

## 第56章 実践のエビデンスの要約方法

### 背景

米国厚生省の傘下団体である医療研究・品質局（The Agency for Health Care Research and Quality：以下 AHRQ）は、この報告書において評価、確認された患者安全の実践についてランク付けもしくは格付けを行う職務を UCSF-Stanford Evidence-based Practice Center に委託した。この報告書は、一般的なトピック（パート 3、セクション A-H）に分類されるものから、個別の実践として章の中で区分されるものまで、広範でかつ様々な患者安全の実践を扱った集大成である。情報をまとめることは困難であったが、読者や治療方針を決定する者がその実践を導入するか、もしくはさらに研究を進めるかを判断するためにはこのランク付けは重要である。これらの二つのタイプの読者を念頭におき、我々は“ランク付け”について以下の3つの目標を設定した。

- ・ 患者安全の介入について実践を是認したり、さらなる研究に資金を提供するかどうかの意思決定を行う者に対して、入手可能な文献の範囲内で、できるだけ実利的な情報を提供し、実践の主要な要素をランク付けするための枠組みを策定する。
- ・ 報告書に今後の研究課題を示すために、不均一性が高い分野で実行可能な実践を比較する場合には、ランク付けの限界を明記する。
- ・ 注意書きにランク付けを過度に簡略化したことを明確に示すために、ランク付けの結果は、整理がされ、見た目に分かりやすく、そして理解しやすいやり方で伝える。

最終目的は、それぞれの実践を緻密に格付けされたスケールで測る能力を示すことではなく、様々な局面について、エビデンスに基づいて、実践に重きを置くことである。これらを適切に比較する（費用効果分析）には、現時点で文献で入手できるものだけでなく、より多くのデータを要する。

### 実践のランク付けのためのデータ入力

実践のランク付けで述べているように、おのこの実践について、最終的な蓄積に対する様々なインプットに関する情報は、先を見越して決定されている。どのような情報を集めるかは、実践を要約した表を使用することでその可能性が見えてくるようにした。ランク付けを周知するため、実践を3つのカテゴリーに分けた。

- ・ *実践の予想される影響*：対象となる患者の安全性の重大性と頻度に基づき、現在の実践

の実施状況に関する情報を示す。

- ・ *実践を支持するエビデンスの強さ*：エビデンスの相対的重みや、効果の大きさを評価した情報、そして付随的に負の効果が起きないように実践の実施の際に注意を要する情報を含む。
- ・ *実施*：コスト、実践遂行上の障害、政策課題について考慮した情報。

3つのカテゴリーについての説明は順序だてておこなわれている。実践を行うことで予想される影響をカテゴリーに分けるため、各々の実践について、対象となる安全問題の頻度と重大性が報告された。次に、*現在の実施状況*についてエビデンスのレビューを行って、実践の潜在的な影響を評価した。より広範囲に実行可能な実践がその対象とする有害事象をさらに減少させるかどうかと言うと、介入が既に広く行われている場合だと、現在あまり行われていない実践よりも改善の余地が少ない。それゆえ、実践の潜在的な影響の大きさは、対象となる安全問題の頻度と重大性（医療過誤）と現在の実践の実施状況によって規定される。

もちろん、*エビデンスによって支持されている実践*に関連する要因が評価されれば、実践の実際の影響を評価することができる。本報告書は患者安全の実践の集大成となっているので、関連する研究の要約を行い、さらに研究デザインとアウトカムに関する詳細なデータ要素をまとめることを各章の著者に指示した。（3章参照。）全般的には、この情報はそれぞれの実践について*研究の強さを全体的に評価する際に用いられる*。効果（*effectiveness*）は、ルーチンとして実践を行った場合の正味の肯定的な効果であると共通して定義されている。しばしば研究で報告されている効力（*efficacy*）に関するデータは、介入群と対象群に分け実験的な状況下における正味の肯定的な効果である。直接的なエビデンスがない場合、効力を効果と言い換えることは一筋縄ではいかず、それゆえ実施された特定の研究の一般化可能性に関する判断に基づく。ランク付けを行うために研究から集約されたもののうち、最も重要なのは*介入効果の大きさ*である。

最後に、エビデンスに基づいたレビューでは、医学的介入の潜在的な害について考察を行い、著者には実践で得られた共通認識や文献から収集された関連するエビデンスのみならず詳細な理由をつけた関連事項も報告するよう求めた。

実社会の環境と患者安全分野の動きに対するの社会の要望に対応するために、実践についての章では、コストと他の予想される*実行面*での障壁について情報を示した。そこでは著者は、コスト削減を論じたり、費用効果分析を報告する一方で、焦点は現在実行中の実践の*立ち上げコストと年間支出を焦点とした*。初期費用や実践の継続のための、実践を導入しようとする各々の環境（たとえば医療ネットワークもしくは施設の規模）によって大

きく左右されるが、予想されるコスト削減は実践を導入するところの状況（例えば、患者安全の対象となる問題の頻度）にむしろ制約されているようだ。主な投資の決定については、トレードオフの評価は局所的なレベルでより適切である。我々の意図するところは、単に初期費用と経常コストの“概算の”見積もりを報告することである。特定の実践の実施における経済学的影響と分けて考えられるのは、*政策的・技術的考察*である。

実践のランク付けにこれら全てのデータを投入することについては、第一の目標は、出版物や他の情報源から入手可能な最良のエビデンスを見つけることである。これらの領域それぞれについて同じように考察した文献は、これまでのところまとめられてないため、さらに研究が行われれば概算の数値はより改善するであろう。可能な場合は、これら全ての要素について要約する際に、我々は限られたデータを基に主要な部分についての評価を試みた。

## 評価プロセス

患者の安全性の確保、実践の説明、一般的なランク付けの分類（高・中・低）（いくつかの要素については前述の通り）についての評価書式の策定は4人の編集者が行った。評価の段階では個人的な判断がされているが、それぞれの分類には一般的な経験則が明記されている。この書式では、評価を行う者が評価を行う際に気になった点や個々の判断を記録できるようコメント欄が設けてある。章ごとにおのおのの編集者が個人でエビデンスの存在する実践についての評価を行った。編集者らは3日間集まり、お互いの採点を比較し、採点の相異を話し合いを行った結果、分類によるランク付けと要約のランク付け - あるいはレビューを行った実践について合意を得ることができた。

## 意思決定の基準と判断を行う際の考察についての詳細

### *実践の潜在的影響の大きさ*

上記に関しては、患者安全の対象となる問題や現在実施されている実践の頻度と重大性の評価を考慮して、潜在的な影響の評価を行う。編集者らは各章のデータや臨床的知識を使って、潜在的な影響を<高><中><低>もしくは<情報が不十分>の順にランク付けした。その実践が<高>と認められるには、入院患者の1%以上の患者集団（年間で約30万人）を対象としている場合や、もしくは、対象となる患者安全の問題が死亡や障害につながる恐れがある場合である。<低>とされる実践は、効果的な実践が行われないことで入院患者の0.01%以下の患者集団（年間で約3千人）が予防可能な有害影響を被るとされる場合である。患者安全の対象が<高>と<低>のカテゴリーの間に該当する実践は、潜在的な影響は<中>程度であると判断した。

頻度と重大性を基に主たる評価が実施された後、追加された決定の基準は影響の大きさの評価に適用される。実践が現在でも広く実施されている（75%以上の病院）なら、その時は評価を一段階下げる（高から中へ、もしくは中から低へ）。こういう状況が起こった場合は、予想される影響のレベルについては現在の実施状況がすでに広範で行われていると判断したためである旨を注記するようにした。

<情報が不十分>という分類は、頻度と重大性に関する情報が限られている場合や、患者安全の対象となるべき問題の定義がきちんとなされていない場合である。

### *実践を支持するエビデンス*

潜在的な害に起因する研究の強さ、対象に対する効果の大きさ、そして警戒レベルの高さは、前述の決定基準というよりは判断に基づいてランク付けされている。どのケースにおいても評価を行う者はランクを選択した理由を記すことにした。

研究の強さについては、研究デザインとアウトカムレベル（階層については、3章を参照。）研究の数、研究対象となった患者数、一般化の可能性、そして他の方法論的問題などが、個々の実践における研究の相対的強さを測る上で考慮すべき因子となっている。研究の強さは<強><中><弱>に分類した。研究の強さは実際の研究成果を考慮せずに対象についての効果の大きさだけで評価した。実践についてエビデンスが全くない、もしくは最小限しかない場合は、研究の強さは<弱>と評価され、研究結果に過度の「信用」を与えないように、評価者は実践を支持するエビデンスの他の二つの要素は評価しなかった。

対象に対する効果の大きさの評価は、エビデンスの効果に関するレビューの対象となった研究に報告されている相対リスクの減少もしくはオッズ比を基にしている。評価は実践の章で扱われている研究結果のみを利用し、追加分析（メタ分析）は加えなかった。全研究のうち、もしくは報告されている多数の研究のうち、圧倒的多数が肯定的であり、効果の大きさが大きい場合には（相対リスク比の減少が15%以上）予想される効果の大きさを<頑健>と分類した。明らかに効果がない場合や、ごく最小限の効果しかない場合（相対リスク比の減少が5%以下）は、予想される影響の大きさは<わずか>と評価した。実質的な影響が示唆されているが、明らかに<頑健>ではないと考えられる研究結果については、<適度>という分類を用いた。最後の分類である<不明>は、効果の大きさの結果が一貫していない実践について用いた。

ある有害事象を減少させるという結果が示された実践についても、実践が実際に実行された際には、新たな問題が引き続いて発生することが予想される。従って、実践が実際に実行された場合にその実践が否定的な結果に終わらないよう、我々は警戒レベルの高さを

基に有害性への関心度合いを主観的に評価し、その分類を、〈低〉〈中〉〈高〉とした。〈低〉とされる実践は、潜在的な有害性に対して注意を全く必要としない、もしくは若干の注意を必要とする場合とし、〈高〉とされるものは潜在的な否定的影響についてモニターを強化するに値するとされるものである。疑わしい場合は警戒レベルの高い分類を選択するように、これらの評価は控えめに行った。

### 実行

実際に実施するために3年間必要だと見積もって、患者安全実践をコストと複雑さについて評価を行った。コストは標準的な規模の病院や医療機関において全面的に実施した場合の初期の立ち上げに要する費用と年間の支出額に基づいている。評価の際には予想されるコスト削減は考慮しなかったが、文献に記されている場合には実践について扱った章で報告した。実践の実施に100万ドル以上の経費がかかることが予想される場合は、〈高〉とした。概算の支出が10万ドルから100万ドルのものは、〈中〉と分類した。このレベル以下のものは、コストの点に関しては〈低〉と評価した。

実践の実現可能性については、予想される政策変更(だれが医療ケアを提供するかという大きな転換)、技術面での障害(完全に捨て去ることのできない古いシステムの集積、日々最新のコンピューターシステムが普及していくこと)を考慮して評価した。実践の複雑性について評価したデータは比較的少数しか存在しないので、我々は二つの分類のみ使用した。〈低〉と〈高〉、すなわち比較的容易であることと比較的困難であることを意味する。単に金銭を費やせば実施できる場合は、コストは〈高〉とし、実現の可能性は〈低〉とした。

### 影響・エビデンスの全体評価

加えて、実践を支持するエビデンスや予想される影響についての情報を0から10で(エビデンスの強さ)編集者らがそれぞれ採点を行い、総合的に考慮した。この評価については、大規模な医療法人(病院や総合医療システム)のリーダーという視点を盛り込み、「施設が今後三年間で患者安全の向上を望んでいるとして、資源についてはほとんど配慮されていない場合、この実践をどのようにランク付けするか?」と尋ねた。この評価は、我々は実践の実施の難しさやコストに関してはきちんと考慮しなかった。むしろ、評価は単にエビデンスの強さ、具体的には医療ケアに関連する有害事象の減少に対する安全実践の効果と予想される影響についてエビデンスの強さを反映させた。患者安全の対象となる問題について、影響が〈大〉とする場合は、特定の実践が否定的な結果(院内感染)を有意に減少させる(影響の大きさが〈高〉)とする有力なエビデンスがあつて(相対的研究の

強さが<強>) 評価者が実践をほぼ 10 点と採点する場合である。研究が説得力に欠け、効果の大きさも取るに足りないものであるか、潜在的な有害性のために警戒を要するレベルが<中>や<高>とされる場合は評価を下げた。

#### 研究の優先度に関する全体評価

同じように、我々はそれぞれの実践についてさらに研究を行う有用性、特に研究プログラムが上手に取り組みられる可能性があるかどうか(研究の優先度)を重視して評価した。これは、具体例としては「もしあなたが患者安全の向上を義務付けようとする大きな機関や財団のリーダーだとしたら、今後新たに実施する研究を推進するために研究費の分配を検討しているとしたら、この実践をどうランク付けするか?」という質問である。調査によって対処可能な単純な隔たりがエビデンスにあったり、あるいは実践が多面的であったり、効果的な要素を決定することで実施が容易になるような場合は、研究の優先度は<高>とした。予想される影響が大きいとされる領域の 1 つであり(有害事象による死亡や疾病リスクがあるとされる患者数が多数であること) 実践について研究が十分されていない場合も、研究の必要性は比較的高い評価をした。見込みがほとんどない場合(実践によって対処する安全性の問題が比較的少数の患者にしか影響しない、もしくは有意な知見ではあるが実践が実用性に欠くと示される場合)は、そのような実践に対する研究の必然性(優先度)は低く評価される。逆に、明らかに有効な実践で、低コストでかつ、実施が容易なものは、さらに研究する必要はないので、優先度は低くなる。

#### **評価の注意点**

すべての要素に対する評価については、評価者である 4 人の編集者の間で評価が一致しないことは稀であり、合意が得られるまで話し合った。しかし、それぞれの最終的な分類では、解釈に食い違いがあり、議論の後でも食い違いが続く時には、これらの食い違いについてのコメントをプロトコルに記載するようにした。(59 章参照。)特定の追加情報によってその評価の正確さに懸念がでることが明らかな場合にもコメントを記した。二、三の事例では、特定していない分類は特別な状況の場合に使用し、そのような時は分類を説明するコメントを再度記載した。

#### **ランク付けの表**

##### 情報収集

ランク付けはデータ表に記録し、コメントは脚注に記した。過去の議論を要約しながら、57 章から 59 章に示されている様々なデータ表は、以下の 11 の要素のいくつか、もしくは

全てが含まれ、ランク付けしカテゴリーごとに、強いものから弱いものへと並べた。

1. 章の番号
2. 患者安全の対象
3. 患者安全の実践の説明
4. 予想される影響：高、中、低、情報が不十分
5. 研究の強さ：強、中、弱
6. 効果の大きさ：頑強、適度、わずか、不明
7. 警戒レベル：低、中、高
8. 実施コスト：低、中、高
9. 実施の複雑性（政策面、技術面）：低、高
10. 影響 / エビデンスの総合評価：0 - 10（10 が最も高い）、0.5 ごとの評価
11. 研究の必要性の総合評価：0 - 10（10 が最も高い）は 0.5 ごとの評価

#### 報告された情報

上記の 1-9 の要素を含む各分類の詳細なデータ表を 59 章に示す。総合的な評価について具体的なスコアを報告することは、スコアリングが精緻であることを意味する。可能な場合は、分類に関する注意事項を脚注に示した。

エビデンスの強さと研究の優先度の両方の評価について、我々の目的とする所は、1-10 のスコアを正確に報告することではなく、一般的な「区分」や実践のグループ分けを發展させることである。最初に記したように、実践に関する研究情報が利用できる場合には相対評価する際により良い方法を利用することができる。エビデンスが相対的に不足している場合にはより正確で、精巧ではあるが結局のところ実行不可能なアプローチを利用しなかった。

57 章では、影響と有効性のスコアについて「エビデンスの強さ」の総合評価を要約し、実践を 5 つの区分に細分化した。実践はそれぞれのランク付け区分で、高いスコアから低いスコアの順に並べた。ランクは、<強い>（8-10 のスコア）、<やや強い>（6-7.5 のスコア）、<中>（4-5.5 のスコア）、<やや弱い影響 / エビデンスの実践>（2-3.5 のスコア）、<弱い影響 / エビデンスの実践>（0-1.5 のスコア）とした。あるランクの下限に近い実践は、隣接した下の区分の上部近くにくることが適当であるようにした。同じように、区分の上限にある実践は、隣接した上の区分と実際に比較できるようにした。区分間の境界はいくぶん人為的ではあるが、実践の効果を支持（あるいは否定）するエビデンスや影響について、一般的で合理的なデータ合成を可能にするものである。読者は、最も高い区分の実践は最も低い区分に属しているわけではないと確信してよい。

58 章では、「研究の優先度」のスコアについてその総合評価を要約し、役に立つであろう研究タイプの例を示した。実践は次の 3 つの分類に分けた：＜さらに研究を行うことで高い便益をもつ＞（7 以上のスコア）、＜さらに研究を行うことで便益をもつ＞（4-6.5 のスコア）、＜研究の優先度は低い＞（4 以下のスコア）。研究の優先度の上位 2 分類に含まれる実践はリストに示した。