

平成 30 年 6 月 19 日現在

機関番号：33916

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26460755

研究課題名(和文) 診療ガイドライン作成のためのシステマティックレビュー法の最新化のための調査研究

研究課題名(英文) A survey for modernizing systematic review methods for developing clinical practice guidelines

研究代表者

寺澤 晃彦 (TERASAWA, Teruhiko)

藤田保健衛生大学・医学部・准教授

研究者番号：30399597

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：証拠に基づく医療先進国の公的医療技術評価機関では2000年代後半よりエビデンス統合法が包括的に評価され、その運用法が整理されてきている。英米の同機関の運用マニュアルを調査研究し、推奨されている新規統合法をがん検診ガイドライン改訂のためのエビデンスレビューに応用、その運用性を評価した。従来法より生じた問題点について、新規法は理論的に改良された発展的なモデルに拠っていた。モデル構築が柔軟なベイズ統計も積極的に採用されていた。新規法が容易に運用できる環境は十分に整備されておらず、従来法より多くの労力や専門的知識を必要とした。新規法が必須な状況を適切かつ効率的に選択するリソースの確立が急務である。

研究成果の概要(英文)：Since the late 2000s, AHRQ-EPC program in the US and NICE in the UK, two major health technology assessment agencies in the leading countries of evidence-based medicine, have comprehensively evaluated methods for synthesizing clinical evidence, and developed methods guides. We investigated the usability and relevance of the recommended novel statistical approaches, and applied selected statistical methods in our evidence reviews conducted for updating cancer prevention guidelines in Japan. The novel methods recommended were typically based on advanced statistical models and Bayesian approaches were frequently adopted. Because platforms that allowed novice clinician investigators to easily carry out these methods were limited, application of the novel methods required more time and effort than the conventional methods, in addition to statistical expertise. Establishing resources to efficiently select more appropriate methods to adopt in particular situations is an urgent task.

研究分野：臨床疫学

キーワード：臨床エビデンス統合 メタ解析 ベイズモデル解析 がん検診 がん予防 比較効果 精度

1. 研究開始当初の背景

(1) 人口は高齢化の一途をたどっており、疾患罹患者数も増加している。また、これに対応する新規医療技術の開発は検査・治療薬・医療機器の高度化を要し、医療費総額もさらに増加の一途をたどっている。適切な資源分配は急務であり、効果に関しては新たな医療技術による介入を既存のものと比較し、患者個人、さらには集団において効果の違い(比較効果)のエビデンスを経験的・系統的に検証し、証拠を明確化する必要がある。欧米諸国では医療費問題を検討する手段としてシステマティック・レビューの重要性が広く認知されており、公的医療技術評価機構が主導となってシステマティック・レビューを実施し、結果を医療政策に反映していくシステムがすでに確立していた。2000年代後半にはエビデンス統合法のオーバービューが実施され、最新のエビデンス統合法を採用したエビデンス・レビューの運用マニュアルが整備され始めていた。

(2) 2000年ごろより我が国でも証拠に基づく医療(EBM)の積極的な導入に伴い、診療ガイドライン作成は一般的なものとなった。しかし、ガイドラインの作成は学会をはじめとする学術団体が中心となり、限られた資源の下に行われることが典型的であった。EBMの方法論やシステマティック・レビュー、メタ解析の研鑽機会は十分でなく、標準的なガイドライン作成手順書も近年開発されたエビデンス統合法を十分にカバーできていない状況があった。

2. 研究の目的

研究代表者は2012年より厚生労働省がん研究助成金研究班(現在は国立がん研究センター研究開発費研究班)によるがん検診ガイドライン文献レビュー委員会委員としてわが国のがん検診ガイドライン策定をサポートするエビデンス・レポートの作成を担当している。一般地域住民を対象としたがん検診ではスクリーニング検査とその検査結果に基づき後続する医療介入からなる一連の医療介入が評価の対象となる。臨床検査評価のFrybackモデルで要素を評価すると、必要なアウトカムには診断・スクリーニング精度()、予後予測因子または予後予測精度()、診断・予後予測モデル()、検査結果に基づく介入全体としての健康アウトカム()があり、これらの比較効果について適切なエビデンス統合が必要となる。これらのアウトカムについて、本邦においてガイドラインの新規策定や更新をサポートするエビデンス・レビューを作成する際、世界標準レベルでのレビューが実施できるよう、近年開発された新規エビデンス統合法を調査研究し、その運用性を検証することが目的であった。

3. 研究の方法

(1) 欧米EBM先進国の公的医療技術評価機構である米国アメリカ合衆国保健福祉省医療研究品質庁エビデンス・プラクティス・センタープログラム(AHRQ-EPC)、英国国立医療技術評価機構(NICE)、コクラン共同体を中心に、公表されているエビデンス・レビュー実施ガイドの最新版を調査し、採用・推奨されているエビデンス統合方法の方法論・枠組みを検討した。さらに、同機関が主導して実施した実際のエビデンス・レビューを参考にし、採用・推奨されている方法論の運用状況を観察した。特にガイドライン策定目的に実施したシステマティック・レビューに注目した。

(2) 上記調査研究で同定された方法論を研究代表者、共同研究者が実施する新規システマティック・レビューに採用し、運用性を検証した。適用したシステマティック・レビューはがん検診ガイドラインの新規策定・更新を目的としたトピックスを中心としたが、研究代表者がこれとは独立に実施するシステマティック・レビューにも積極的に応用した。

4. 研究成果

(1) 米国では頻度の多い疾患、予防医療、高額な医療介入、保険診療の適応についてAHRQ-EPCプログラムでシステマティック・レビューを実施し、エビデンス・レポートとしてエビデンスの統合を実施している。本プログラムでは2008年時点の最新のエビデンス統合法を収集・編纂しMethods Guide for Effectiveness and Comparative Effectiveness Reviewsを公表した。2013年にメタ解析の実施に議論がある特定項目について追補が公表されたが、2017年メタ解析法の推奨が全面的に改定された。新規改訂項目はメタ解析を実施すべき要件の図式ガイドによる明示化、固定効果モデル、ランダム効果モデルを除く新たなモデルの紹介、従来まで標準とされたDerSimonian-Lairdランダム効果モデルから状況に応じた特定のランダム効果モデル(特にプロファイル尤度法とベイズ階層ランダム効果モデル)の推奨、稀なイベント・データに対するベイズ階層モデルの推奨、ベイズ階層モデルをリソースが許せば従来法の代替アプローチとして全体として推奨、ランダム化比較試験のネットワークメタ・アナリシスの公式採用であった。

一方NICEでは早期から数学的モデルによる費用効果分析を医療技術評価に採用しており、モデルに使用するパラメータ算定手段として比較効果のメタ解析を利用してきた。2008年および2013年にはMethods Guideが公表され、その運用を具体的かつ実践的とすることを目的としたマニュアルEvidence Synthesis Technical Support Documents seriesが公表された。本マニュアルではベイ

ズモデルを用いた介入に関する階層ランダム効果モデル・メタ解析を紹介しており、一般化線系モデルを基本とするため、リンク関数と尤度関数のペアを組みかえることによって臨床研究から得られる種々のアウトカム・メトリックに対応可能であり、単純な2つの介入間の比較から始まり、ネットワークメタ・アナリシスまで具体的な方法を推奨している。エビデンス統合法については従来からのコントロール群を対照とした相対指標のモデル統合を推奨しているが、従来法でモデル収束が問題となる場合、その安定化を目的とし、対照群リスクをモデルに組み込んだ統合法も紹介した。

診断・スクリーニング検査の研究は一般的な介入研究にはない中間的なアウトカム指標である診断（つまり対象コンディションの有無）、診断後の介入やその変更、そして最終的に介入後の健康アウトカムの変化の評価という臨床判断に沿った固有で複雑なプロセスを有する。AHRQ-EPC プログラムでは通常の介入のエビデンス統合法とは別個に、検査研究に特化したエビデンス統合法を編纂、Methods Guide for Medical Test Reviews として 2012 に公表した。採用された特徴には

複雑なプロセスを定型化する図式的解析枠組みの推奨 診断・スクリーニング検査に特化された文献検索戦略 診断・スクリーニング精度のメタ解析法がある。メタ解析では従来の対象コンディションの有無を 100%評価できる方法（gold standard）の有無に拘わらず一律的にエビデンス統合していた慣習的方法から gold standard の有無によって大きく2分した。gold standard が仮定できる状況では標準的な2変量ランダム効果モデル・階層統合 ROC モデルにより統合感度・特異度および統合 ROC 曲線による「統合点と領域」、あるいは「統合線」によるエビデンス統合が推奨された。一方、gold standard が存在しない場合は多角的なアプローチが必要とされる。検査に関するエビデンスが成熟している場合は診断・スクリーニング精度アウトカムを超えた健康アウトカムの変化を介入研究の枠組みでエビデンスを統合評価（稀なケース）競合する既存検査が存在する場合は新旧検査の一致率（値、陽性・陰性一致率）のランダム効果メタ解析 完全な参照テストを仮定した慣習的方法の踏襲 不完全な参照テストをモデルに取り込んだベイズメタ解析の実施が推奨された。

エビデンス統合には評価したいアウトカムについて、エビデンスの妥当性（バイアスのリスク）を研究デザインごとに適切に評価し、結果の信頼性を吟味しなければならない。AHRQ-EPC プログラムおよび NICE の両機関ではエビデンス・レビューに採用する臨床研究の妥当性評価にコクラン共同が開発した標準的なバイアス評価ツールを採用していた。具体的には 診断精度研究（QUADAS-2） 予後予測因子研究（QUIPS-2） 予後予測

モデル（PROBAST）、ランダム化比較試験（RoB 2.0 tool）、介入の非ランダム化比較研究（ROBINS-I）である。これら最新の妥当性評価ツールの評価項目を吟味し、がん検診ガイドラインのエビデンス・レビュー用に実際に運用可能な形に調整した。

（2）新規エビデンス統合法の実践への応用

胃がん発症抑制効果を評価した健常者におけるピロリ菌除菌に関するランダム化比較試験では、胃がんは稀なイベントであるため、長期のフォローアップで綿密に胃がん発症を観察する必要があるが、フォローアップ方法は多様であり、さらに多数のイベント評価脱落者が存在することが問題の1つであった。ここでは独立した競合イベントを時間生存解析の枠組みでモデルにあてはめたベイズ階層モデル多変量メタ解析を実施した。対照群・介入群ともに0イベントのアウトカムを持つ研究が1件あり、モデル収束の安定化を目的として対照群のベースライン・リスクをモデルに組み込んだ。既報の同じテーマを扱ったメタ解析の結果と異なり、除菌により胃がん発症が減少する証拠は得られなかった（ハザード比 [HR]=0.65; 95%信用区間 [CrI], 0.41-1.0; $I^2=11%$ ）（現在論文投稿中）。

近年、子宮頸がん検診は細胞診による検診からヒト・パピローマウイルスを検出する方法（HPV 検査）による検診が代替法として検討・実用されてきている。現在では細胞診単独、HPV 検査単独、細胞診と HPV 検査併用など、多様な検診方法が可能である。一方、検診の評価対象病変は前がん病変である子宮頸部上皮内腫瘍（CIN）を含んだハイリスク病変であり、病理検査で確認される。検診陽性者の2次検査（精密検査）のコルポスコピーおよび有所見時に必要によって実施する生検（病理検査）は侵襲的な検査であり、通常は検診陽性者のみに実施され、検診参加者全員には実施されない。さらに、コルポスコピー自体は全ての評価対象病変を検出できない不完全な参照テストである。ここでは、細胞診単独、HPV 検査単独、細胞診と HPV 検査併用の3通りの検診についてスクリーニング精度の枠を包括する、浸潤がん発症抑制効果という介入研究全体の枠組みでとらえ、評価した。HPV 検査単独と細胞診および HPV 検査併用での検診は既報のランダム化比較試験では直接比較が実施されておらず、ネットワークメタ・アナリシスを実施して間接エビデンスとして検討した。点推定値としては併用スクリーニング法が平均 32%の浸潤がん減少効果がみとめられるものの、信頼区間は広く、どちらがより優れるという証拠は得られなかった（HR=0.68 [95% CrI: 0.23-1.76]）（現在論文投稿中）。

上記子宮頸がん検診のスクリーニング精

度研究では評価対象病変を検出できる参照基準が全員に実施されておらず確認バイアスの可能性がある。既報研究の中には確認バイアスの可能性を考慮し、全員に病理検査を実施した研究、全員に少なくともコルポスコピーまでは実施した研究、精密検査を全く実施していない対象者に統計モデルで補正を実施した研究に限定し、完全な参照テストが全員に実施されていると仮定した診断精度の2変量ベイズメタ解析を実施した。細胞診のCIN2+をアウトカムとする研究については統合スクリーニング感度は0.635 (95% CrI:0.493-0.762)、統合特異度は0.946 (95% CrI:0.913-0.967)であった。HPV検査については最も頻用されているプラットフォームであるハイブリッド・キャプチャー2については、統合感度は0.860 (95% CrI:0.724-0.937)、統合特異度は0.911 (95% CrI:0.873-0.938)であった。プラットフォームを問わず全HPV検査を対象とする場合、統合感度は0.865 (0.758-0.933)、統合特異度は0.911 (0.881-0.934)とほぼ同様の結果であった(現在論文投稿中)。

認知症診断の完全な参照テストは脳組織の病理診断である。しかし、生前に生検検査や死後剖検として病理検査が実際におこなわれることは稀で、臨床診断の結果で代用するのが実際である。近年検討されているレビー小体型認知症(DLB)を診断する髄液・核医学バイオマーカーの診断精度を評価する目的で、完全な参照テストが実施されない実情を考慮し、標準的なDLB臨床診断基準との値および陽性・陰性一致率のメタ解析を実施した。DLBを除いた各種認知症とDLBを鑑別するMIBG心筋シンチグラフィ(遅延相)の診断精度は臨床診断基準を参照テストとして、統合値は0.85 (95%信頼区間:0.74-0.96)、統合陽性一致率0.98 (95%信用区間:0.92-1.0)、陰性一致率0.95 (95%信用区間:0.88-0.99)、であった(Mishima et al. J Alzheimers Dis 2015)。

子宮頸がんの細胞診検査では頻度は少ないが、適正な細胞診評価が不可能な不適正検体が生じ、運用上の問題となる。近年導入された液状検体法は従来からの塗抹検体法(パニコロウ染色法(従来法))と比較し、不適正検体の頻度が低下すると報告されている。本邦での液状検体法導入に伴い、不適正検体率が低下しているかを検証する目的で、低イベント率に対し混合効果モデルを用いた2項-正常分布を仮定するランダム効果メタ解析を実施した。従来法と液状検体法の不適正検体率を比較した研究では液状検体法に不適正検体が少ない傾向が認められるものの、明確な証拠は得られなかった(統合オッズ比 $=3.5 \times 10^{-2}$ [95% CI: 6.9×10^{-4} -1.7]) (Hosono et al. Cancer Sci 2018)。

先進国における子宮がん検診の対象となる子宮頸部上皮内腫瘍を含んだハイリスク病変の有病率は非先進国と比べ低いと考えられている。しかし、有病率の低さに加え、精密検査の侵襲性から、スクリーニング参加者全員に評価対象病変の精密検査は実施されない。先進国で実施された観察研究ではスクリーニング参加者全員に病理検査による有病率を確認したデータは非常に限られている。一方、先進国以外では有病率の高さもあり、スクリーニング参加者全員に病理検査を実施して有病率を確認したデータも存在する。そこで、稀なイベント率に対するベイズ階層モデルを用いた2項-正常分布を仮定したランダム効果メタ解析を実施した。メタ回帰分析を用い、先進国サブグループにモデル予測を応用し、先進国および非先進国のすべての研究データから推定した先進国の有病率を統合した。メタ回帰解析から得られたモデルでは、先進国でスクリーニング参加者全例に病理検査が実施される状況を仮定した検診1ラウンドにおける平均有病率は2.28% (95%信頼区間:1.01-5.06%)と推定された(現在論文投稿中)。

(3) 新規エビデンス統合法の運用検証
従来のエビデンス統合法の問題点を検討し、理論的に改良した高度な統計学的モデルに基づき新規エビデンス統合法は開発されている。特にモデル構築が柔軟なベイズ統計法を応用した統合法が積極的に採用される傾向にあった。AHRC-EPCプログラムおよびNICEのエビデンス統合法についてのMethod guideではこの傾向が明確に示されていた。複雑な理論とモデル構築に対応できる利点がある一方、高度な専門的知識と経験を有さない一般の研究者でも運用できるソフトウェアは平易なものではなく、解析には時間と労力を要する。具体的にはモデル調整、初期値設定、収束評価、複数の感度分析による剛健性の評価等、専門知識と経験なしには実用性・運用性に大きな問題があった。実際のメタ解析への応用では従来の古典的手法と比し解析に圧倒的な労力と時間を要した。システムティック・レビューやメタ解析の研鑽機会が十分でない研究者へのハードルは高く、汎用性が高いとは言い難い。重要な点としては、統合する対象研究数が多く、研究間の不均一性が低い単純な事例の場合、従来からの古典的手法も新規統合法と結果に大きな違いはないデータも論じられている。具体的にどのような状況で新規エビデンス統合法が必須か適切かつ効率的に選別できるアルゴリズムの構築など喫緊の課題である。また、エビデンス統合を実施する研究者の応用力の育成も必要と考察された。

5. 主な発表論文等
(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計7件)

1. Hosono S, Terasawa T, Katayama T, Sasaki S, Hoshi K, Hamashima C. Frequency of unsatisfactory cervical cytology smears in cancer screening of Japanese women: A systematic review and meta-analysis. *Cancer Sci*. 2018 Apr;109(4):934-943. (査読あり)
2. Trikalinos NA, Nihashi T, Evangelou E, Terasawa T. Positron emission tomography (PET) for prediction of glioma histology: protocol for an individual-level data meta-analysis of test performance. *BMJ Open*. 2018 Feb 17;8(2):e020187. (査読あり)
3. Takeuchi M, Dahabreh IJ, Nihashi T, Iwata M, Varghese GM, Terasawa T. Nuclear Imaging for Classic Fever of Unknown Origin: Meta-Analysis. *J Nucl Med* 2016;57(12):1913-1919. (査読あり)
4. Mishima A, Nihashi T, Ando Y, Kawai H, Kato T, Ito K, Terasawa T. Biomarkers Differentiating Dementia with Lewy Bodies from Other Dementias: A Meta-Analysis. *J Alzheimers Dis*. 2015;50(1):161-174. (査読あり)
5. 寺澤晃彦, 二橋尚志. EBMでみる鑑別診断に使う検査の臨床研究(1): 対象者の参加方法. *健康文* 2016;51:157-163. (査読なし)
6. 二橋尚志, 寺澤晃彦. EBMでみる鑑別診断に使う検査の臨床研究(2): 結果のいろいろな評価方法. *健康文* 2016;51:164-169. (査読なし)
7. 寺澤晃彦. 放射線科学: EBMでみる予後を予測する検査. *健康文* 2016;52:159-163. (査読なし)

〔学会発表〕(計5件)

1. 馬越 弘泰, 二橋 尚志, 岡田 有美子, 寺澤 晃彦. ヨード・ガドリニウム造影剤アレルギーに対する前投薬の有効性: systematic review. 第77回日本医学放射線学会総会. 2018年4月12日 - 15日. 於: 横浜国際平和会議場(神奈川県横浜市).
2. 二橋 尚志, 馬越 弘泰, 岡田 有美子, 平澤 直樹, 寺澤 晃彦. グリオーマ診療における PET 検査研究の臨床エビデンスマップ. 第57回日本核医学会学術総会. 2017年10月5日. 横浜国際平和会議場(神奈川県横浜市).
3. Takeuchi M, Gafter-Gvili A, Garcia-Gomez FJ, Andres A, Blockmans D, Terasawa T. Prediction of spontaneous remission by 18F-FDG PET or PET/CT in classic fever of unknown origin: a meta-analysis. Society of General Internal Medicine Annual Meeting 2017. April 19-22, 2017. Washington Hilton. Washington D.C., USA
4. 小川広晃, 寺澤晃彦. 市中肺炎患者における喀痰グラム染色の起病菌を診断する診断精度の再評価: システムティックレビューとメタ解析. 第86回日本感染症学会西日本・第59回日本感染症学会中部支部合同学術集会・第64回日本化学療法学会西日本支部総会. 2016年11月24日 - 26日. 沖縄コンベンションセンター(沖縄県宜野湾市).
5. Takeuchi M, Nihashi T, Dahabreh I, Terasawa T. 18F-FDG PET or PET/CT and 67Ga-citrate Scintigraphy in the Diagnosis of Classic Fever of Unknown Origin: A Systemic Review and Meta-Analysis of Test Performance and Diagnostic Yield. Society of General Internal Medicine Annual Meeting 2015.

April 22-25, 2015. Toronto, ON, Canada

〔図書〕(計2件)

1. **濱島ちさと**, 加藤勝章, 高久玲音, **寺澤晃彦**, 西田 博, 宮代勲, 吉川貴己.
胃がん検診エビデンスレポート 2014 年度版 . 胃がん検診文献レビュー委員会 . 平成 27 年 3 月 31 日 . p1-221. 国立がん研究センター .
2. 祖父江友孝, 井上和彦, 齋藤 博, 佐川元保, 渋谷大助, 中山健夫, 中山富雄, 成澤林太郎, **濱島ちさと**, 加藤勝章, 高久玲音, **寺澤晃彦**, 西田 博, 宮代勲, 吉川貴己. 有効性評価に基づく胃がん検診ガイドライン 2013 年版・ドラフト . 平成 25 年 7 月 31 日 . P1-94. 国立がん研究センター .

〔産業財産権〕

出願状況 (計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年月日 :
国内外の別 :

取得状況 (計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :

〔その他〕

ホームページ等
該当なし

6. 研究組織

(1) 研究代表者

寺澤晃彦 (TERASAWA, Teruhiko)
藤田保健衛生大学, 医学部, 准教授
研究者番号 : 3 0 3 9 9 5 9 7

(2) 研究分担者

濱島ちさと (HAMASHIMA, Chisato)
国立研究開発法人国立がん研究センター,
社会と健康研究センター, 室長
研究者番号 : 3 0 2 8 6 4 4 7

(3) 連携研究

該当なし

(4) 研究協力者

二橋尚志 (NIHASHI, Takashi)
名古屋大学, 医学系研究科, 講師
研究者番号 : 0 0 4 2 6 4 9 6