。言い換えると、OFCを実施する前の時点で

診断がCMAである可能性がどの程度高ければ、OFCを実施せずに治療を開始するだろうか。

OFC検査の利益と不利益

たとえば、OFCを実施することで、CMA患者が牛乳や乳製品を摂取し続け、アナフィラキシー反応のリスクを抱えながら症状を経験し続けるリスクは軽減される。さらに、CMAでない患者が牛乳や乳製品を除去した食事を不要に摂取し、場合によってはアレルギー用ミルクの摂取によって家族または医療システム（アレルギー用ミルクが還付対象ならば）に経済的負担がかかるリスクを軽減できる。

食物負荷にはアナフィラキシーのリスクがあることから（食物アレルギーの検査前確率が50%未満の小児で陽性結果が出たものの10%～30%に発生すると推定される）、蘇生器具が揃った管理された医療セッティングにおいて実施されるべきである。

食物負荷試験による合併症のリスクは、試験中のアナフィラキシーショックのリスクによって左右される。われわれは、OFC検査中にショックを起こすリスクのある患者集団を、高リスク集団（例：最近重篤なアナフィラキシー反応を起こした患者、喘息の既往患者、複数のアレルゲンに感作された患者、sIgEが非常に高値の患者）と、軽度なアナフィラキシーを含む比較的軽度な反応を起こすと考えられる低リスク集団（すなわち、上述の危険因子を持たない患者）の2つに大別する。

CMAの予後

診断されないままだと、患者は次の症状を経験する。

1. アナフィラキシー（11%）[そのうち半数（すなわち、CMA患者の5%程度）は軽度ないし重篤なアナフィラキシー反応を起こす]
2. じんましん（18%）
3. 喘息または喘鳴（21%）
4. 鼻炎（20%）
5. 疝痛（27%）
6. 湿疹（35%）
7. 嘔吐（48%）
8. 下痢（53%）

CMA治療の利益と不利益

CMAを持つ年長児、青年、成人は、牛乳や乳製品を摂取しないため、ビタミンDとカルシウム欠乏のリスクがある。CMAの幼児を持つ授乳中の母親が自身の食事で乳タンパク質の摂取を控えている場合は、自身の食事の栄養状態を評価する必要があるかもしれない。

アレルギー用ミルクには高額の費用がかかる場合がある。小児1人の治療にかかる推定コストの概算を以下の表に示す（国や地域によって異なるかもしれないが、コ

ストについて検討する際にはこれらの概算値を参考にしたい。

| 牛乳/アレルギー用ミルク | 1リットルあたりのコスト [米ドル(ユーロ)] | 半年あたりのコスト [米ドル(ユーロ)] |
|--------------|----------------------------|-------------------------|
| 通常の牛乳 | 1.2 (0.9) | 100 (75) |
| 通常のアレルギー用ミルク | 2.5 (2.0) | 230 (160) |
| 高度加水分解乳 | 9 (6) | 800 (550) |
| 大豆乳 | 7 (5) | 750 (450) |
| 米乳 | 9 (6) | 800 (550) |
| アミノ酸乳 | 20 (14) | 1800 (1250) |

課題

CMA疑いの患者にOFC検査を実施または実施しないことの望ましい帰結と望ましくない帰結のバランスについてのあなたの予測を考慮した上で、次の質問に教えてください。0% (CMAでないことと確信できる) から100% (CMAであると確信できる) の尺度で評価してください。

研究セッティングではなく、日常診療を想定して以下の質問について考えてください。

検査閾値

Q1

OFCの実施中にアナフィラキシーショックを起こすリスクが低い患者では、牛乳アレルギーの診断の確率が何% (0%~100%) であれば、検査 (SPT, sIgE, OFC など) を中止し、CMAの可能性は十分低いと判断し、患者や両親にCMAではないことを告げることができるか。

A : %

Q2

OFCの実施中にアナフィラキシーショックを起こすリスクが高い患者では、牛乳アレルギーの診断の確率が何% (0%~100%) であれば、検査 (SPT, sIgE, OFC など) を中止し、CMAの可能性は十分低いと判断し、患者や両親にCMAではないことを告げることができるか。

A : %

治療閾値

Q3

OFCの実施中にアナフィラキシーショックを起こすリスクが低い患者では、牛乳アレルギーの診断の確率が何% (0%~100%) であれば、牛乳を使用したOFC検査を実施せずに治療 (牛乳除去食士アレルギー用ミルク) を開始するか。

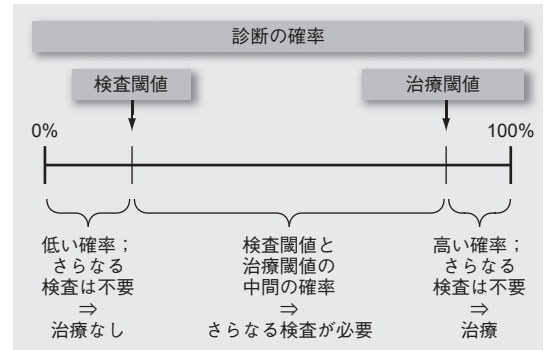
A : %

Q4

OFCの実施中にアナフィラキシーショックを起こすリスクが高い患者では、牛乳アレルギーの診断の確率が何% (0%~100%) であれば、牛乳を使用したOFC検査を実施せずに治療 (牛乳除去食士アレルギー用ミルク) を開始するか。

A : %

検査閾値と治療閾値の概念の詳細については以下を参照すること。



良好な検査に対して臨床医が期待するのは、その検査結果によって、診断の確定または除外を可能とするような確率の変化がもたらされることである。しかし検査は、懸案の疾患が存在する確率を推移させるものに過ぎない。

検査前確率が検査閾値を下回る場合は、疾患を除外するためのさらなる検査は無用である。

検査前確率が治療閾値を上回る場合は、疾患の確率をさらに上げるためのさらなる確認検査は無用である。

確率が検査閾値と治療閾値の間にある場合は (診断の不確実性がある患者)、検査の実施は有用である。検査結果により、確率がいずれかの閾値をまたぐとき、その検査結果の価値は最大となる。

診断の不確実性 (検査閾値と治療閾値の間) がある患者において、ある検査結果により、懸案の状態の確率が検査閾値以下となる場合、それは当該状態の可能性が極めて低いこと、当該状態のための追加的な検査や治療に関連した不利益が予期される一切の利益を上回ること、当該状態のための検査や治療がこれ以上行われるべきでないことを示す。

検査結果により、疾患の確率が治療閾値以上となる場合、それは当該状態の可能性が極めて高いこと、当該状態の確率をさらに上げるための確認検査は不要であること、治療から予期される利益が考えられる害を上回ることを示す。