

## 附録 4.3 GRADE ワークショップ資料 (2008)

本資料は、GRADE ワークショップ (UMC Utrecht, 2008) で利用されたものを和訳したもので、GRADE pro を利用できる環境で活用するのが望ましい<sup>k</sup>。本資料に含まれるワークシートをそのままコピーして利用して GRADE の理解に役立てることも可能である。

### ワークシート

・ワークシート 1：アウトカムのチェックリスト	136 頁
・ワークシート 2：各アウトカムについての研究全般に関するエビデンスの質評価	137 頁
・GRADE エビデンスの質評価基準と定義	138～140 頁
・ワークシート 3：エビデンス・プロファイル	141 頁
・ワークシート 4：ガイドラインパネルが検討する推奨草案	142 頁

### 課題

- ・コクランレビューを使って課題を完了させる。
  - ・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。
  - ・時間管理
  - ・次の指示を読む。
1. 与えられたシステマティック・レビュー (RCTs) に目を通しておく。抄録を読む
  2. クリニカル・クエスチョンを PICO (患者集団, 介入, 比較, アウトカム) 形式で定式化する。小グループ内で行う
  3. 比較において重要な関連アウトカムを 7 個まで選択する (次の提案を考慮すること)

#### 提案

- (a) 関連アウトカムのリストを作成する (本資料の 3 頁目の[ワークシート 1](#)を参照)。
  - ・ 1 次アウトカム：システマティック・レビューの著者が 1 次アウトカムであると位置づけたアウトカムをリストアップする (「レビューの研究について検討するための基準 (Criteria for Considering Studies for This Review)」を参照すること)。
  - ・ 報告されたアウトカム：別のアウトカムについてデータが報告されている場合は、そのアウトカムを追加する (「結果と抄録 (Results and Abstract)」を参照すること)。
  - ・ レビューで報告されていないその他のアウトカムで、介入の使用・非使用の決定者にとって重要であると考えられるアウトカムがあれば、追加する (利益と有害作用の両方を含め、また関連があればコストも含める)。
  - ・ 各小グループ内で、どのアウトカムが重要でエビデンス・プロファイルに含めるべきか、について合意すること ([ワークシート 1](#))。
- (b) そのリストの中から、ガイドラインパネルや推奨作成者にとって最も重要で、エビデンス・プロファイルに含めるべきであると考えられるアウトカムを最多で 7 件まで選択する。これを空のエビデンス・プロファイルに転記するか、GRADE pro を使用する (本資料の 8 頁目の[ワークシート 3](#)を利用、または GRADE pro にアクセス)。小グループまたはペアを組んで作業すること。

<sup>k</sup> 2008 年 9 月, Yngve Falck-Ytter 氏, Holger J. Schunemann 氏らが, GRADE ワークショップのために作成したものである。

4. GRADE アプローチに従い、各アウトカムに関するエビデンスの質を評価する。

提案

- 当該アウトカムに関するエビデンスの質を決定するため、[ワークシート 2](#)（この資料の 4 頁目）記入。
- この資料の 5～7 頁の、[GRADE エビデンスの質評価基準と定義](#)を参照すること（138～140 頁）。
- 関連情報についてはレビューの次の項に記載されている。  
 「研究の定義説明 (Description of Studies), 方法論的質 (Methodological Quality), 考察および結果について (Discussion and Results)」
- エビデンス・プロファイル（141 頁）の中の「質」列（右から 2 列目）に記入する。

5. GRADE pro を使用せずに紙面でエビデンス・プロファイルを作成した人は、本ステップをとばして、「GRADE pro を使って GRADE エビデンス・プロファイルを作成」に進む。

提案

- わからないことがある場合、GRADE pro のヘルプファイル / ハンドブックを参照すること。不明の用語にカーソルを当てて（カーソルを当てるとクエスションマークが表示）クリックすると、ヘルプファイルが開く。
- ノート型パソコンをもっている参加者が何人いるのかにより、ファシリテータのパソコンを利用してグループでデータ入力を行ったり、2 人 1 組に分かれて各組に 1 つまたは複数のアウトカムを割りあてたりしてもよい。
- 当該アウトカムに関するエビデンスの質については、[GRADE エビデンスの質評価基準と定義](#)に基づいて判断すること（GRADE pro ではこれが自動計算される）。
- さらに、質の定義に関しては、「各アウトカムについての研究全般に関するエビデンスの質」も参照すること（137 頁）。

6. エビデンス・プロファイルを使用してエビデンスから推奨へと移行する。

提案

- [ワークシート 4](#) を使用して推奨度を「強い」または「弱い」に判定する。
- 「強い推奨と弱い（条件付き）推奨の定義」および「強い推奨と弱い（条件付き）推奨の意味」も参照すること（143 頁）。

## ワークシート1：アウトカムのチェックリスト

システマティック・レビューからのアウトカムのリスト			
システマティック・レビューのタイトル			
<p>エビデンス・プロファイルや SoF テーブルに含むべき重要なアウトカムは 7 個を上限として選択し、相対的重要性 (1～9 点) を参考にする。</p> <ul style="list-style-type: none"> <li>・ 1～3 点：意志決定に重要ではなく、エビデンス・プロファイルに含めない。</li> <li>・ 4～6 点：意志決定には重要だが重大ではない (エビデンス・プロファイルに含めるべきかどうかは、他の重要なアウトカムとの兼ね合いである)。</li> <li>・ 7～9 点：意思決定にとって重大で、必ずエビデンス・プロファイルに含めるべきである。</li> </ul>			
アウトカム	重要性	エビデンス・プロファイルに含めるか	
1.		はい	いいえ
2.		はい	いいえ
3.		はい	いいえ
4.		はい	いいえ
5.		はい	いいえ
6.		はい	いいえ
7.		はい	いいえ
8.		はい	いいえ
9.		はい	いいえ
10.		はい	いいえ
11.		はい	いいえ
12.		はい	いいえ

**ワークシート 2：各アウトカムについての研究全般に関するエビデンスの質評価**

\* もし、GRADE pro を利用しているならワークシート 2 はとばす  
 (本資料の 5 頁目、「エビデンスの質評価基準と定義」を参考)

エビデンスの質判定基準	グレーディング (各項目についていずれかの判定を 1 つ選択)	コメント (質を下げる理由を説明)	エビデンスの質 (アウトカムごとに 1 つ選択)
<b>第 1 アウトカム</b>			
研究の限界	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕⊕⊕⊕ 高
結果の非一貫性	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕⊕⊕○ 中
エビデンスの非直接性	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕⊕○○ 低
データの不精確さ	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕○○○ 非常に低
出版バイアス	なさそう ありそう (-1) 非常にありそう (-2)		
<b>第 2 アウトカム</b>			
研究の限界	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕⊕⊕⊕ 高
結果の非一貫性	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕⊕⊕○ 中
エビデンスの非直接性	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕⊕○○ 低
データの不精確さ	なし 深刻 (-1) 非常に深刻 (-2)		<input type="checkbox"/> ⊕○○○ 非常に低
出版バイアス	なさそう ありそう (-1) 非常にありそう (-2)		

(本来第 1 ~ 7 アウトカムまでのチェックリスト表があるが、紙面の都合で第 3 以後は割愛)

## GRADE エビデンスの質評価基準と定義

エビデンスの質 (Quality of evidence)	研究デザイン (Study design)	グレードを下げる (*) (Lower if)	グレードを上げる (*) (Higher if)
高 high	RCT	<ul style="list-style-type: none"> <li>研究の限界 (limitations)               <ul style="list-style-type: none"> <li>-1 深刻 (serious)</li> <li>-2 非常に深刻 (very serious)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>効果の程度が大きい (large magnitude of effect)               <ul style="list-style-type: none"> <li>+1 大きな効果 RR &gt; 2 あるいは &lt; 0.5</li> </ul> </li> </ul>
中 moderate			
低 low	観察研究	<ul style="list-style-type: none"> <li>結果の非一貫性 (Inconsistency)               <ul style="list-style-type: none"> <li>-1 深刻</li> <li>-2 非常に深刻</li> </ul> </li> <li>エビデンスの非直接性 (indirectness)               <ul style="list-style-type: none"> <li>-1 深刻</li> <li>-2 非常に深刻</li> </ul> </li> <li>データの不精確さ (imprecision)               <ul style="list-style-type: none"> <li>-1 深刻</li> <li>-2 非常に深刻</li> </ul> </li> <li>出版バイアス (publication bias)               <ul style="list-style-type: none"> <li>-1 ありそう (likely)</li> <li>-2 非常にありそう (very likely)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>+2 極めて大きい効果 RR &gt; 5 あるいは &lt; 0.2</li> </ul>
非常に低 very low			<ul style="list-style-type: none"> <li>用量-反応勾配 (dose-dependent gradient)               <ul style="list-style-type: none"> <li>+1 あり</li> </ul> </li> <li>すべての交絡因子 (plausible confounder)               <ul style="list-style-type: none"> <li>+1 提示された効果を減弱させている</li> </ul> </li> </ul>
* 1 = 1 段階グレードを下げる, あるいは上げる (たとえば, 「高」から「中」へ) 2 = 2 段階グレードを下げる, あるいは上げる (たとえば, 「高」から「低」へ)			
<b>アウトカムについての研究全般に関するエビデンスの質の定義</b> (Definition: Quality of evidence across studies for the outcome)			
<ul style="list-style-type: none"> <li>「高」 = 今後の研究によって効果推定値に対する確信性が変わる可能性は低い (very unlikely)。</li> <li>「中」 = 今後の研究によって効果推定値に対する確信性に重要な影響がおよぶ可能性が高く (likely), 推定値が変わる可能性がある (may)。</li> <li>「低」 = 今後の研究によって効果推定値に対する確信性に重要な影響がおよぶ可能性が非常に高く (very likely), 推定値が変わる可能性が高い (likely)。</li> <li>「非常に低」 = あらゆる効果推定値が不確実である (very uncertain)。</li> </ul>			

例：あるアウトカムに関する複数の RCTs からのシステマティック・レビューを想定すると、初めエビデンスの質は「高」から開始する。もし深刻な限界 (-1), さらに非常に深刻な不精確さ (-2) がある場合には、最終的なエビデンスの質は、「非常に低」となる。

## 附：エビデンスの質を下げる 5 要因 (RCTs)

### 研究の限界 (limitations)

- ・ 割り付けの隠蔽なし
- ・ 盲検化の欠如（特に、アウトカムが主観的でアウトカム評価バイアスの可能性が高い）
- ・ 追跡からの大きな脱落
- ・ ITT 解析を遵守していない
- ・ 利益が確認されたとして試験が早期中止となっている
- ・ 選択的なイベント報告：研究者が報告されているアウトカムを見逃す  
（典型的には、効果がないアウトカムのことが多い）

### データの非一貫性 (inconsistency)

異なる研究間で介入効果の推定値が大きく異なる場合（結果に異質性やばらつきがある場合）は、真の介入効果に差異が存在する可能性を意味する。異質性 (heterogeneity) が存在するにも関わらず、研究者が、それについて妥当な説明ができない場合は、結果における非一貫性の程度に従い、エビデンスの質を 1 ～ 2 段階下げる。非一貫性は、以下の項目における差異などによって生じる。

- ・ 集団（例：重症の集団の方が薬の効果が高い場合）
- ・ 介入（例：薬の用量が多いほど効果が大きい場合）
- ・ アウトカム（例：時間が経過するにつれて治療効果が減少する場合）

### エビデンスの非直接性 (indirectness)

非直接性には次の 3 種類がある。

- ・ 非直接的なアウトカム：たとえば、代理アウトカムなど
- ・ 非直接的な比較：介入 A と介入 B の比較ができないために、介入 A を介入 C と比較し、介入 B を介入 C と比較して間接的に比較する場合などがこれに当てはまる。このような研究では、介入 A と介入 B の効果の大きさを間接的に比較することができるが、介入 A と介入 B を直接つぎ合わせて比較した場合のものとは比べてエビデンスの直接性の程度は低くなる
- ・ 集団、介入、比較対照：ガイドラインパネルやシステマティック・レビューの著者が取り組む疑問が、集団、介入、比較対照、に関する入手可能なエビデンスと異なる

### データの不精確さ (imprecision)

研究に含まれる患者数やイベント数が比較的少なく、そのために効果推定値の信頼区間が広い場合には、結果は不精確であるとされる。

#### 1. 2 値アウトカムの場合

- ・ 合計（累計）サンプルサイズが、算出された最適情報量（optimal information size: 単一の試験におけるサンプルサイズの算出に相当する）よりも小さい
- ・ イベント発生件数の合計が 400 件未満、あるいは総サンプル数が 4,000 人未満（シミュレーションに基づく経

験則で、本来はベースラインリスクおよび効果サイズに依存する) である

- ・統合された推定効果または最良の推定効果を取り巻く95%信頼区間に、「効果なし」と、「相当の利益」または「相当の害」が、ともに含まれている。このエビデンスの質評価の根拠とされる「相当な利益」または「相当な害」について、GRADE では、相対リスク減少 (RRR) または相対リスク増加 (RRI) が 25% を上回るという閾値を提案している

#### 例外

イベント発生率が非常に低い場合、相対効果を取り巻く95%CI が非常に広くても、絶対効果を取り巻く95%CI は狭い場合がある。こういった場合、不精確さを理由にエビデンスの質を下げることはしない。

#### 2. 連続アウトカムの場合

- ・95%CI に「効果なし」が含まれ、なおかつ上限信頼限界または下限信頼限界が利益または害の最小差異 (MID: 意味のある介入効果の最小の差) をまたいでいる
- ・MID がわからない場合、あるいは異なるアウトカム指標を使用した場合で効果サイズ (ES) の計算が必要となったときは、上限信頼限界または下限信頼限界がいずれかの方向に 0.5 の効果サイズを超えた場合は、評価を下げるおよびだろう

#### 出版バイアス (publication bias)

仮説を肯定する有意差がある結果が出た肯定的研究 (positive study) のみが論文出版され、そうでない否定的研究 (negative study) は出版されない傾向があること。出版バイアスがありそうな場合、エビデンスの質を下げる。



## ワークシート4：ガイドラインパネルが検討する推奨草案

推奨草案		
推奨に関連する価値観や好み (検討した各アウトカム別に、一連の価値観を想定する)		
(例：この推奨の作成に当たっては、致死率の高い疾患における死亡の回避を重要視した)		
重大なアウトカム全般に関する全体的なエビデンスの質		
推奨度の判定 (以下の4項目について判定し、その説明を記載する)		
推奨度を弱くする要因	判定	説明
アウトカム全般に関する全体的なエビデンスの質が低い (全体的なエビデンスの質が高いほど推奨度は「強い」とされる可能性が高くなり、逆に全体的なエビデンスの質が低いほど、推奨度は「弱い」とされる可能性が高くなる)	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
利益と害・負担のバランスが不確実 (コストは含まず) (望ましい効果と望ましくない効果の差が大きければ大きいほど、推奨度が強くなる可能性が高い。正味の利益が小さければ小さいほど、利益の確実性が減じられ、推奨度が「弱い」とされる可能性が高くなる)	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
患者の価値観や好みの不確実さ、あるいは相違 (価値観や好みにばらつきがあればあるほど、または価値観や好みにおける不確実性が大きければ大きいほど、推奨度が「弱い」とされる可能性が高くなる)	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	
正味の利益がコストや資源に見合ったものかどうか不確実 (介入のコストが高ければ高いほど、すなわち消費されるリソースが多ければ多いほど、推奨度が「強い」とされる可能性が低くなる)	<input type="checkbox"/> はい <input type="checkbox"/> いいえ	

「はい」の回答が多いと、推奨度が「弱い」とされる可能性が高くなる。

討議によって合意に至らなかった場合、推奨を策定するパネルは、以下の表 (GRADE grid) を活用することによって、入手可能なエビデンス、利益と不利益、価値観と好み、そしてコストの分析結果に基づき、ある特定の介入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。この評価結果は各介入の使用・非使用を示す推奨度にマッピングされる。表の各カテゴリの推奨に対する投票数を入力すること（附録 4.2, 「推奨度評価の合意形成 (GRADE grid)」, 130 頁参照）。

GRADE grid による合意（附表 4.2-1 再掲）

GRADE の推奨度	強い	弱い	弱い	強い
介入の望ましい効果と望ましくない効果に関する評価者の見解	望ましい効果が望ましくない効果を明らかに上回る	望ましい効果が望ましくない効果をおそらく上回る	望ましくない効果が望ましい効果をおそらく上回る	望ましくない効果が望ましい効果を明らかに上回る
推奨	“実施する” ことを推奨する	“実施する” ことを提案する	“実施しない” ことを提案する	“実施しない” ことを推奨する
推奨度				
上記の 4 つのカテゴリから 1 つを選択する				
最終的な推奨				
全パネルメンバーからの意見をまとめ最終的に決定する				
注意				