

ve/30366

4) Pearson SD, Rawlins MD, Quality, innovation, and value for money: NICE and the British National Health Service. *JAMA* 23;294(20):2618-22, 2005.

5) Guidelines Manual-Appraising Orphan Drugs, NICE, <http://www.nice.org.uk>, 2006.

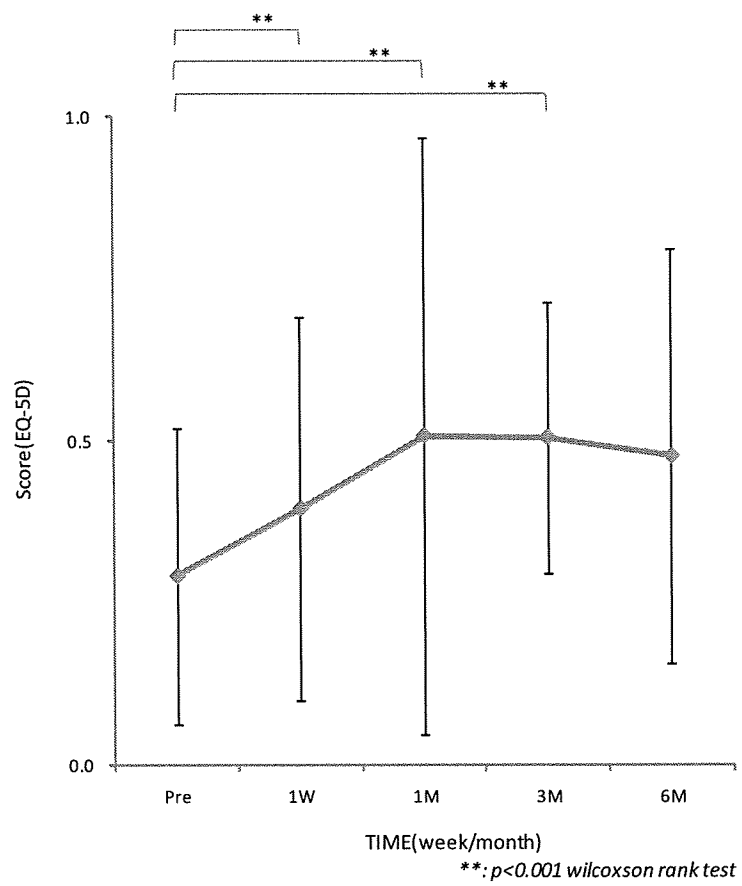


図 1. PVP の介入後の EQ-5D スコアの推移

表 1. PVP の介入に係わる健康関連 QOL 間の相関分析

Time: Pre				
<i>r/p</i>	PCS-8	MCS-8	EQ-5D	VAS
PCS-8	-	**		
MCS-8	-0.605	-		
EQ-5D	-0.279	-0.188	-	*
VAS	0.078	0.207	-0.410	-

Time: 1W				
<i>r/p</i>	PCS-8	MCS-8	EQ-5D	VAS
PCS-8	-			
MCS-8	-0.293	-	*	*
EQ-5D	-0.221	-0.435	-	*
VAS	0.199	0.363	-0.399	-

Time: 1M				
<i>r/p</i>	PCS-8	MCS-8	EQ-5D	VAS
PCS-8	-	**	**	*
MCS-8	-0.551	-	**	
EQ-5D	-0.828	0.710	-	
VAS	0.486	0.140	-0.291	-

Time: 3M				
<i>r/p</i>	PCS-8	MCS-8	EQ-5D	VAS
PCS-8	-	*	**	**
MCS-8	0.448	-	*	**
EQ-5D	-0.624	-0.491	-	**
VAS	0.702	0.668	-0.594	-

Time: 6M				
<i>r/p</i>	PCS-8	MCS-8	EQ-5D	VAS
PCS-8	-		*	*
MCS-8	0.377	-	*	
EQ-5D	-0.784	-0.787	-	
VAS	0.726	0.613	-0.649	-

*: $p < 0.01$, **: $p < 0.001$ Simple Correlation

「HTA Core Model for Diagnostic Technologies, 1.0R, Work Package 4」からの一部抜粋

Clinical effectiveness (pp.85-87)

「臨床上的有効性」

Iris Pasternack, Tuija Ikonen, Sigurdur Helgason, Sami Kajander, Heikki Ukkonen, Marjukka Mäkelä

Methodology

方法

診断技術の有効性を評価するために最も理想的な調査デザインとは、無作為化臨床試験であるが、診断技術を対象とした無作為化臨床試験は稀にしか行われていないのが現状である。しかし診断技術の有効性を直接的に測定した臨床試験が行われていないとしても、診断技術の安全性や診断精度、患者管理、治療への影響力などに関するデータが存在すれば、それらのデータを組み合わせて臨床現場における有効性を間接的に算出することができる。注意すべきなのは、複数の調査データを組み合わせる場合、それらの調査が果たして同一の疾患を対象としていると見なせるのか、という論点である。

Optimal study types

最適な研究デザイン

1. Direct trial evidence

直接的な試験エビデンス

最も信頼度が高いのは、無作為化臨床試験（RCT）である。母集団となる患者を、既存の診断技術を用いるグループと新しい診断技術を用いるグループとへと無作為的に分け、健康状態の相違を評価する方法である。他の比較研究、例えばコーホート研究や症例対照研究は、調査バイアスが生じる可能性が高いことに留意すべきである。

2. Linked evidence

間接的なエビデンス

有効性を直接的に測定した研究が存在しないのであれば、他の要素を測定している研究を組み合わせて、有効性を間接的に算出することができる。以下の図で、調査要素ごとの推奨される研究デザインを示しておく。

調査要素	推奨の研究デザイン
安全性	特になし
診断精度	コホート分析
患者管理	前後比較試験
治療への影響力	無作為化臨床試験

診断精度に関する研究ではまず、対象としている患者や疾患、診断技術、その他の条件が、診断精度や治療の有効性という観点から同一とみなせるか、という点をチェックする。場合によっては、診断精度データのみで診断技術の有効性を算出できることがある。即ち、新たな診断技術が、コストも低く、安全性、特異度が向上している場合である。

患者管理前後比較研究とは、新たな診断技術の導入が、患者管理の効率の向上に与える影響を測定するものである。この要素が重要になるのは特に、追加的な診断技術の導入の場面¹である。

代替的な診断技術導入の場面では、基本的に診断テストの結果は患者管理には影響を及ぼさないと仮定が置かれている。しかし他の場合では、診断テストの結果が、患者管理に関する医師の意思決定に影響が存在することが示されねばならない²。

また患者管理研究が必要となるのは、治療に関する意思決定に、診断テストの結果以外の要素、例えば患者個人の特性や選好が、影響を及ぼす場合である。さらに、診断テストの影響力が不確かな場合、例えば複数の鑑別診断結果を弁別する診断テストや、有病率や疾患の重症度が異なる患者を対象とした診断精度研究を行った場合にも患者管理研究は重要となる。

新しい診断技術がメリットとデメリットの双方をもたらす場合、例えば侵襲性の低下による安全性の向上は見込まれるが、特異度の低下が伴う場合、擬陽性に起因するデメリットを算出するため、意思決定分析が有用である。また、有病率の相違による診断テストの有効性の比較にも、意思決定分析は有用である。ただし、意思決定分析を用いることができるのは、診断精度分析が治療効果分析と等値できる場合に限定されていることに留意すべきである。診断精度の向上が治療効果の向上に結びついていない場合には、無作為化臨床試験を行わなければならない³。

Quality Assessment 質の評価

1. Direct trial evidence 直接的な試験エビデンス

¹ Guyatt GH, Tugwell PX, Feeny DH, Drummond MF, Haynes RB. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J.Chronic Dis.* 1986;39(4):295-304.

² Medical Services Advisory Committee. *Guidelines for the assessment of diagnostic technologies.* August 2005

³ Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *The Lancet*, 2000 11/25;356(9244):1844-1847.

診断テストの有効性評価研究のバイアスは、対象とした患者のサンプリング過程、患者に対するケアの偏り、調査結果の測定・解釈、追跡不能者の偏りに起因する⁴。

2. Linked evidence

間接的なエビデンス

安全性評価と診断精度、治療に対する有効性といった個別の要素に関しては、本報告書の該当する部分に記載してある。

間接的な測定手法が無作為化臨床試験に対して持つメリットとデメリットに関しては、MSAC⁵の69頁にチェックリストを記載されている。

前後比較患者管理研究については、症例シリーズ研究と同様の基準（MSAC70頁を参照のこと）で評価すればよい。バイアス要因は対象となる患者の選択、診断テスト実行の客観性、テスト結果の評価である。

Costs and Economic evaluation (pp.93-97)

「費用と経済的評価」

Methodology 方法

以下の表1は、HTAを構成する4種の経済学的評価軸をまとめたものである。それぞれの評価軸は、健康状態の測定手法、及びにその評価という側面で区分されている。評価の目的をどこに置くのか、どのようなデータが手元に存在するのか、具体的にどのようなコンテキストでその評価を行うのか、という観点に応じて、これら4種の評価軸を使い分けるべきである。

⁴ Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998 Aug 22;352(9128):609-613.; Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995 Feb 1;273(5):408-412

⁵ Medical Services Advisory Committee. Guidelines for the assessment of diagnostic technologies. August 2005.

表 1 HTA を構成する 4 種の経済学的評価軸

評価軸の種類	その評価軸が適切な場合	コストの評価単位	医療アウトカムの評価単位	対応する調査課題
コスト最小化分析	対象となる診断技術の効果性に差がない;コストに関するデータが十分	通貨単位	なし	コストが最小となる介入技術の特定
コスト効果分析	対象となる診断技術の効果性が異なる	通貨単位	自然単位(例えば余命や障害日数, 血圧)	介入結果を1単位改善するため投入すべきコストの算出
コスト効用分析	健康関連QOLが医療アウトカムの重要な構成要素, かつ/または, 異なる医療部門間の比較をする場合	通貨単位	QOL調整済み生存年数, 健康年数相応値	介入結果を1単位改善するため投入すべきコストの算出
コスト便益分析	健康指標以外からの効果にも関心が有る場合(例えば, 治療プロセスそのものや, 情報の効果), もしくは特定の診断技術のみを評価する場合(正味便益値), もしくは個人の生命を通貨単位で換算する場合, もしくは社会の複数のセクター間での比較を行う場合	通貨単位	通貨単位	複数の調査対象が社会に及ぼすコストと便益の算出

経済学的な分析の基本的な考え方とは、ヘルスケアにおける資源配置の最適化、即ち、所与の資源の下で集団の健康状態を極大化する、という発想である。

Perspective 指標の評価観点

どのようなコスト、どのようなアウトカムを分析に組み込むべきなのか、という問題は、いかなる観点から経済学的な評価を行うのか、という問題と密接に関連している。例えば、病院という観点から考えるならば、入院期間の短縮はコストの削減に繋がるかもしれないが、在宅医療のコストまでを勘案するならば、社会的な観点から考えて、コストが余計に増大する可能性も存在するのである。

理想的にはなるべく広汎な観点から経済学的な評価を行うべきである。最も包括的に考えるならば、社会全体の観点から、その診断技術に伴うあらゆるコストとアウトカムを特定・測定・評価すべきである⁶。評価の範囲を絞り込んでゆくのであれば、ヘルスケア・セクターの観点や、第三者からの観点、病院の観点、患者の観点、といったものが考えられるだろう。最終的にどの観点をを用いるのかは、経済学的な評価の目的によって異なってくる。社会全体の資源配分が問題ならば、社会全体の観点が必要になり、病院の医療技術評価ならば、病院の観点を選ぶのが当然である。

Cost 費用

コスト評価の手続きは3つに大別できる。即ち、コスト要因の同定、コストの算出、資源使用の評価である。まずはどのような資源が用いられているのかを同定しなければならない。その次に具体的に投入されている資源の量の測定が行われる。最後にその投入量の評価が必要である。

直接的なコストとは、疾患や医療技術に直接関連しているあらゆるコストを指す。直接的なコストの内訳として、ヘルスケア・セクター内部のコスト（例えば、薬剤、機材、スタッフなど）やヘルスケア・セクター外部のコスト（例えば、通院に要する時間）が存在する。

間接的なコストとしては、例えば疾患に起因する患者の失われた労働時間、作業能力の低下分、死亡による失われた所得、などが存在する。失われた所得を算出する手法としては、人的資本モデルや摩擦コストモデルが存在する。失われた所得という間接的なコストは独立に分析されることが多いが、コスト効果分析やコスト効用値分析にもそのコストは組み込んで分析すべきである。また失われた所得という概念を、疾病手当のような社会保障給付金と混同してはいけない。

使用資源量は、通貨価値に換算した使用資源コストとは別個に報告すべきである。そうでなければ、異なる条件下でそのデータが活用できなくなってしまうからである。付け加えると、直接的なコストと間接的なコストとは区分して報告することが望まれる。

Outcome アウトカム

⁶ Drummond M.F, Sculpher M.J, Torrance G.W., O'Brien B.J., Stoddart G.L, *Methods for the Economic Evaluation of Health Care Programmes. Third Edition.* Oxford University Press, 2005.; Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: Recommendations for the design, analysis, and reporting of studies. *International Journal of Technology Assessment in Health Care* 2005;21: 165-71

介入による健康アウトカムの測定単位には、自然単位（死亡、余命）、健康状態値や効用値、通貨単位が存在する。

介入による効果が余命の延長のみならず、QOLの向上にも繋がるのであれば、複合アウトカム指標（例えば、QOL調整済み余命年数(QALYs)や健康年数相応値(HYEs)）を用いるべきである。複合アウトカム指標を用いた分析は、医療政策分析や医療プログラム意思決定に有用である。なぜなら、複合アウトカム指標は汎用性が高く、複数の介入技術や疾患間の比較が可能となるからである。また、対象集団や疾患、介入の相違にも原理的には関係なく用いることができる指標なので、同じ方法を用いている研究であれば、複数のプログラム間の比較研究も可能である。健康関連 QOL 指標(HRQOL)とは、健康に関連した QOL のことである。健康関連 QOL 指標を算出するには複数の手法があるが、いまのところ、決定的と見なされる手法はまだ登場していない。患者アウトカム指標とは、死亡率や疾病率といった旧来の指標とは異なり、個人の健康状態に関連する身体的・社会的・感情的側面をも含んだ指標である。患者アウトカム指標を算出するためにも、疾患ごとに提案されている手法や汎用的な手法、効用値に注目した手法など、複数の手法が存在している。ともあれ経済学的な評価に組み入れるためには、何らかの手続きにより数値化する必要があるため、なるべく汎用性の高い手法を今後は用いるようにすべきである。

コスト有効性増分率(ICER)

どの医療技術がコストに比して有効性が高いのか、という問いに結論を出すためには、コストと有効性の総量を、少なくとも2つの介入に関して比較しなければならない。比較結果としては、以下の表の9類型が存在する。

類型1や類型9においては、どちらの技術が望ましいのか、追加の分析が必要である。そのためには、アウトカムの相違に対する介入コストの比であるコスト有効性増分率の算出を行わなければならない。コスト有効性増分率とは、異なる介入技術が、1単位分の健康アウトカムの改善に要するコストのことである。社会全体が追加の便益のためにコストを負担しようとする（類型9）、もしくは、社会全体が、有効性の低下が生じたとしてもコスト削減が重要だと判断するのであれば（類型1）、新しい介入技術はコスト効果的だと言える。

表2 コスト有効性分析結果の9類型

従来のもとの比較して新しい技術は	有効性低下	同等の有効性	有効性の向上
コスト削減	1. この段階では不明. 追加の分析が必要	4. 新しい技術を採用すべき(低い優位性)	7. 新しい技術を採用すべき(高い優位性)
同等のコスト	2. 従来技術を継続すべき(低い優位性)	5. 両者に差はなし	8. 新しい技術を採用すべき(低い優位性)
コスト増大	3. 従来技術を継続すべき(高い優位性)	6. 従来技術を継続すべき(低い優位性)	9. この段階では不明. 追加の分析が必要

Threshold cost-effectiveness and net benefit approach

コスト効果性閾値と正味便益分析

介入のコスト効果性の判断は、アウトカム1単位当たりにつき、最大でどれだけのコスト負担を行うのか、という観点に左右される。この観点とは即ち、コスト効果性閾値のことである。介入のコスト効果性が閾値より低ければ、その介入はコスト効果的（例えば、ヘルスケアの効率性を向上させる）であると判断され、逆に介入のコスト効果性が閾値より高ければ、その介入はコスト効果的だとは判断されず、この介入に対する資源配分は、ヘルスケアの効率性を向上させることにはならない。

確かにコスト効果性増分率は広く用いられている分析手法であるが、増分率の数値そのものは、検討対象となっている介入の効果性の度合いには無関係の指標であることに注意すべきである。具体的な量を勘案した評価を行いたいのであれば、正味便益分析が有用である。コスト有効性分析の難点を解決するには、正味通貨便益(NMB)や正味健康便益(NHB)といった指標が有効⁷である。

Modeling モデル化

経済学的なデータや臨床データが手元に存在しない場合、もしくは短期的な臨床データから長期的な臨床データを外挿しなければならない場合、モデルを用いた経済学的評価が必要となる。このような場合に広く用いられているモデルとして、決定木解析モデルやマルコフモデルが存在する。

⁷ Drummond et al., 2005; Drummond et al., 2005.

Sensitivity analysis 感度分析

経済学的な評価は多くの場合、統計学的な手法に基づいて算出された推定値を用いている。パラメータそのものの不確実性は別にしても、経済学的な評価、特に経済学的なモデル分析は、複数のパラメータ間の、必ずしも確実とはいえない関係を前提としていることに留意すべきである。従って、経済学的な評価には、パラメータ間関係の不確実性を考慮に入れるべきなのである。つまり感度分析により、経済学的分析によって得られた結果の信頼性を知ることができるのである。経済学的評価を行う場合には、決定論的分析であれ、確率論的分析であれ、必ず感度分析を伴って行うべき⁸である。特に、評価に用いる経済学的モデルに関し、不確実性を伴っている個々の入力変数が分析結果に及ぼす影響を特定するためには、必ず十分なだけの感度分析を行わなければならない。特定の入力変数につき感度分析を行わないのであれば、その理由を必ず明示しておかなければならない。不確実性を把握するための手法については、Briggs et al 1994, Briggs et al 2006⁹を参照のこと。

Discounting 割引率

未来のある時点でのコストやアウトカムに関して経済学的な評価を行う場合は、その数値に一定の割引率を加味しなければならない。割引率とは、未来のある時点におけるコストや帰結を現在価値に換算する手法である。割引率を用いることで、異なる時点間のコストやアウトカムの考量が可能になるのである。また、割引率とインフレ率とを混同してはならない。

Transferability of resource utilization and unit cost elements

資源利用情報と単位コストの転用可能性

一般的に言って、特定の国に関して得られた診断技術のコスト・データを別の国での分析に転用するのは不可能だが、コスト・データの個々の要素については、転用できる場合もある。コスト・データの要素ごとの転用可能性をまとめておいたものが、表3である。確かに資源利用データや単位コスト要素データは部分的にしか転用できないものだが、それらは経済学的な評価の重要な側面なので、このようなデータは有用であろう。また、特定の国では、どのような資源をどれだけの量、用いているのか、というデータは、他国のHTA研究者にも有用な情報である。

⁸ *Guidelines for the Economic Evaluation of Health Technologies*: Canada, 3rd edition, 2006;
Guidelines for pharmacoeconomic evaluations in Belgium: Brussels, 2008

⁹ Briggs A, Schulpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. *Health Economics* 1994;3:95-104; Briggs A, Claxton K, Sculpher M. *Decision Making for Health Economic Evaluation*. New York : Oxford University Press, 2006.

表3 コスト・データの要素ごとの転用可能性

データ要素	転用可能性
診断技術評価において組み込まれている資源の種類	大部分は可能. ほとんどの場合, 分析に組み込まれている資源の種類は, 他国でも参考にできるが, 一応のチェックは必要である.
診断技術評価において測定されている資源の量	部分的に可能. よく知られていることだが, 経済学的な分析の基礎データとなる投入資源量データ(例えば特定の治療に要する平均通院日数)は, 国ごとに大きな相違が見られるのである. もちろん国ごとの相違がそれほど大きくないデータもあるが, 資源投入量データ転用可能性については, 経験的な調査に基づき, 慎重に判断する必要がある.
診断技術評価において算定されている資源の量単位当たりコスト	転用は不可能. もちろん, 資源によっては, 国ごとの単位価格の比悪が可能であるかもしれないが, 一般的には単位当たりコストは転用不可能なデータである.

「HTA Core Model for Medical and Surgical Interventions, 1.0R, Work Package 4」からの一部抜粋

Safety (pp.60-63)

安全性

Nicholas Hicks, Marta Lopez de Argumedo, Alberto Ruano Ravina, Nieves Sobradillo, Alison Price, Antti Malmivaara

方法論

概論

介入技術の安全性評価には体系的なアプローチが必要であり、安全性に関する因子の特定、データの抽出・合成・解釈には、HTA で用いられているあらゆる手法が必要なのである。

安全性の評価において考慮すべき側面は多岐に渡るが、まずはそれらの側面に関する定義を示し、アウトカムの測定方法を特定しなければならない¹⁰。また、評価対象となる介入技術が用いられる集団の人口統計学的特性も特定しておかなければならない。

介入技術評価において活用可能な安全性データの量を左右する要因は、以下の通りである。それがどのような技術であるのか、ライフサイクルのどの時点でその技術を評価するのか、その技術を用いている患者の数、特定の有害事象に関する重症度と発生頻度などである。また、有害事象に関する報告が不適切であったり、報告の用語法や整理法が一貫していない、といった原因により、適切なデータを特定できない場合も存在することに留意しておかなければならない。

従って、様々な事情により必要なデータの一部が、HTA の基準に満たない場合もあり得るし、計量データが十分には得られていない状態で、危険性に関する質的な報告を行わざるを得ない場合も存在する¹¹。個々の調査デザインの特長については、Jefferson and Demicheli によるレビュー¹²を参照のこと。

臨床試験で対象とする患者数は比較的に小さなものに留まってしまうため、発生頻度の

¹⁰ Kong Loke Y, D Price, A Herxheimer on behalf of the Cochrane Adverse Effects Subgroup. Including adverse effects. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.5 [updated May 2005]; Appendix 6b.

¹¹ Busse R, Orvain J, Velasco M, Perleth M, Drummond M, Gürtner F, Jørgensen T, Jovell A, Malone J, Rütther A, Wild C. Best practice in undertaking and reporting health technology assessments. *International Journal of Technology Assessment in Health Care* 2002; 18(2): 361–422.; Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *Journal of the American Medical Association* 2001; 285: 437-43

¹² Jefferson T, Demicheli V. Balancing benefits and harms in health care: observational data on harm are already included in systematic reviews. *BMJ* 2003; 327: 750

低い有害事象の多くを臨床試験で検知するのは難しい。また、臨床試験についてはたいていの場合、調査期間も限られているため、遅発性の有害事象の検知も同様に困難である。従ってそのような有害事象を評価に組み入れるためには、症例報告やコーホート分析、登録簿分析、医薬品安全管理システム、市販後調査などのデータが必要である¹³。

有害事象の評価につき、どのアプローチを選択するかで、有害事象の特定に用いるデータが変わってくるため、有害事象の評価そのものが異なってくるのである¹⁴。

Sources of evidence on safety (harms)

安全性・危険に関するデータ

いかなるデータに基づいて安全性を評価したのかは、正確に報告しておかなければならない。データとして用いることのできる情報源は様々なものが存在するので、以下にその例を挙げておく。

- ・ 査読付きの学術論文や灰色文献
- ・ 製薬会社の作成した製品データ、製薬承認申請書
- ・ 有害事象に関する医薬品の安全管理モニタリングシステムのデータ
- ・ 患者の疾病登録情報
- ・ 有害事象に関する個別のデータベースや定期的な安全性報告
- ・ 病院や保健機関で定期的に刊行する統計データ

もちろん、さらに詳細な情報を得たいのであるならば、製薬会社や規制当局、専門家集団への個別の問い合わせも有意義である。

Search strategies for safety information for HTA

HTAに必要な安全性情報の収集方針

データ収集の方針については明確に示しておく必要があり、既に多くの文献がこの主題について論じている。共通して指摘されているのは、情報収集の網羅性と情報の品質とはトレード・オフの関係にある、という事実である。この点に関しては、Cochrane Handbookが包括的なレビュー¹⁵を行っている。一般的に言って、データの収集範囲を広げること（刊

¹³ Loke YK, Derry S, Aronson JK. A comparison of three different sources of data in assessing the frequencies of adverse reactions to amiodarone. *British Journal of Clinical Pharmacology* 2004; 57: 616- 21

¹⁴ Loke et al., 2004.; McIntosh HM, Woolacott NF, Bagnall AM. Assessing harmful effects in systematic reviews. *BMC Medical Research Methodology* 2004; 4: 19

¹⁵ Kong Loke Y, D Price, A Herxheimer on behalf of the Cochrane Adverse Effects Subgroup. Including adverse effects. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.5 [updated May 2005]; Appendix 6b.

行資料，未刊行資料，製薬会社，規制当局，登録情報）は可能だが，収集範囲を広げるならば，調査設計の手法に関しても，幅広く検討すべきである。

文献やデータベースのキーワードが一貫していない，有害事象の報告範囲がまちまちである，といった問題への対処方法に関しては，所定のデータ収集戦略¹⁶が既に存在している。

また，HTAを進めている途中で，目下収集していた情報が不十分だと判明した場合には，評価対象となっている介入技術につき，その有害事象に関する体系的レビューを行うことを検討すべきである。

Extracting evidence

エビデンスの抽出

必要であれば，エビデンス抽出に関する標準化されたアプローチを用いることができる。この段階に関しては，エビデンスの質についての評価を行っておかなければならない。

Synthesizing and analysing evidence

エビデンスの合成と分析

必要であれば，エビデンスの合成と分析に関する標準化されたアプローチを用いることができる。この段階に関しては，エビデンスレベルについての評価を行っておかなければならない。

Interpreting evidence

エビデンスの解釈

エビデンスの解釈に際しては，データ源や収集方法，データそのものや分析に用いた手法につき，質的・量的を問わず，その限界を明確に述べておかななくてはならない。

また，調査の結果得られた知見がどの程度の一般性を持っているのか，という点につき，調査の対象とした母集団の特性を合わせて，報告をしておかなければならない。

¹⁶ Busse R, Orvain J, Velasco M, Perleth M, Drummond M, Gürtner F, Jørgensen T, Jovell A, Malone J, Rütther A, Wild C. Best practice in undertaking and reporting health technology assessments. *International Journal of Technology Assessment in Health Care* 2002; 18(2): 361–422; Golder S, McIntosh H M, Duffy S, Glanville J. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Information and Libraries Journal* 2006; 23: 3-12; Golder S, Yoon Loke, McIntosh H M. Room for improvement? A survey of the methods used in systematic reviews of adverse effects. *BMC Medical Research Methodology* 2006; 6: 3.

参考文献

Busse R, Best practice in undertaking and reporting health technology assessments. Working Group 4 Report. *Int. J Tech Ass in Health Care* 2002;18; 361-422. (Table 17 p 395)

Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Annals of Internal Medicine* 2004; 141: 781-8

Kong Loke Y, D Price, A Herxheimer on behalf of the Cochrane Adverse Effects Subgroup. Including adverse effects. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.5 [updated May 2005]; Appendix 6b. <http://www.cochrane.org/resources/handbook/hbook.htm> safety (accessed 31st May 2005)

Clinical effectiveness

臨床上の有効性 (pp.68-75)

Antti Malmivaara, Pekka Kuukasjärvi, Chris de Laet, Hans Van Brabandt, Regina Kunz

Methodology

方法論

PICO

介入技術評価を行う場合、まず最初に必要なのは、調査課題の特定である。その作業には以下に挙げる PICO 図式が有用である。PICO とは

P : 母集団 (調査対象となる疾患を持った患者群)

I : 介入技術 (評価対象となっている介入技術)

C : 比較対象 (評価対象となっている介入技術の比較先)

O : アウトカム (効果性と安全性を評価する上でのエンドポイント)

の頭文字である。

PICO 図式に基づき調査課題を立てることで、臨床現場における有効性の評価を具体的にを行うことができるのである。PICO 図式に基づいた調査課題は、安全性や経済学的な評価にも応用可能である。

一般的に母集団や介入技術、比較対象は同一であり、アウトカムのみが調査ごとに異なってくる。もちろん場合によっては、安全性や有効性についての詳細を調査するため、部分母集団に着目することもあり得る。

Systematic review of literature

データの体系的レビュー

介入の効果を評価するためには何よりもまず、無作為化臨床試験の結果に関する体系的レビューが必要である¹⁷。体系的レビューの方法論に関しては今日広汎な合意が得られている。最も広く用いられている方法論は、Cochrane Collaboration のもの¹⁸である。無作為化が行われていない比較研究や観察研究に関しても同様の方法論が適用可能なので、無作為化臨床試験の数が少ない介入技術に関しても、HTA は可能である¹⁹。

本章では、臨床現場における有効性について論じるが、そこでの重要な目的とは、特定のヘルスケア環境に関して得られた HTA の知見を、他のヘルスケア環境（一国内部の別の環境や他国）へと容易に転用できるようにすることである。転用可能性の拡大という目的を達成するため、評価結果の報告形式や評価項目、データの提示方法、評価フォーマットや規則に関し、そのガイドを以下に明示しておく。このガイドについては、介入技術の有効性に関して研究を行ってきた様々な専門機関や専門家集団の公刊物において、広汎な範囲で得られているコンセンサスを基にして作成した。具体的には、体系的レビューに関する Cochrane Handbook²⁰、体系的レビューに関する CRD ガイド²¹、CONSORT 声明²²、体系的レビューの報告に関する QUOROM 声明²³、INAHTA による HTA 報告チェックリスト²⁴、データ品質と推奨基準のグレード付けに関する GRADE ワーキング・グループの提言²⁵である。

¹⁷ *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing, 2001.; Khan KS, Kunz R, Kleijnen J, Antes G. *Systematic Reviews to Support Evidence-based Medicine: How to review and apply findings of health care literature*. London: The Royal Society of Medicine Press, 2002.

¹⁸ *Cochrane Handbook for Systematic Reviews of Interventions*. 2005.

¹⁹ Canadian Coordinating Office for Health Technology Assessment. *Guidelines for Authors of CCOHTA Health Technology Assessment Reports*. 2001.

²⁰ *Cochrane Handbook for Systematic Reviews of Interventions*. 2005.

²¹ Khan, K. S., ter Riet, G., Glanville, J., Sowden, A. J., and Kleijnen, J. Undertaking systematic reviews of research on effectiveness: *CRD's guidance for those carrying out or commissioning reviews*. <http://www.york.ac.uk/inst/crd/report4.htm> CRD report no. 4(2nd). 2001. NHS Centre for Reviews and Dissemination; University of York.

²² Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.; Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152-60. ; Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann.Intern.Med.* 2001;134:657-62.

²³ Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999;354:1896-900.; Clarke M. *The QUOROM statement*. *Lancet* 2000;355:756-7.

²⁴ Hailey D. Toward transparency in health technology assessment: a checklist for HTA reports. *Int J. Technol. Assess. Health Care* 2003;19:1-7.

²⁵ Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.; Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B et al. Grading strength of recommendations and quality of evidence in clinical guidelines: *report from an American College of Chest Physicians Task Force*. *Chest* 2006;129:174-81.

本章では、介入に関する体系的レビューの構造を提示する。しかし注意しておいて欲しいのは、本章で述べる内容は決して、体系的レビューのやり方を述べた単なる実践的教科書ではない、ということである。具体的な手法に関しての情報は、詳細な参考文献を付しておいたので、そちらを参照してほしい。

臨床現場における医療技術の有効性の報告に関する実践的なガイドとなることを目的としており、本章の内容は、調査課題の策定からデータの解釈や評価、提言の作成まで、HTAに関するあらゆる段階を網羅している。

How to plan and conduct a systematic review on effectiveness

有効性に関する体系的レビューの計画と実行

Background work

準備的作業

介入技術の有効性に関する体系的レビューを行う際にはまず、その技術に関する HTA 報告や体系的レビューが既に公表されていないか、その点を調査しておくのが賢明であろう。Cochrane 図書館には例えば、体系的レビューに関する Cochrane データベースなど有用なデータベースが存在し、NHSCRD(Centre for Reviews and Dissemination)も HTA データベースを公開している。後者に関してはオンラインでの閲覧も可能である。運がよければ、調査しようとしていた介入技術に関して、包括性も十分であり、方法論も水準を満たしているような調査結果を発見できるであろう。さらにその調査結果が転用可能と判断できれば、その時点で調査は完了である。

データベースを調査しても先行の調査報告が見つからないのであれば、新たに体系的レビューを行う必要が生まれる。以下にそのためのガイドを提示しておく。

Phrasing the problem as focussed and standardised questions

標準化された調査項目への調査課題の割り振り

具体的にレビューを始めるとき、最初に必要となるのは、その介入技術の特性、及びに利点を適切に把握できる調査項目へと、調査課題を割り振ることである。そのためには先に提示しておいた PICO 図式が有用である。即ち、患者の特性やヘルスケア環境など調査対象となる母集団の特定、介入技術の特性の記述、比較対象となり得る現用の介入技術、新たな介入技術の評価に関するアウトカム基準の定義を行わなければならない。もちろんいま述べてきたような主題を余すことなく調査項目に盛り込むのは煩雑すぎとなるだろう。介入技術のメリットの判断に有益である限りで、調査項目を設定すべきである。

最後に、体系的レビューを行うものは、レビューの対象とした個々の研究につき、その調査デザインを明示しておく必要がある。効果性に関するレビューは確かに無作為化された調査のみを対象とする傾向が強く見られるが、無作為化臨床試験のデータが十分な数だけ得られない場合には、無作為化を行っていない調査デザインでの研究もレビュー対象に

組み入れる必要があるだろう。場合によっては、当該のレビューが対象としていない調査課題を明示しておくことも有用であろう。ともかく調査項目を明確に作成しておくことは、レビューを行う上でも、また他の人がそのレビューを活用する上でも、特にその知見の一般化可能性や転用可能性を判断する上で、重要なのである。

従って具体的にレビューを進めてゆくためには、前もってプロトコルを策定しておく必要があり、プロトコルからの逸脱はその旨を記載しておかなければならない。

Locating and selecting studies

調査研究の位置づけと選定

調査項目を具体的に作成し終われば、次に行うべきなのは、文献調査である。文献調査はなるべく包括的で、かつ明確な基準に基づいて行うべきだが、それには様々な情報源が存在している。医療全般に関するデータベースとしては例えば、Medline, Embase, Cochrane Controlled Trials register などが存在し、特定に主題に関するデータベースとしては例えば、CINAHL, PSYCINFO, ASSIA, SOCIOLOGICAL ABSTRACTS などが存在している。臨床試験に関しては、Current Controlled Trials (<http://www.controlled-trials.com/>) や Clinical Trials (<http://www.clinicaltrials.gov/>) が便利である。もちろん学術雑誌や本の要旨を自ら調査することや、いわゆる灰色文献を調査することも有用である。補足的な調査を行う上では、SCIを活用し、引用関係を辿って雑誌論文を探すのも、重要な手段である。また補足調査には、製薬会社への聞き取りや監督機関からの助言も活用できる。調査を包括的なものとするのに必要となる資源（調査時間や費用）は具体的な調査課題ごとに異なってくるであろう。もちろん当該調査研究の位置づけとその限界（調査期間、調査対象とした言語、対象とした集団など）については、明示しておく必要がある。体系的レビューに関する先行文献の調査には、以下のデータベースが推奨される。HTA, DARE, NHS EED, CDSR, and Cochrane CENTRAL.

調査対象とする文献の選定してゆくプロセス、具体的には調査方針や選別基準に関しては、明示的に記載しておかなければならない。

文献選定のプロセスは2回以上反復して行う必要がある。まず第1回目は、タイトルや要旨に基づいた選定を行い、2回目以降は、論文全体を対象とした選定を行う。レビューの最後には同様のプロセスで最新文献の補充を行うことになる。もちろん文献選定の過程には主観的要素が入り込まざるを得ないため、選定プロセスの透明性と客観性を保つためには、2人以上の調査者が関与すること、可能であれば、調査者間の見解の一致点と相違点、相違を解決したプロセス（例えば第3者への相談など）に関しても記述しておくことが望ましい。

文献選定に関する QUOROM 推奨のフロー・チャート²⁶は、非常に効果的との評価を得ている。このフロー・チャートに記載するのは、選定の各プロセスについての、選定した

²⁶ Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: *the QUOROM statement. Quality of Reporting of Meta-analyses*. Lancet 1999;354:1896-900.

文献の総数と、選定から外した文献の総数、各文献の被引用数である。また、選定から外した文献については、必ずその理由を書き添え、最終的な文献総数と、どのような研究デザインに基づいた文献を選定したのか、その類型を記して完成となる。

Assessing study quality

研究の質の評価

体系的レビューを行う上では、検討と対象とした先行研究につき、その研究の方法論的な質を適切に評価しておかなければならない。例えば無作為化臨床試験だと、2重盲検が適切に行われているのか、患者やアウトカムの評価は適切なのか、追跡調査が行われているのか、という点のチェックは、その研究のバイアス評価に欠かせない要素である。その体系的レビューがどのような調査課題を掲げているかによってチェックする項目は当然異なってくるが、各種の研究デザインにつき、バイアスが入り込む要因を下で概観しておく。

無作為化臨床試験の方法論的バイアスに関しては、様々なチェックリストが存在する。主なものとしては、User's Guides to the Medical Literature²⁷, the Consort Statement²⁸, Cochrane Handbook²⁹を挙げることができる。

無作為化臨床試験以外の研究に関する方法論の評価基準については議論が少なく、必ずしも十分に見解が統一されているとは言えないが、John Deeksによるレビュー³⁰は非常に有益である。

個々の研究についての方法論的な評価を終えた後、今度は検討対象とした研究全体の質を把握する必要がある。その作業については、GRADE Working Groupの提案したアプローチ³¹が現在のところ最も包括的である。即ち、個々の研究の質を問題とするだけでなく、

²⁷ Guyatt GH, Rennie D. Users' Guides to the Medical Literature. *A Manual for Evidence-Based Clinical Practice*. AMA Publishing Group; in print, 2007.

²⁸ Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94; Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med*. 2001;134:657-62; Rennie D. How to report randomized controlled trials. The CONSORT statement [editorial; comment]. *JAMA* 1996;276:649.

²⁹ *Cochrane Handbook for Systematic Reviews of Interventions*. 2005.

³⁰ Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:1-186; MacMahon S, Collins R. *Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies*. *Lancet* 2001;357:455-62; Radford MJ, Foody JM. How do observational studies expand the evidence base for therapy? *JAMA* 2001;286:1228-30; Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008-12.

³¹ Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490; Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B et al. *Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians Task Force*. *Chest* 2006;129:174-81.