

本資料は、Holger Schunemann 教授が「診断研究における GRADE システムの活用」を指導するために作成した workshop package(2010 年)を翻訳したものです。診断研究に関するシステムティックレビューや診療ガイドラインにおいて GRADE を利用する場合は、治療介入に関する GRADE の基本的な理解が必須です。また、診断 GRADE に関しては、既報の BMJ 論文または今後発表される JCE 論文を参考にしてください。(相原、2011 年 5 月)

ワークショップ: []

開催場所: []

ファシリテータ: []

課題

- 小グループで作業を行う。
- グループ全体への報告を行う担当者を選ぶ(メモをとる)。
- 時間を管理する。
- 次の指示に従う。

指示 (要点: 指示の詳細は表示された各頁をみるとよい)

1. システムティックレビュー(付録)の抄録を読む。
2. レビューで提起された**臨床的クエスチョン**とその構成成分を同定する(2頁目、3頁目のワークシート)。取り組みたいレビューから**主な比較**を同定する(例、虚血性脳卒中に対する MRI vs CT)。レビューの関連セクションを読み、欄外でマーキングした文面に着目する。
3. 各検査結果に関する患者にとって重要な帰結を検討する(3、4頁目)。各アウトカムの相対的重要性を決定する。
 - アウトカムの相対的重要性を判断する。
 - 自身の判断を同僚と比較する。
 - 不一致について検討し合意に達するようにする。
4. エビデンスプロファイルの中のエビデンスの質評価を完了させる(5頁目、6頁目と7頁目の指示)。
 - システムティックレビューのマーキングされた文面へ進み、エビデンスの質を評価する。
 - 各アウトカムのエビデンスの質について判断を下す。
 - 全体的なエビデンスの質について判断を下す。
5. 望ましいアウトカムと望ましくないアウトカムの大きさに関する情報を提供する(8頁目)。真陽性(TP)、真陰性(TN)、偽陽性(FP)、偽陰性(FN)を計算するために感度と特異度を得る。
 - 仮定の疾患事前確率(有病率)として 60%を使う。
 - 真陽性、真陰性、偽陽性、偽陰性を計算する(8頁目)。
 - アウトカムに関する情報をテーブルに記入する。
6. 当該検査の使用に関する推奨を作成する(10 頁目と 11 頁目)。

課題の途中で、診断に関するシステムティックレビューの SoF テーブルまたはエビデンスプロファイルがどのように提示されるか、または最終的に提示するものがどれほど改善されるかを考えていただきたい。

1. レビューの詳細と提起された臨床疑問を特定する。

レビューのタイトル (どのレビューを利用するかを、グループ内で合意する): _____

対処している疑問(PICO)を考えてみる。主な比較を選択する(例、CT vs 参照基準、または DWI vs 参照基準)。

次の要素を同定する。

患者 (P: 集団) および有病率:

新規/指標検査 (I: 介入):

既存/比較/参照検査 (C: 比較):

アウトカム (O): [治療の有無にかかわらず、患者が経験するアウトカムを全て考慮する(次頁ワークシート 1 を使用)]

カットオフ基準(関連があるならば):

最終的な疑問(PICO):

2. 意思決定にとって最も重要なアウトカムを選択する。

状態、疾患あるいは問題の帰結(例、死亡率、障害、出血合併症など)という意思決定にとって最も重要なアウトカムを選択する。正しい診断がなされたかどうかに限らず全ての帰結(例、偽陽性とその後の治療による帰結の有無)を検討すること(偽陰性は潜在的に有益な治療を受けることはなく、偽陽性は潜在的に有害な治療を受ける可能性がある)。それが不慣れなトピックならば、隣の人と相談するか、仮説や最も有力な推測を立てること。

アウトカムについて以下のことを考慮する。

- アウトカムは疾患の自然経過の一部である。
- アウトカムはシステマティックレビューまたは個別研究において報告されている可能性がある。
- アウトカムは、比較されている検査の実施に伴う合併症を含めて、検査の使用・非使用の決定者にとって重要かもしれない(利益と有害作用の両方、また関連があればコストも含める)。
- すべてのアウトカムに関する治療の有効性と有害作用の発生率を考慮する。

帰結に関する想定の一覧をまとめる(本頁の以下のワークシートを参照)。

- 例、死亡率の相対リスク減少: 50%
- 例、重症有害アウトカムの相対リスク増加: 20%

考えの参考になるのであれば、何が患者に起こりうるかのフロー図を作成すること。

すべてのアウトカムを列記し、それらが4つの各カテゴリでどれくらい頻繁に起こるかを指定する(灰色の例を上書きする)。

アウトカム	適切な治療の効果についての想定 (RRR)	TP, TN, FP, FN に記入する (正しい治療のありなしにかかわらず、どんな率でそのアウトカムは起こるか?)			
		TP (%)	TN (%)	FP (%)	FN (%)
1. 30日後の死亡率	50%	35% ¹	5% ²	7% ³	70% ⁴
2.					
3.					
4.					
5.					
6.					
7.					

脚注:

¹ 疾患ありと正しく診断され治療も成功する(RRR 50%)。

² 30日以内とは関係なく死亡する割合。

³ 治療の有害作用により死亡率が2%増える。

⁴ 治療が受けられない、または治療が遅れるために死亡する割合。

3. アウトカムの相対的重要性を評価する。

疾患があるかどうかの分類が正しいか正しくないかに限らず、**患者にとって重要な帰結**を考慮する。各アウトカムの**相対的重要性**を評価し、どのアウトカムがその検査を使うための決断に重大かを決定する。ワークシート1で想定したパーセンテージを使うこと。帰結とパーセンテージを列記することから始め、次のスケールにしたがってアウトカムを評価する。

重要性を 9-point scale で評価する(数回にわたって同じ評価ができるように注意していただきたい)。

1～3	重要でない(エビデンスプロファイルに含めない)
4～6	重要だが意思決定には重大ではない(エビデンスプロファイルに含める)
7～9	意思決定に重大である(エビデンスプロファイルに含める)

患者にとって重要な帰結	重要性
真陽性 (TP): (例、死亡率 35%、重症脳卒中 5%、など)	
真陰性 (TN):	
偽陽性 (FP):	
偽陰性 (FN):	
不確かな結果*:	
コスト:	

* 解釈不能、不確定、あるいは中間的な検査結果のことをさす。

4. レビューにおける各アウトカムについて研究全般のエビデンスの質を評価する。
(このパッケージの最後に示す例を参照のこと)

診断精度研究の質の評価

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					最終的な質 (記号)
			バイアスのリスク	非直接性(システマティックレビューに基づく)	非一貫性	不精確さ	出版バイアス	
真陽性 (TP)								
真陰性 (TN)								
偽陽性 (FP)								
偽陰性 (FN)								
不確かなもの								
コスト								
脚注:								

診断精度研究の GRADE 評価基準

根底の研究デザイン

診断が不確実な患者において、検査結果を適切な参照基準と直接的に比較した妥当性のある診断精度研究(横断研究あるいはコホート研究)は最初、「高」の質のエビデンスと評価される。しかしこのような研究は稀である。

エビデンスの質を下げるかもしれない要因

- ↓ 研究のデザインや実施における限界(バイアスのリスク)
- ↓ 非直接性(比較対照または集団、新検査、比較検査、およびアウトカム)
- ↓ 研究結果の非一貫性
- ↓ 不精確な結果
- ↓ 出版バイアスの可能性が高い

もしグレードを下げることを正当化する要因があるならば、その限界が**深刻か**(1段階グレードを下げる)、または**非常に深刻か**(2段階グレードを下げる)を検討する。

診断精度研究のエビデンスの質を決定あるいは下げる要因

質を決定する要因	治療の選択肢に関するエビデンスとの違い
研究デザイン	精度研究の、マネジメント試験の場合とは異なる基準 診断が不確実な患者において、適切な参照基準と直接的に比較した横断研究またはコホート研究は、「高」の質と考えられるが、次の要因によっては、「中」、「低」または「非常に低」になりうる。
研究のデザインや実施における限界(バイアスのリスク)	精度研究の、マネジメント試験の場合とは異なる基準 単一のコホートとして連続した患者が組み込まれているべきであり、疾患状態によって分類されるべきではなく、選択ならびに紹介プロセスが明確に説明されているべきである。新しい検査法ならびに詳細に説明された参照基準が、同一の患者集団に属する全患者を対象に実施されるべきであり、代替検査や参照基準の結果が評価者にわからないようになっているべきである。
エビデンスの非直接性 アウトカム 患者集団、診断検査・介入、比較検査・介入、非直接的な比較	精度研究の、マネジメント試験の場合と同様の基準 精度研究は患者にとって重要なアウトカムに関する直接的なエビデンスを提供することはない。真陽性、真陰性、偽陽性、偽陰性の違いが患者にとって重要なアウトカムに与えると想定される影響のバランスについて、検査の合併症やコストと照らし合わせながら、推論しなければならない(診断精度研究は一般に、治療の代理アウトカムの場合と同様、アウトカムの非直接性により推奨作成のためのエビデンスとしては質が低い)。 精度研究の、マネジメント試験の場合と同様の基準 以下の場合、エビデンスの質は下げられる。 1) 研究の集団と推奨の対象集団との間に重要な違いがある(例、疾患の範囲または併存症)。 2) 研究の検査、および研究検査を適用する担当者の専門知識と、推奨の対象となる状況との間に重要な違いがある。 3) 比較されている複数の検査が、それぞれ異なる研究の中で参照基準(ゴールドスタンダード)と比較されており、同一の研究内で直接比較されていない場合は、エビデンスの質は下げられることがある。
結果の非一貫性	精度研究の、マネジメント試験の場合と同様の基準 診断精度研究では、(相対リスクや平均差ではなく)感度、特異度、または尤度比における説明のつかない非一貫性により、エビデンスの質を下げることもある。
結果の不精確さ	精度研究の、マネジメント試験の場合と同様の基準 診断精度研究では、検査精度、または真陽性・真陰性・偽陽性・偽陰性の推定値における信頼区間が広い場合、エビデンスの質を下げることもある。
出版バイアス	精度研究の、マネジメント試験の場合と同様の基準 出版バイアスのリスクが高い場合(例、新しい介入や検査に関する小規模研究からのエビデンス、またはファンネルプロット上の左右非対称性)、エビデンスの質を下げることもある。

参考文献: 1) Guyatt GH et al. What is "quality of evidence" and why is it important to clinicians? BMJ 2008;336:995-8. 2) Schünemann HJ et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 2008;336:1106-10.

QUADAS による診断精度研究の質基準

(Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. 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. Available from:

<http://srdta.cochrane.org/>)

#	項目	はい	不明	いいえ
1	患者の範囲は、実際の診療で検査を受けることになる患者を代表していたか？(代表的な範囲か)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	参照基準は標的状態を正確に分類できると考えられるか？(参照基準は許容できるか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	参照基準と指標検査の間の期間が短く、合理的に考えて、検査と検査の間の期間で標的状态に変化はなかったといえるか(検査間の遅延は許容できるか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	診断の参照基準によって検証されたのは、サンプル全体かまたはサンプルから任意に選択された者か？(部分的検証が回避されたか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	患者は指標検査の結果に関わらず、同じ参照基準による検査を受けたか？(鑑別的検証が回避されたか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	参照基準は指標検査と独立していたか(すなわち、指標検査が参照基準の一貫として行われていなかったか)？(混同は回避されたか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	参照基準の結果の解釈は、指標検査の結果がわからない状態で行われたか？(指標検査の結果は盲検化されたか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	指標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？(参照基準の結果は盲検化されたか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	実際の診療で検査が使われる場合と同様の臨床データが入手可能だったか？(関連する臨床情報)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	解釈不能または中間的な検査結果は報告されていたか？(解釈不能の結果は報告されたか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	研究からの脱落については説明があったか？(脱落は説明されたか)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. 想定有病率の患者集団における、検査を使用した効果推定値を提示する。

想定した典型的な検査前確率(有病率): 60%

検査所見	
統合された感度	_____ (95% CI: _____ to _____)
統合された特異度	_____ (95% CI: _____ to _____)
帰結	
	検査を受けた 1,000 人 あたりの人数*
TP	
TN	
FP	
FN	
不確かな結果**	
コスト	

* すべての結果は有病率 %、および統合された感度と特異度に基づき、1,000 人あたりの患者数として示されている。

** 不確かな結果とは、解釈不能、不確定あるいは中間的な検査結果のことである。

メタアナリシスからの統合された感度と特異度、および標的状態の想定有病率に基づいた検査結果をもとに、2x2 表を使って患者の絶対数を計算する。

		参照基準	
		疾患あり	疾患なし
新しい検査	陽性	TP _____	FP _____
	陰性	FN _____	TN _____
有病率: _____ %		_____	_____
			1000

想定有病率 20% の場合の計算例

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

5. 患者にとって重要なアウトカムと、それらが診断精度にどれほど直接的に関連しているかを検討する。

患者にとって重要なアウトカムに基づいた結果要約(SOF)テーブル**

集団/セッティング:

新検査/カットオフ値:

比較検査/カットオフ値:

参照検査:

患者にとって重要なアウトカム (重要性) (2)	患者 1,000 人年あたりの結果			診断検査精度のエビデンスの質	診断検査の精度: 感度 __ (CI __ - __) 特異度 __ (CI __ - __)	想定	患者にとって重要なアウトカムに関するエビデンスの質
	検査前確率 __%	検査前確率 __%	検査前確率 __%				
死亡率 (9)	0 人超の患者が生存	1 人超の患者が生存	1 人超の患者が生存	⊕⊕○○ 低 (4)	真陽性は早期診断によって死亡率を改善し、偽陽性は診断を遅れのために死亡率を悪化させる。	異なるステージの CKD で一様に、毎年 10% の死亡リスクが増加すると推定した。CKD の早期診断についても、死亡率 RRR 10% と推定した。	⊕○○○ 非常に低 (6)

脚注: **

(1)

6. エビデンスから推奨へ

6.1. 全体的なエビデンスの質を決定する。

すべての**重大な**アウトカムに関する全体的なエビデンスの質は:

1つを選択する	記号	質	解釈
<input type="checkbox"/>	⊕⊕⊕⊕	高	今後の研究によって効果または精度の推定値に対する確信が変わる可能性は非常に低い。
<input type="checkbox"/>	⊕⊕⊕○	中	今後の研究によって効果または精度の推定値に対する確信に重要な影響がおよぶ可能性があり、その推定値が変わる可能性がある。
<input type="checkbox"/>	⊕⊕○○	低	今後の研究によって、効果または精度の推定値に対する確信に重要な影響がおよぶ可能性が非常に高く、推定値が変わる可能性がある。
<input type="checkbox"/>	⊕○○○	非常に低	効果または精度のあらゆる推定値が非常に不確実である。

6.2. 価値観と好み (検討された各アウトカムに関する一連の価値観を想定する)

例: 関連する治療には重要な利益と害のいずれもあるため、真陽性、偽陰性、および偽陽性に高い価値を置くが、真陰性と資源活用には比較的低い価値を置く。

アウトカム	価値観と好み
TP	
TN	
FP	
FN	
不確かな結果	
コスト	

6.3. 推奨草案

7. 推奨の強さを決める(強い、または条件付き・弱い)

判断を下すために以下の表を使う。推奨が「強い」、または「条件付き」(別名、「弱い」)かどうかは本表の中の4要因によって決まることになる。はいの回答が多いと、強い推奨の可能性が高くなる。自身の判断についての説明を必ず追加するようにする。

推奨の強さを弱くする要因	判定	説明
<p>高いまたは中等度の質のエビデンス (「高」または「中」の質のエビデンスはあるか?)</p> <p>エビデンスの質が高いほど、強い推奨の可能性が高くなる。エビデンスの質が低いほど、条件付き/弱い推奨の可能性が高くなる。</p>	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
<p>利益と害・負担のバランスに関する確実性 (確実性があるか?)</p> <p>望ましい帰結と望ましくない帰結の差が大きいほど、その差についての確実性が高くなり、強い推奨の可能性が高くなる。正味の利益が小さいほど利益の確実性が低くなり、条件付き/弱い推奨の可能性が高くなる。</p>	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
<p>価値観の確実性または類似点 (確実性があるか?)</p> <p>価値観や好みにばらつきが少ないほど、または確実性が大きいほど、強い推奨の可能性が高くなる。</p>	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
<p>資源の影響 (消費される資源は期待される利益に見合うか?)</p> <p>考慮された代替よりも介入のコストが高く、意思決定に関連したほかのコストが高いほど、すなわち消費される資源が多いほど、条件付き/弱い推奨の可能性が高くなる。</p>	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	

討議によって合意に至らなかった場合、パネルは以下の表を活用することによって、入手可能なエビデンスとその質、利益と不利益、価値観と好み、そして資源の利用(コスト)の解析に基づき、ある特定の管理選択肢に関する推奨についての自らの見解(投票)を記録することができる。この評価は各介入の使用・非使用についての推奨の強さにマッピングされる。

もし投票の必要があるならば：各カテゴリの推奨の投票数を記入する。				
介入の望ましい帰結と望ましくない帰結のバランスに関する評価者の見解	望ましい帰結が望ましくない帰結を明らかに上回る。	望ましい帰結が望ましくない帰結をおそらく上回る。	望ましくない帰結が望ましい帰結をおそらく上回る。	望ましくない帰結が望ましい帰結を明らかに上回る。
推奨の強さ	介入支持の強い推奨	介入支持の条件付き(弱い)推奨	介入反対の条件付き(弱い)推奨	介入反対の強い推奨
推奨の表現	「～をする」ことを推奨する。	「～をする」ことを提案(条件付きで推奨)する。	「～をしない」ことを提案(条件付きで推奨)する。	「～をしない」ことを推奨する。
投票数				

推奨の強さ

- 強い
- 条件付き(弱い)

最終的な推奨

強さ:	エビデンスの質:

根底にある価値観と好みについての想定

--

注意

--

他の例

Hamon M et al. Diagnostic performance of multislice spiral computed tomography of coronary arteries as compared with conventional invasive coronary angiography: a meta-analysis. J Am Coll Cardiol 2006;48:1896–1910.

疑問: 冠動脈性疾患(CAD)の診断に、従来の冠動脈造影よりもマルチスライススパイラル CTを使用すべきか?

患者または集団: 冠動脈疾患の疑いの成人患者

セッティング: 組み込まれた試験は欧州と北米において実施されたものである。

診断精度研究の質評価 — 例

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
			バイアスのリスク	非直接性	非一貫性	不精確さ	出版バイアス	
真陽性 (CADの患者)	21件 (1570人)	横断研究 ¹	深刻な限界なし	不確実性はわずかがほとんどない	深刻な非一貫性 ³	深刻な不精確さなし	検出されない ⁴	⊕⊕⊕○ 中
真陰性 (CADのない患者)	21件 (1570人)	横断研究 ¹	深刻な限界なし	不確実性はわずかがほとんどない	深刻な非一貫性 ³	深刻な不精確さなし	検出されない ⁴	⊕⊕⊕○ 中
偽陽性 (CADありと誤って分類された患者)	21件 (1570人)	横断研究 ¹	深刻な限界なし	不確実性はわずかがほとんどない	深刻な非一貫性 ³	深刻な不精確さなし	検出されない ⁴	⊕⊕⊕○ 中
偽陰性 (CADなしと誤って分類された患者)	21件 (1570人)	横断研究 ¹	深刻な限界なし	いくらかの不確実性あり ²	深刻な非一貫性 ³	深刻な不精確さなし	検出されない ⁴	⊕⊕○○ 低
不確かな結果	–	–	–	–	–	–	–	–
コスト	–	–	–	–	–	–	–	–

¹ 全ての患者が従来の冠動脈造影を受けるように選択されたため、全体的に冠動脈性疾患の確率が高かった(組み込まれた研究における有病率の中央値は63.5%、範囲6.6~100%)。

² 診断の遅れや心筋梗塞による有害な影響を招きかねない偽陰性に関わるエビデンスの直接性にいくらかの不確実性があるため、偽陰性の検査結果の帰結に関するエビデンスの質を「高」から「中」に下げる。

³ 感度(冠動脈造影が陽性でCTスキャンが陽性の患者の割合)、特異度(冠動脈造影が陰性でCTスキャンが陰性の患者の割合)、尤度比、診断オッズ比の結果に統計的に有意で説明のつかない異質性があったため、真陽性、真陰性、偽陽性の検査結果の帰結に関するエビデンスの質を「高」から「中」に、偽陰性の検査結果に関しては「中」から「低」に下げる¹³。

⁴ 出版バイアスの可能性が除外されたわけではないが、エビデンスの質のグレードを下げるほどではない。

結果の要約 (Summary of findings) – 例 想定された事前確率(有病率)は20%

検査の所見	
統合された感度	0.96 (95% CI: 0.94~0.98)
統合された特異度	0.74 (95% CI: 0.65~0.84)
帰結	
	1,000人あたりの人数 ¹
TP ²	192
TN ³	592
FP ⁴	208
FN ⁵	8
不確かな結果 ^{6,7}	–
コスト ⁷	–
	重要性
TP ²	8
TN ³	8
FP ⁴	7
FN ⁵	9
不確かな結果 ^{6,7}	5
コスト ⁷	5

¹ すべての結果は、有病率20%、ならびに統合された感度と特異度に基づいて、1,000人あたりの患者数として示されている。

⁶ 不確かな結果とは、解釈不能、不確定、あるいは中間的な検査結果のことを指す。

² 薬物療法、血管形成術とステント、バイパス手術の実施が決まるため、重要である。

³ 患者が有害作用との関連がある不要な介入をうけなくてもすむため、重要である。

- ⁴ 患者が薬物や侵襲性の高い処置により不要な有害作用にさらされるため、重要である。
- ⁵ 患者が有効な治療を受けないことで冠動脈イベントのリスクが増加するため、重要である。
- ⁶ 解釈不能、不確定、あるいは中間的な検査結果のことを指す。これが原因で、不安、対処方法についての不確実性、精査、治療の実施または非実施によるマイナスの帰結の可能性が生じるため、重要である。
- ⁷ これらの帰結に関する結果については、入手可能なデータの観点からは厳密には確かでないため、報告されていないが、重要である。