

「有効性評価に基づくがん検診ガイドライン」のための
エビデンスレポート作成方法
2018 年度版

2019 年 3 月 31 日

文献レビュー委員会 編集
国立がん研究センター 社会と健康研究センター 検診研究部
がん検診ガイドライン事務局 発行

目次

要旨

I. 背景	1
II. 目的	3
III. 文献レビュー委員会の役割	5
IV. 文献レビュー委員会の編成	7
V. 評価の基本方針	8
1) これまでの評価	8
2) 最終アウトカム	9
3) 中間アウトカム：がん罹患率	9
4) 中間アウトカム：感度・特異度	9
5) 評価方針	13
VI. 文献レビュー方法	15
1) 概要	15
2) 有効性評価のアウトカム	15
3) Analytic Framework 作成	15
4) 研究デザイン	17
5) 適応・除外基準	18
6) 文献検索	19
7) 文献レビューの進め方	23
8) 研究の質評価	23
9) 結果の統合	26
10) メタアナリシス	28
11) サマリーテーブルの作成	36
VII. 証拠のまとめ	42
1) 利益の評価	42
2) 利益と不利益の対比	43
VIII. 対象年齢と検診間隔	46
1) 対象年齢	46
2) 検診間隔	46
IX. ガイドラインのための付加情報	49
1) 医療資源・費用	49
2) 受診者の価値観	49
X. エビデンスレポートの作成と公開	50
文献レビュー委員会 名簿	51
Appendix : Risk of Bias 評価ツール	55

要旨

「有効性評価に基づくがん検診ガイドライン」作成の基本となるエビデンスレポート作成方法を提示する。本レポートでは、システマティックレビューの手順とともに、がん検診の評価の基本方針を明らかにする。

文献レビューの開始に先立ち、がん検診全体の流れを俯瞰する Analytic Framework (AF) を作成し、各段階における検討課題を PICO (Population、Intervention、Comparator、Outcome) 形式で作成する。

AF の各検討課題の適応基準・除外基準を設定し、キーワードを選択、MEDLINE、Embase、Cochrane Central Register of Controlled Trials、医中誌 Web 等の複数の検索エンジンを用いて候補となる研究を抽出する。抽出された研究について、文献レビュー委員会のメンバーが題名や抄録を確認し、候補となる研究を絞り込む。さらに、文献本文のレビューを行い、検討課題の科学的根拠となる文献を抽出し、その成果をサマリーテーブルとして整理する。最終的に抽出された研究は、各研究のチェックリストによる質評価を行う場合がある。また、研究件数、質等の条件がそろえばメタアナリシスを行う。

がん検診の本来の利益である死亡率減少効果の有無を判断し、さらに研究の質評価も踏まえ、Body of Evidence について 3 段階の評価を行う。がん検診の利益があると判断された検診方法は利益と不利益の対比を行い、4 段階で評価する。AF に対応するすべての検討課題の科学的根拠を整理し、最終的にエビデンスレポート案を作成する。エビデンスレポート案は関連分野の臨床医を含む専門家の外部評価を受け、その後の追加修正を経て、ガイドライン作成委員会に提出する。エビデンスレポートは、「有効性評価に基づくがん検診ガイドライン」とともに国立がん研究センターから公開する。

I. 背景

わが国におけるがん検診の有効性評価は、1998 年の厚生省老人保健推進費補助金老人保健福祉に関する調査研究等事業「がん検診の有効性評価に関する研究班」報告書（主任研究者 久道茂）に始まる。以降、1999 年、2001 年と単年度の報告書が作成された。これらの報告書は、各臓器別のがん検診について、当時の各種がん検診関連のエキスパートが高名な評価研究の概要をまとめたものであり、系統的な文献検索や研究の質評価は行われていなかった。

2003 年度から、厚生労働省がん研究助成金によるがん検診ガイドライン作成のための研究班が設置され、がん検診ガイドライン事務局を国立がん研究センター検診研究部検診評価研究室が担当してきた。諸外国のガイドラインを参照し、作成方法を定式化し、その方法に基づき、胃がん検診（2005 年度版）、大腸がん検診、肺がん検診、子宮頸がん検診（2009 年度版）、前立腺がん検診ガイドラインを作成した。2014 年度以降は、国立がん研究センターから胃がん検診、乳がん検診ガイドラインが公開されている。

がん検診の有効性評価は、1976 年の Canadian Task Force on Preventive Health Care による評価レポートに始まる。以降、U.S. Preventive Services Task Force や International Agency for Research on Cancer (IARC) 等がエビデンスレポートを公表している。しかしながら、ガイドラインとエビデンスレポートの役割が明確ではなく、エビデンスレベルと推奨グレードの識別がされていなかった。このため、科学的根拠の評価が即推奨グレードに繋がるという誤解から、科学的根拠の評価が歪曲される可能性があった。こうした誤解は依然として続いており、新たな研究成果が公表されるたびに、社会的混乱を誘発している。

2011 年、Institute of Medicine (IOM) が診療ガイドラインの定義を変更し、科学的根拠の系統的評価とともに利益と不利益の比較が必須となった¹⁾。さらに、文献レビューをガイドライン作成から分離し、その成果はエビデンスレポートとしてまとめることを提案した。ガイドライン作成委員会は、エビデンスレポートを吟味し、さらに、患者の価値観や医療資源等の観点も踏まえ、総合的判断から、最終的に推奨グレードを決定する。IOM では、システマティックレビューの標準化も提案し、その質の向上を牽引した²⁾。この結果、文献レビューは独立して科学的根拠の判断を行うことが明確になるとともに、評価の厳密性が求められることになった。

「有効性評価に基づくがん検診ガイドライン」作成においても、IOM の提案に基づき、「胃がん検診ガイドライン 2014 年度版」では、ガイドライン作成委員会と文献レビュー委員会を分離し、独立した評価を目指した。さらに、2014 年から作成に着手した「子宮頸がん検診ガイドライン」の作成過程において、エビデンスレポートの作成方法を全面的に見直し、今後のがん検診ガイドライン作成のための標準的な文献レビューとその評価方法を提示した。

文献

1. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines; Editors: Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E. Clinical practice guidelines We Can Trust. Washington D.C., The National Academies Press; 2011.
2. Institute of Medicine; Board on Health Care Services; Committee on Standards for Systematic Reviews of Comparative Effectiveness Research; Editors: Eden J, Levit L, Berg A, Morton S. Finding what works in health care: Standards for systematic reviews. Washington D.C., The National Academies Press; 2011.

II. 目的

「有効性評価に基づくがん検診ガイドライン」は、わが国のがん検診、とりわけ国の予防対策として行われる対策型検診の方向性を決定するうえで重要な役割を担ってきた¹⁾。ガイドラインの作成方針は、わが国のがん検診が科学的根拠に基づき提供され、国民にとって不利益を上回る利益をもたらすことを目標としている。わが国における診療ガイドラインの作成は各専門学会を中心に行われているが、その作成方法や判断基準は必ずしも統一されてはおらず、エキスパートオピニオンの集積に留まるものが多い。しかし、がん検診は、一定年齢の無症状の健常者を対象とし、主として地域住民に提供されることから、より確実な科学的根拠が求められる。

1983 年に老人保健法により、国の予防対策が導入された当初には科学的根拠に基づきがん検診を導入するという概念は一般的ではなかったが¹⁾、現在では、がん検診に限らず、すべての医療技術は科学的根拠が検討され、その結果に基づく政策決定が行われることが理想とされる²⁾。特に、新薬の評価では、保険償還の可否や価格調整などの政策判断にも大きな影響をもたらしている。

がん検診で用いられている多くの医療技術は、すでに臨床現場で用いられており、必ずしも高度な技術を要しない。このため、その導入は一見容易に考えられがちである。しかし、有症状者を対象とした臨床現場から、無症状の健常者を対象とした検診に対象を拡大するためには、段階を経た評価が必要となる³⁾。最終的に、検診対象となるがんの死亡率減少効果を証明することにより、がん検診の有効性が評価されることとなる。さらには、がん検診の過程で発生するさまざまな不利益、特に偽陽性、過剰診断、偶発症も配慮されなくてはならない。

科学的根拠の評価を盤石とするためには、評価の再現性、公平性、透明性が必要である⁴⁾。がん検診に関連するエビデンスの評価過程でもこうした基本条件に従ったプロセスが求められる。本レポートは、IOM をはじめとする諸外国のシステマティックレビューの方法を参照しながら、「有効性評価に基づくがん検診ガイドライン」を作成するために、わが国におけるがん検診にとってのベスト・エビデンスを公平かつ適切に評価するための方法を定式化することを目標とする。

文献

1. Hamashima C. Cancer screening guidelines and policy making: 15 years of experience in cancer screening guideline development in Japan. *Jpn J Clin Oncol*. 2018;48:278–286.
2. Baicker K, Chandra A. Evidence-Based Health Policy. *N Engl J Med*. 2017;377:2413–2415.
3. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst*. 2001;93:1054-1061.
4. Institute of Medicine; Board on Health Care Services; Committee on Standards for Systematic

Reviews of Comparative Effectiveness Research; Editors: Eden J, Levit L, Berg A, Morton S.
Finding what works in health care: Standards for systematic reviews. Washington D.C., The
National Academies Press; 2011.

III. 文献レビュー委員会の役割

文献レビュー委員会の役割は、ガイドラインの作成とは独立して、科学的根拠の評価を行うことにある。その成果であるエビデンスレポートの作成にも透明性が求められる。

国立がん研究センターのがん検診ガイドライン事務局はガイドライン作成の責任を担い、全過程の進捗状況を管理し、継続的な作成支援を行う。このため、文献レビュー委員会を支援し、エビデンスレポートを作成する。エビデンスレポートは、ガイドラインとともに、国立がん研究センターから公開される。

がん検診ガイドライン作成に先立ち、国立がん研究センターのがん検診ガイドライン事務局は文献レビュー委員会の設置を検討する。同事務局は、文献レビュー委員会の運営のために、研究費の配分、人材育成の支援、文献収集等を行うが、詳細は両者が協議して決定する。また、システマティックレビューに先立ち、国内外の研究やがん検診の実施状況等を確認するために専門家へのヒアリングを行う場合がある。しかし、システマティックレビューの過程で外部の専門家への意見聴取は行わない。がん検診ガイドライン事務局は、文献レビュー委員会にオブザーバーとして陪席はするが、以降の科学的根拠の評価過程には直接関与しない¹⁾。

がん検診ガイドライン事務局と文献レビュー委員会は、システマティックレビューの実施に先立ち、対象とする検診方法、Analytic Framework (AF)、対象年齢や検診間隔の検討の必要性を協議する。この協議結果をもとに、がん検診全体を俯瞰するための AF を作成し、検討課題を決定する。複数の検索エンジンにより必要となる文献を検索し、検討課題の解答となりうる科学的根拠を抽出し、研究結果や質を評価する。検討課題の科学的根拠を **Body of Evidence** としてまとめる。さらに、利益（死亡率減少効果、子宮頸がん検診では浸潤がん罹患率減少効果）と、利益と不利益の対比に証拠のレベルを付与する。これらの結果をまとめて、最終的な成果としてエビデンスレポート案を作成する。エビデンスレポート案は関連分野の臨床医を含む専門家の外部評価を受ける。

ガイドライン作成委員会は、科学的根拠として利益と不利益以外の要因も考慮し、わが国におけるがん検診の実施の可否を検討する。一方、文献レビュー委員会の評価は科学的評価に留まる。ただし、ガイドライン作成に必要な情報について追加的なレビューを行う場合がある。文献レビュー委員会では科学的根拠の評価から推測される推奨グレードを意識し、科学的根拠が歪曲されることがないように公平な立場から検討する。ガイドライン作成委員会は、エビデンスレポートの吟味を行い、利益と不利益、そのバランスをもとに、他の要因も考慮し推奨グレードを決定する。ガイドライン作成委員会は推奨グレード決定に必要な科学的根拠や追加情報を求める場合があるが、文献レビュー委員会はその要求に応じ、追加的な検討を行い、ガイドライン作成を支援する。

ガイドライン作成委員会の開催には、文献レビュー委員会のメンバーが臨席し、科学的根拠の評価過程や内容の説明を行う。ただし、ガイドライン作成委員会のメンバーを兼ねず、会議には出席するが、推奨グレード決定には関与しない。ガイドライン・ドラフトおよびエ

ビデンスレポート・ドラフト公開後に開催される公開フォーラムに出席し、エビデンスレポートの説明を行い、科学的根拠に関する質疑には回答する義務がある。エビデンスレポートにおける科学的根拠の評価に関する指摘に適切な追加修正を行い、最終レポートをガイドライン作成委員会に提出する。

文献

1. IARC Handbooks of Cancer Prevention, PREAMBLE FOR SCREENING, International Agency for Research on Cancer, Lyon. 2019.

IV. 文献レビュー委員会の編成

文献レビュー委員会のメンバーは、対象となるがんを専門とする臨床医だけではなく、システマティックレビュー、文献検索、メタアナリシス、医療技術評価、診療ガイドライン等の専門家を含む。また、対象とするがん以外を専門とする臨床医や総合内科医等を含む場合もある。

文献レビュー委員会のメンバーは利益相反を明らかにする必要がある。製薬会社や機器メーカー等がん検診に関連する特定の企業から研究助成等を受けている場合には、委員となることができない場合もある。利益相反が認められた場合、委員となることはできても、委員長となることはできない。また、研究助成等の金銭的な援助以外にも、競争的資金の研究助成を受け、対象となるがん検診の評価研究を行っている研究班の研究代表者あるいは研究プロトコル作成者は委員になることができない。ただし、文献レビューを担当できる適切な人材が見いだせない場合には、文献レビュー委員会の委員となることもある。個別のケースについては適宜検討する。

文献レビュー委員会のメンバーは、最終的なエビデンスレポートが公開されるまで、ガイドライン作成委員会、がん検診ガイドライン事務局以外に内容を公開してはならない。エビデンスレポート案を関連学会や臨床医等により外部評価を受ける場合には、守秘義務を確認する。公開方法は、学会、講演、SNS 等のすべての方法を含む。また、公開フォーラムやガイドライン作成委員会でエビデンスレポートを説明する場合でも、がん検診ガイドライン事務局と協議し、説明内容の範囲を決め、それを逸脱してはならない。事前に内容に関する漏洩が発覚した場合には、その処分について、文献レビュー委員会とがん検診ガイドライン事務局が協議する。

V. 評価の基本方針

がん検診の有効性評価は最終アウトカム指標であるがん死亡率を用いることが原則である。しかし、がん検診への早期導入を見据えた評価では中間アウトカムの利用が期待されている。「有効性評価に基づくがん検診ガイドライン」のこれまでの評価と諸外国での評価方法を参照し、エビデンスレポート作成の新たな基準を提示する。

1) これまでの評価

がん検診における新技術の評価を進めていくうえで、同様の問題に直面する可能性は極めて高い。現段階において、生物学的機序に同様であり、かつ検出病変のスペクトラムも同様と考えられる検査方法に限定する Lord らの概念に従うと、感度・特異度の評価で効果判定ができるものと死亡率減少効果の証明が必要な場合の二者に分類される¹⁾。

表 1 にその例を提示した。子宮頸がん検診では液状検体法の死亡率減少効果に関する検討は行われておらず、細胞診従来法 (Pap smear) と液状検体法の精度評価が基本となり、多くのガイドラインで同等の推奨グレードとなっている。ただし、精度評価は横断研究に留まらず、コホート研究、無作為化比較試験が行われている。これらの結果は一貫して細胞診従来法 (Pap smear) とほぼ同等以上の感度・特異度が得られた結果が推奨グレードの科学的根拠となっている²⁾。大腸がん検診における便潜血検査でも、化学法は無作為化比較試験で大腸がん死亡率減少効果が示されている。その化学法と感度・特異度を比較することにより、免疫法も同等の大腸がん死亡率減少効果があると推測され、同様に推奨されている。これらの方法は、いずれも比較対照となる方法の有効性がすでに確立しており、さらに複数の研究から同等以上の感度・特異度が得られている³⁾。

表 1. 死亡率減少効果の証明が不要とされた検診方法

検診	効果の確立した方法	新技術	導入に必要な評価研究	
			精度評価	死亡率減少効果 (進行がん罹患率減少効果)
子宮頸がん検診	細胞診 (Pap smear)	液状検体法	必須	望ましい
	細胞診 (Pap smear)	HPV検査	必須	必須
大腸がん検診	便潜血検査化学法	便潜血検査免疫法	必須	望ましい
	便潜血検査化学法	便DNA検査	必須	必須
	シグモイドスコーピー	全大腸内視鏡検査	必須	必須
	シグモイドスコーピー	CTコロノグラフィ検査	必須	必須

2) 最終アウトカム

がん検診の有効性評価は最終アウトカム指標であるがん死亡率を用いる。全死因死亡率をアウトカム指標とすべきという意見もあるが⁴⁾、一般的には対象となるがんの死亡率が用いられる⁵⁾。特に、マンモグラフィによる乳がん検診について、北欧コクランセンターでは全死因死亡率を評価指標とした場合に検診の効果が認められないというメタアナリシスの結果を公表して以来、国際的な議論が続いている。近年、北欧コクランセンターの評価を支持するスイスがん協会等の対応が国際的にも注目された⁶⁾。しかし、U.S. Preventive Services Task Forceをはじめとして対象となるがんの死亡率を最終アウトカムとして用いている機関が多い。

3) 中間アウトカム：がん罹患率

子宮頸がん検診では、例外的に浸潤がん罹患率を最終アウトカムとして用いてきた European Commission が作成した子宮頸がん検診精度管理ガイドライン第2版⁷⁾ および米国予防医療専門委員会のエビデンスレポート 2018⁸⁾ では子宮頸がんスクリーニングによる効果を検討するアウトカムとして、浸潤がん罹患減少（アウトカムのレベル・ランク3）獲得に至る下位の中間アウトカムとして CIN3+の減少（アウトカムのレベル・ランク4）、CIN2+または3+の検出率上昇（アウトカムのレベル・ランク5）を提唱している（表2）。がん罹患率による評価は、乳がん検診や大腸がん検診でも行われている。

表2. 子宮頸がん検診によるアウトカム減少の効果を示すランキング³⁾

アウトカムのレベル	アウトカムの内容
ランク1	子宮頸がん死亡の減少
ランク2	子宮頸がん罹患（病期1B+の発症）の減少
ランク3	子宮頸がん罹患（微少浸潤がんを含む）の減少
ランク4	CIN3+病変罹患の減少
ランク5	CIN3+病変（またはCIN2+病変）の検出増加 ①累積CIN3+病変検出増加 ②CIN2病変検出増加確認後の検診でCIN3+病変検出の減少
ランク6	スクリーニング陽性数の増加（陽性反応適中度の増加または不変を必要とするが、わずかの減少まで許容する）

注) 国際的にはランク1～3を介入研究で検証可能な最高レベルの証拠とみなすことが一般的であるが、研究の実現性から、ランク4～6に相当する証拠も許容する提唱がされている。

4) 中間アウトカム：感度・特異度

早期がん比率、がん発見率、感度・特異度等は中間アウトカムの指標とはなりうるが、最終的にがん検診の有効性を証明するには脆弱である。しかしながら、各がん検診の特性を勘

案し、代替指標による評価を行う場合もあり、また最終アウトカムの結果を予測するうえでも重要な結果を示す可能性が高い。このため、中間アウトカムについても AF 上でその位置づけを明らかにし、検討を行うことが望まれている⁹⁾。

近年、新技術の開発とともに、がん検診のより早期導入に期待が高まっている。新技術には、現在の技術を改良したものや類似性のある技術もあり、必ずしもすべての技術に最終アウトカムを用いた評価が期待されているわけではない。例えば、子宮頸がん検診では細胞診従来法 (Pap smear) が標準的方法として行われてきたが、液状検体法の導入では浸潤がん罹患率減少までは評価せず、従来法とは感度・特異度を比較することでその有効性が検証された経緯がある。

Pepe らはがん検診におけるバイオマーカー開発について、5 段階の評価過程を設定し、各段階での評価指標を提示している (表 3)¹⁰⁾。その中で、検診の有効性評価の第 1 段階に位置づけられているのが、感度・特異度の算出である。Lord らも診断検査において、新技術の感度・特異度、治療効果、安全性や費用の 3 つの観点から効果判定のための無作為化比較試験の必要性の有無を判断し、判断基準をフローチャートで表示している (図 1)¹⁾。この際、感度・特異度や追加的な発見病変の評価の際の比較対照となるのはすでに効果の確立した標準的手法である。便潜血検査では、新技術がより病変検出に特異的であり、発見病変のスペクトラムが同じであり、相対的に安全で、費用については問題なく、偽陽性を避けることができるメリットが得られるのを予測されることが判断基準となっている。このため、便潜血検査の新技術 (便潜血免疫法) は無作為化比較試験による評価が行われない場合でも導入の可能性ありという判断を示している。

World Endoscopy Organization (WEO) でも、すでに評価の確立した方法と新技術の感度・特異度の測定を評価の第 1 段階に位置づけている¹¹⁾。ただし、評価はそこに留まらず、最終アウトカムによる評価を継続して行うことが望まれている。ただし、ここでいう新技術とは生物学的機序が類似した検査に限定され、まったく異なる検査方法は該当しないことに留意すべきである。IARC も WEO の考え方を踏襲し、生物学的に類似の検査は感度・特異度により有効性を検証可能という見解を示している¹²⁾。一方、U.S. Preventive Services Task Force や American Cancer Society のガイドラインでは大腸がん検診に関しては拡大解釈が行われており、無作為化比較試験により有効性の確立した便潜血化学法を基準として、感度・特異度の高いほぼすべての方法が推奨されている^{13,14)}。ADAPT を用いている American Society of Clinical Oncology (ASCO) ガイドラインもこの考え方に従っている¹⁵⁾。

感度・特異度による評価は、がん検診における新技術の導入をより加速化させる可能性が高いが、一方で不確実性を残しており、真の利益をもたらすかどうか確定的ではない。特に感度が高い検査で検出される追加的病変には過剰診断が含まれる可能性が高いことから、真の治療の評価がない状況での導入は問題がある。WEO でも感度・特異度の評価により有効性の評価は可能としてはいるが、継続的な評価研究を勧めている¹¹⁾。Lord らのフローチャートでは、感度が高いことだけを新技術の導入条件としておらず、同等の感度であっても

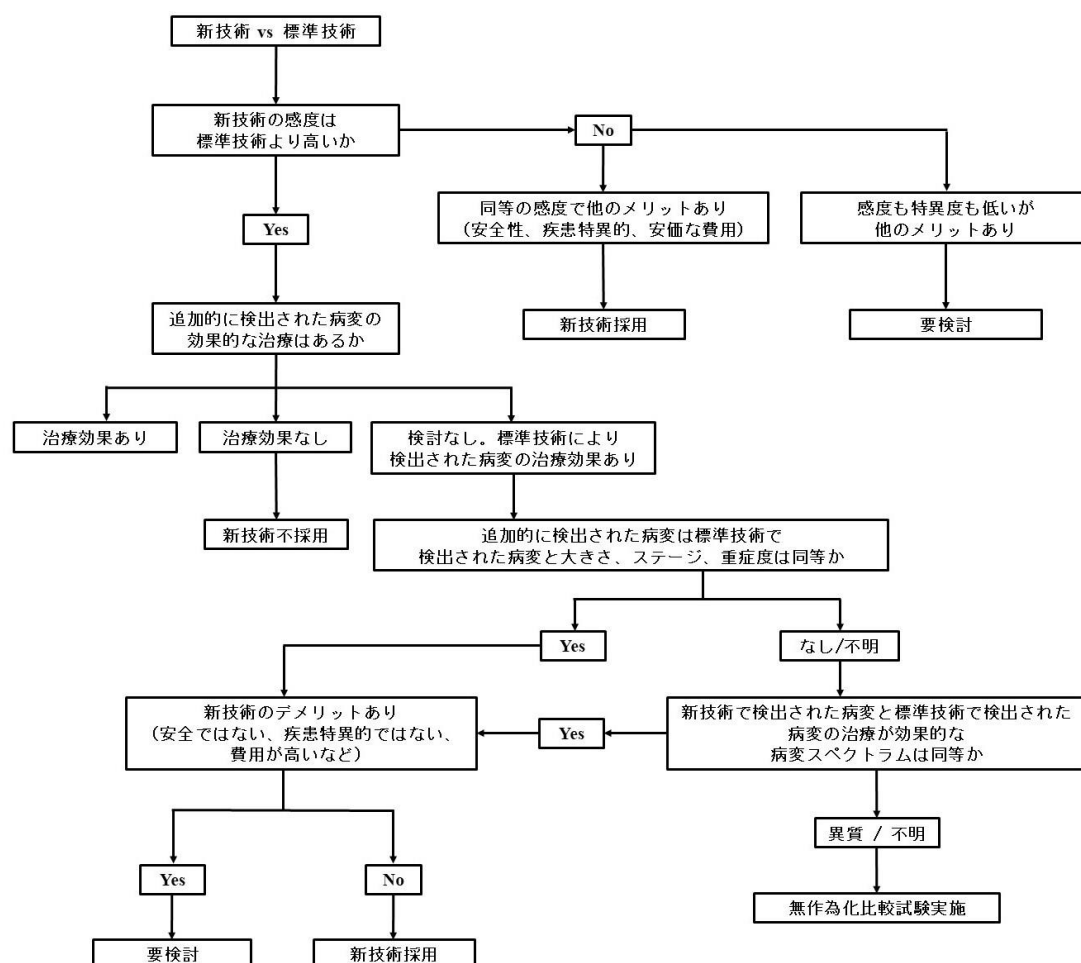
特異度が高く、安全で安価な方法であれば導入可能としている¹⁾。また、感度・特異度の両者が多少劣っている場合でも何らかのメリットがあれば考慮の余地はあると判断される。むしろ、感度が優れている場合には、検出病変のスペクトラムや不利益に慎重に対応することを勧めている。

表 3. 検診のための新技術の有効性評価の過程

段階	分類	研究の対象	概要	期待される成果
Phase 1	探索的研究	患者	臨床への応用を確定する	臨床利用できる測定法の開発
Phase 2	臨床的測定法の妥当性検証	患者	臨床的測定を行い、特定疾患の検出を確定する	患者対象の感度・特異度
Phase 3	遡及的・縦断的研究	健常者 (検診受診者)	バイオマーカーにより症状発現よりも早期の段階で診断可能であること、検診で「陽性」と判断する基準を定める	無症状者対象の感度・特異度
Phase 4	検診の前向き研究	健常者 (検診受診者)	検査により検出できる病変の範囲や特性を検討する。精密検査の対象となる偽陽性率を確定する	検診の実行可能性
Phase 5	がん対策としての検証	健常者 (検診受診者)	検診を行うことで減少できる疾病負担の割合を確定する	罹患率・死亡率減少効果

(Pepe MS, et al.¹⁰⁾ 改訂)

図 1. 新技術の評価研究の必要条件



(Lord SJ, et al.¹⁾ 改訂)

5) 評価方針

がん死亡をエンドポイントとした評価研究を行わず、有効性ありという判断に至る可能性があるのは、標準技術を比較対照として新技術の感度・特異度が正しく測定された場合に限定する。ただし、新技術とは検査の生物学的機序が同様であることと、検出される病変のスペクトラムが同等であることが条件である。例えば、便潜血検査の化学法と免疫法は同等の検査とみなすが、便潜血検査とシグモイドスコーピーは生物学的に同等の検査ではない。感度・特異度は単独の断面調査では十分に判断できない。至適基準（gold standard）を明確にし、以下の過程を経て、標準技術と比較することが原則である。

- ① 検査の感度・特異度を測定するために、同一の対象者に標準技術と新技術の両者を行う。さらに、全対象者に精密検査を行うか、あるいは陰性者に精密検査を行えない場合には補正する。この結果に基づき、検査精度を測定する。遡及的研究と前向き研究の方法があるが、前向き研究がより望ましい。
- ② スクリーニングプログラムの感度・特異度を測定するために、無作為化比較試験あるいはコホート研究により少なくとも 1 ラウンドのスクリーニング、追跡調査を行い、感度・特異度を算出する。さらに、両群とも 2 ラウンドのスクリーニングを行い、prevalence screen と incidence screen の両者の感度・特異度を算出することが望ましい。

上記に該当する質の高い研究が行われ、さらに結果の一致性が確認された場合には、死亡率減少効果を確認できない場合でも有効性ありと判断する。ただし、評価方法は今後変化する可能性もあることから、諸外国におけるガイドライン作成機関等の評価方針を踏まえ、検討していく。

文献

1. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med.* 2006;144:850–855.
2. 平成 20 年度 厚生労働省がん研究助成金「がん検診の適切な方法とその評価法の確立に関する研究」班 平成 21 年度 厚生労働省がん研究助成金「がん検診の評価とあり方に関する研究」班（主任研究者 濱島ちさと）．有効性評価に基づく子宮頸がん検診ガイドライン．2009.
3. 平成 16 年度 厚生労働省がん研究助成金「がん検診の適切な方法とその評価法の確立に関する研究」班（主任研究者 祖父江友孝）．有効性評価に基づく大腸がん検診ガイドライン．2005.
4. Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet.* 2001;358:1340–1342.
5. Jatoi I, Anderson WF, Miller AB, Brawley OW. The history of cancer screening. *Curr Probl Surg.*

- 2019;56:138–163.
6. Biller-Andorno N, Jüni P. Abolishing mammography screening programs? A view from the Swiss Medical Board. *N Engl J Med*. 2014;370:1965–1967.
 7. Arbyn M, Anttila A, Jordan J, Ronco G, Schenck U, Segnan N, Wiener HG, Herbert A, Daniel J (technical editor), von Karsa L. European guidelines for quality assurance in cervical cancer screening. Second Edition. Lyon: International Agency for Research on Cancer; 2008.
 8. Melnikow J, Henderson JT, Burda BU, Senger CA, Durbin S, Soulsby MA. Screening for Cervical Cancer With High-Risk Human Papillomavirus Testing: A Systematic Evidence Review for the U.S. Preventive Services Task Force. Evidence Synthesis No.158. Rockville, MD, AHRQ Publication; 2018.
 9. Jonas DE, Ferrari RM, Wines RC, Vuong KT, Cotter A, Harris RP. Evaluating Evidence on Intermediate Outcomes: Considerations for Groups Making Healthcare Recommendations. *Am J Prev Med*. 2018;54:S38–S52.
 10. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst*. 2001;93:1054–1061.
 11. Young GP, Senore C, Mandel JS, Allison JE, Atkin WS, Benamouzig R, Bossuyt PM, Silva MD, Guittet L, Halloran SP, Haug U, Hoff G, Itzkowitz SH, Leja M, Levin B, Meijer GA, O'Morain CA, Parry S, Rabeneck L, Rozen P, Saito H, Schoen RE, Seaman HE, Steele RJ, Sung JJ, Winawer SJ. Recommendations for a step-wise comparative approach to the evaluation of new screening tests for colorectal cancer. *Cancer*. 2016;122:826–839.
 12. Pinsky P, Rabeneck L, Lauby-Secretan B. The IARC Perspective on Colorectal Cancer Screening. *N Engl J Med*. 2018;379:301–302.
 13. Lin JS, Piper MA, Perdue LA, Rutter CM, Webber EM, O'Connor E, Smith N, Whitlock EP. Screening for Colorectal Cancer: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 2016;315:2576–2594.
 14. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, Etzioni R, McKenna MT, Oeffinger KC, Shih YT, Walter LC, Andrews KS, Brawley OW, Brooks D, Fedewa SA, Manassaram-Baptiste D, Siegel RL, Wender RC, Smith RA. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin*. 2018;68:250–281.
 15. Lopes G, Stern MC, Temin S, Sharara AI, Cervantes A, Costas-Chavarri A, Engineer R, Hamashima C, Ho GF, Huitzil FD, Moghani MM, Nandakumar G, Shah MA, Teh C, Manjarrez SEV, Verjee A, Yantiss R, Correa MC. Early Detection for Colorectal Cancer: ASCO Resource-Stratified Guideline. *J Glob Oncol*. 2019;5:1–22.

VI. 文献レビュー方法

1) 概要

文献レビューの開始に先立ち、がん検診全体の流れを俯瞰する AF を作成し、各段階における検討課題を PICO (Population、Intervention、Comparator、Outcome) 形式で作成する¹⁾。

また、検討課題ごとに適応基準・除外基準を設定する。検討課題ごとにキーワードを選択、MEDLINE、Embase、Cochrane Central Register of Controlled Trials、医中誌 Web 等の複数の検索エンジンを用いて候補となる研究を抽出する。抽出された研究は、文献レビュー委員会のメンバーが題名や抄録を確認し、候補となる研究を絞り込む。さらに、文献本文のレビューを行い、検討課題の科学的根拠となる文献を抽出し、その成果をサマリーテーブルとして整理する。最終的に抽出された研究について、各研究のチェックリストによる質評価を行う場合がある。また、研究件数、研究の質等の条件がそろえばメタアナリシスを行う。

がん検診の最大の利益である死亡率減少効果の有無を判断し、さらに研究の質評価も踏まえ、Body of Evidence について 3 段階の評価を行う。がん検診の利益があると判断された検診方法は利益と不利益の対比を行い、4 段階で評価する。AF に対応するすべての検討課題の科学的根拠を整理し、最終的にエビデンスレポート案を作成する。エビデンスレポート案は関連分野の臨床医を含む専門家の外部評価を受け、その後の追加修正を経て、ガイドライン作成委員会に提出する。エビデンスレポートは「有効性評価に基づくがん検診ガイドライン」とともに国立がん研究センターから公開する。

2) 有効性評価のアウトカム

上記の方針に従い、がん検診の対象となるがん死亡率を最終アウトカムとした研究を優先し抽出する。ただし、子宮頸がん検診では浸潤がん罹患率を、他の検診において基本的な条件が確認できた場合には感度・特異度をアウトカム指標とした研究に基づき、がん検診の有効性を判断する。

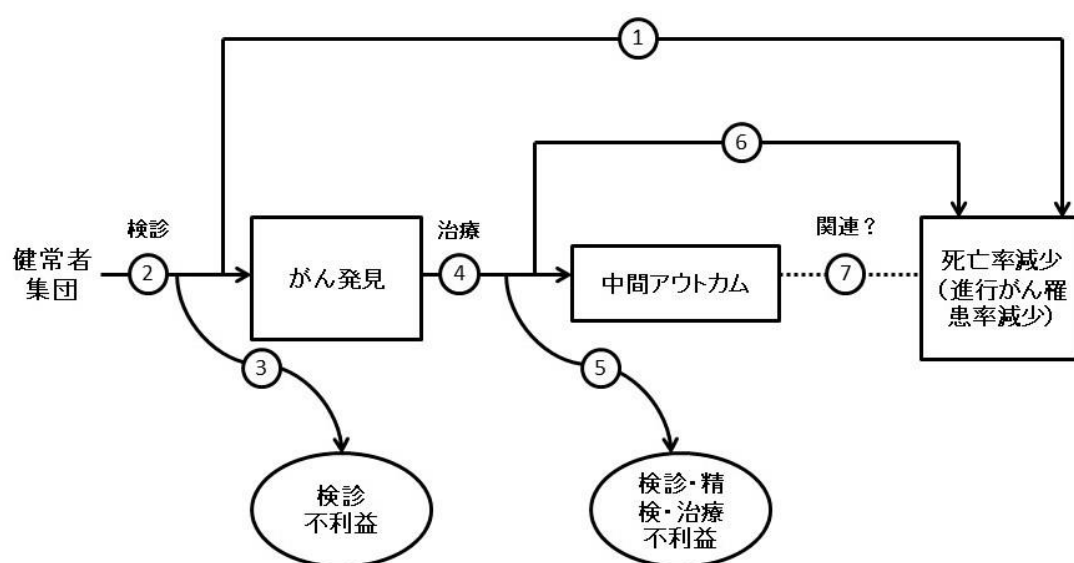
3) Analytic Framework 作成

がん検診全体の流れを把握し、死亡率減少効果のみならず、感度・特異度や偶発症等の不利益を明らかにするための AF を作成する。図 2 には、がん検診の基本的な AF を示した。

がん検診は無症状の健常者集団を対象として実施される。がん検診の実施により、がん死亡率が減少することが科学的に証明できれば、がん検診の有効性があると判断できる。この場合、図 2 の①が相当する。この課題を解決するためのクリニカル・クエスチョンを作成する場合には、PICO の形式に従い、対象 (Population)、検診方法 (Intervention)、比較対照となる方法 (Comparator)、結果 (Outcome) を明らかにする。対象 (Population) は、がん検診の評価の場合は無症状の健常者集団だが、がん検診発見がんに多い治療法の評価等の場合には、がん患者とする場合もある。検診方法 (Intervention) も、検診方法以外の評価、精密検査や手術等の治療法も該当する。比較対照となる方法 (Comparator) は、標準的に行われ

ているがん検診方法や診断・治療である。検診の介入による結果（Outcome）は、がん死亡（子宮頸がん検診の場合は浸潤がん罹患）を原則とする。しかし AF の各段階で必要とされるアウトカムは異なることから、標的となるアウトカムを明確にして、クリニカル・クエスチョンを作成する。すなわち、検討課題①のようにがん検診の直接的な効果を検討する研究もあるが、がん検診の一連の過程の中でその一部に相当する段階、例えば検診の感度・特異度や偶発症発症を検討する場合もある。各検討課題のクリニカル・クエスチョンに基づき、解答となりうる研究を検索するためのキーワードを抽出し、文献検索を行う。

図 2. がん検診の AF



検討課題① 無症状の健常者に新しい検診方法を行うことで、旧来の検診方法に比べ、がん死亡率は減少するか。

- P : 無症状の健常者
- I : 新しい検診方法
- C : 標準の検診方法あるいは検診なし
- O : がん死亡

文献

1. Institute of Medicine; Board on Health Care Services; Committee on Standards for Systematic Reviews of Comparative Effectiveness Research; Editors: Eden J, Levit L, Berg A, Morton S. Finding what works in health care: Standards for systematic reviews. Washington D.C., The National Academies Press; 2011.

4) 研究デザイン

どのような研究デザインを採用するかは、研究課題により異なる。原則的には、教科書的な研究デザインのヒエラルキーに準じる。すなわち、最も信頼性の高い方法は、ポピュレーションベースの無作為化比較試験であり、観察研究がこれに次ぐ（表 4）¹⁾。

表 4. 研究デザインのヒエラルキー¹⁾

優先順位	研究デザイン
1	ポピュレーションベースの無作為化比較試験
2	コホート研究
3	症例対照研究
4	時系列研究、ルーチンで行われる検診データに基づく相関研究

死亡率減少効果に関する研究については、無作為化比較試験を優先する。無作為化比較試験がない場合には、次善の研究としてコホート研究、症例対照研究を検討する場合もある。

死亡率減少効果以外の検討課題は、観察研究を主とした評価となる可能性がある。各検討課題についてベストアベイラブルな科学的根拠を抽出する。この場合も、コホート研究や症例対照研究に限らず、横断研究等の他の観察研究やシングルアームのケースシリーズ研究等も場合によっては検討する必要がある。また、過剰診断の推計等ではモデル研究も含むものとする。

ただし、研究デザインはレベルを決定する絶対的要因ではなく、研究の質を考慮し、最終的な証拠のレベルを決定する。

文献

1. Arbyn M, Anttila A, Jordan J, Ronco G, Schenck U, Segnan N, Wiener HG, Herbert A, Daniel J (technical editor), von Karsa L. European guidelines for quality assurance in cervical cancer screening. Second Edition. Lyon: International Agency for Research on Cancer; 2008.

5) 適応・除外基準

がん検診の対象は、無症状の健常者である。従って、スクリーニングの死亡率減少効果、感度・特異度の評価では、有症状者を含まないことが原則である。ただし、AF 上のがん検診と密接に関わりのある治療やサーベイランス等の評価では、有症状者を対象とする可能性がある。その他の条件は、対象となるがん検診に特有の背景や科学的根拠の公表状況等を勘案し、文献レビューに先立ち、適応・除外条件を設定する。表 5 に図 2 AF_1 の適応・除外基準を示した。

評価研究は原則的には査読ありの学術誌に掲載された原著論文とする。学会抄録やポスター、プロシーディング、レターは対象としないが、情報が限定されている場合には参照する。これらの研究は、いわゆるグレー・ペーパーに相当するが、評価に必要な公表論文が少ない場合には、このような研究と研究登録データベースも活用する。解説や Commentary、Letter、診療ガイドライン、エビデンスレポート、年報・統計は除外するが、システマティックレビューやメタアナリシスは一時的に抽出し、検討の際に参照する。また、メタアナリシスも文献レビュー委員会での独自解析が基本だが、他の研究グループの解析結果をそのまま引用する場合もある。ただし、その採否は、科学的根拠に関する情報が不足している場合やメタアナリシスを行うための情報収集が困難な場合等が考えられる。

評価の対象とする研究は、全言語の研究論文だが、英語ないしは日本語による抄録があり、基本的な内容が確認できる研究とする。ただし、論文題名と抄録で初回の文献スクリーニングを行うことから、抄録のない研究論文は除外する。多くは英語ないしは日本語による論文を対象としていることから、その他の言語の論文については翻訳を依頼する。

表 5. がん検診の死亡率減少効果の適応・除外基準

AF_1	適応	除外
P (Population)	無症状な健常者	患者
I (Intervention)	検診に用いる技術	検診に用いることのない技術 (主として精密検査)
C (Comparator)	標準的で有効性が確立している検診方法	有効性が確立していない検診方法
O (Outcome)	最終アウトカム (全死因死亡、全がん死亡、がん罹患)	中間アウトカム (発見がん、前がん病変発見など)
Design	無作為化比較試験、観察研究 (コホート研究、症例対照研究、時系列研究、地域相関研究)	ケースシリーズ、症例報告
Article	原著論文	総説・解説、ガイドライン、エビデンスレポート、年報・統計などの報告、査読なし報告書等

6) 文献検索

情報収集・文献検索の手法とその結果については、システマティックレビューに関する Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) 声明とそのチェックリストの「情報源と検索」に従い、報告できるよう計画し実行することが望ましい¹⁻³⁾。また、Institute of Medicine (IOM: 現 Health and Medicine Division of the National Academies) は、システマティックレビューのための基準を示しており、その中でエビデンスを特定するための包括的な系統的検索を実施することを推奨している^{4,5)}。さらに、Cochrane Handbook の [7 Searching for studies] が参考となる^{6,7)}。あらかじめターゲットとするべき研究論文等の適格基準が定められ^{6,8)}、それに基づいて検索戦略を決定する。ターゲットとする検診や情報検索について詳しい者の支援を受けながら作業を行い、文献レビュー委員会のメンバーの承認を得ながら進める。

文献検索を行う際に、国際組織や他国にて出版されている既存のガイドラインやシステマティックレビュー、メタアナリシスを調査し、その中で採用されている検索の手法にも留意する。また、システマティックレビューやメタアナリシスにて採用されている検索語は網羅的に含める。さらに、システマティックレビューやメタアナリシスにて採用されている個々の論文のテーマがクリニカル・クエスチョンに一致する場合、それらのキーとなる論文が含まれるよう努める。Cochrane Database of Systematic Reviews (CDSR) の中ですでに出版されている類似のテーマのレビュー中にリストアップされている検索語も参考となる。また、CDSR の中でプロトコールが公開されていれば、エビデンスレポート作成の過程でレビューが発行される可能性もあり、関連テーマについては経過を追う必要がある。Cochrane Handbook for Systematic Reviews of Interventions や Cochrane Reviews of Diagnostic Test Accuracy のプロトコールを含め、システマティックレビューのレジストリーである International prospective register of systematic reviews (PROSPERO) (<https://www.crd.york.ac.uk/PROSPERO/>) の中でシステマティックレビューのプロトコールの参照が可能である⁹⁾。

文献検索は、クリニカル・クエスチョンに沿って行う。クリニカル・クエスチョンごとにターゲットとする情報が違う場合には、検索手法を変える必要がある。

データベースは、MEDLINE、EMBASE (Web サービスである Embase にて検索)、医学中央雑誌 (医中誌 Web にて検索) 等を用いる。無作為化比較試験をターゲットとする場合は、Cochrane Central Register of Controlled Trials (CENTRAL) も検索対象として含める。MEDLINE は、検索プラットフォームとして PubMed や Ovid 等にて検索が可能である。MEDLINE 情報には、米国国立医学図書館 (NLM : U.S. National Library of Medicine) が作成する生命科学用語の統制語 (MeSH : Medical Subject Headings) が付与されているが、新しい文献では、その付与の作業 (インデキシング) 済みでない文献もあり、MeSH terms 以外にも Text Words として自由語検索を組み合わせる必要がある¹⁰⁾。Embase にも Emtree 用語 (統制シソーラス語) が付与されている¹¹⁾。医中誌 Web の「医学用語シソーラス」は、NLM のシソーラス (MeSH) に準拠している¹²⁾。これらのサーチシンタックスは、プラットフォーム (検索

エンジン)により違いがあることに留意する。また、医中誌 Web ではひらがな・カタカナや旧字・新字・異体字などの表記の揺れを吸収するあいまい検索が行われる。異体字や機種依存文字については、1文字1文字をすべて同一視して検索する(例: 頸 <--> 頸、膣 <--> 膣)。「漢字とひらがな」などについては、あらかじめ登録されている文字列に関してのみ、同一視して検索する。例えば、「がん」と「癌」をすべて同一視するのではなく、「子宮がん」など登録されている文字列についてのみ、「子宮癌」「子宮ガン」を同一視して検索する。機種依存文字については、例えば「Ⅱ」と「II」は同一視する。ただし、「CINⅡ」と「CIN2」は同一視しないなど注意が必要である。医中誌 Web にて、検索過程で登録語となっているかヒット数の違いを確認しながら進めるとよい。

検索語は、基本的にはクリニカル・クエスチョンの PICO の、対象 (Population)、検診方法 (Intervention)、比較対照となる方法 (Comparator) に関する語を含める。対象として標的母集団 (target population) は、がん検診を受けるヒトに関する語を選択する。がんの種類により、Target condition に関連する語はさまざまである。検索語には“neoplasm”等の「がん」を指す語、あるいは、前がん病変を指す語等を論理和 (OR) の演算子を用いて含め、その後、身体の一部や臓器を指す語を論理和 (OR) にて検索する。「がん」を指す語のすべてと身体の一部や臓器を指す語のすべてを、論理積 (AND) にて結び絞り込む。これらの論理演算子は、データベースにより利用可能なものに違いがある。さらに検診に関する語に関して、クリニカル・クエスチョンごとにターゲットとする検診方法が違う場合には、一般的な“screening”や“test”等の語に加えて、具体的な検診方法に関連する語を含めれば、さらに網羅的に検索ができる。検診方法 (検査法) に関する語として、目的検査/指標検査 (index tests) や、比較対照となる検診方法 (comparator tests) に関する語を含める。また、必要に応じて至適基準/参照基準 (reference standard) に関連する語等を含める。結果 (Outcome) に関しては、論文のフルテキストにて詳細を確認することになるので、アウトカムに関連する語で絞り込むべきではない。ただし、過剰診断や有害事象等の不利益に関する論文に絞り込みたい場合は、検診方法によって有害事象に関連する語が異なることに留意しつつ検索語に含めるとよい。

組み入れ対象とするエビデンスの研究デザインとして、新しい検診方法の有効性を検討するために、旧来の検診方法あるいは検診なしとの比較を行った無作為化比較試験を検索にて抽出したい場合には、無作為化に関連する語も検索語として採用し絞り込むことになる。一方、検査精度に関するものや、対象集団の価値観や意向についての文献検索においては、さまざまなデザインの観察研究を漏らさないために、一般的にデザインに関する語で絞ることは避けるべきである。観察研究の研究デザインも論文のフルテキストにて詳細を確認することになる。

新しい検診方法についての検索では、その方法が用いられるようになった年代以降で絞り込むことも可能である。

論文の言語は、先進国の住民らのみを対象とする論文や日本国内の対象者に関する研究

論文に絞り込む場合は、英語と日本語にて絞り込むことも可能である。医中誌 Web での検索の場合、英語でのタイトルや抄録を持つ論文も収載されているため、検索語には日本語と英語の両者を含めると、より網羅的な検索となる。

検索の際は、対象のデータベース、検索日、検索式、検索語とヒット数、対象とした出版年等を含む検索履歴を記録し、書誌情報をダウンロードする。

検索テーマに合致する資料（文献）は、適合資料と称されるが、検索して得た資料全体に対する適合資料の割合を「精度（もしくは適合率）（precision ratio）」という。また、適合資料（潜在的に存在するものも含める）の中で検索にて抽出された割合を「再現率（recall ratio）」という。検索で得た資料の中で、適合資料でない部分は「ノイズ（noise）」となり、適合資料の中で検索されなかった部分は「モレ（drop-out）」となる。高再現率（つまりモレが少ない）と高精度（つまりノイズが少ない）が両立することは難しく、これらの指標は逆方向に動くといわれている¹³⁾。モレの少ない網羅的な検索を心がけると、大量の情報からスクリーニング作業を行うことになるので、論文の適格基準を明確にしてスクリーニング作業に臨む必要がある。また、適合について検討された Search Filter も公開されているので参考となる¹⁴⁾。

データベースごとに検索した結果から重複文献を除いて書誌情報の論文タイトルや抄録の情報から 1 次スクリーニングを開始する。フルテキストを取り寄せる論文を決定し、フルテキストによる 2 次スクリーニングを行う。スクリーニング作業は、2 名以上で独立して行うことが推奨される。2 次スクリーニングにて不採用になった論文の不採用理由も記録する。このようにしてスクリーニング作業にて文献が絞られた過程は、ハンドサーチにて収集した論文の数とともに、フロー図として記録する^{1,2)}。

文献

1. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 2009;6:e1000097. doi: 10.1371/journal.pmed.1000097.
2. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM; the PRISMA-DTA Group, Clifford T, Cohen JF, Deeks JJ, Gatsonis C, Hooft L, Hunt HA, Hyde CJ, Korevaar DA, Leeftang MMG, Macaskill P, Reitsma JB, Rodin R, Rutjes AWS, Salameh JP, Stevens A, Takwoingi Y, Tonelli M, Weeks L, Whiting P, Willis BH. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018;319:388–396.
3. 相原守夫. 診療ガイドラインのための GRADE システム 第 3 版. 東京, 中外医学社;2018;291–294.
4. Institute of Medicine March(IOM). Finding What Works in Health Care: Standards for Systematic Reviews. 2011. <http://nationalacademies.org/hmd/Reports/2011/Finding-What->

- Works-in-Health-Care-Standards-for-Systematic-Reviews/. 日本語訳：医療における解決策の模索. システマティック・レビューのための基準（相原守夫 訳）. http://www.grade-jpn.com/iom/Finding_What_Works_in_Health_Care-Standards_for_Systematic_Reviews-j.pdf. Accessed May 7, 2019.
5. 相原守夫. 診療ガイドラインのための GRADE システム, 第 3 版. 東京, 中外医学社;2018:280–283.
 6. Editors: Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. [updated March 2011]. The Cochrane Collaboration. 2011. <https://training.cochrane.org/handbook>. Accessed May 7, 2019. (Version 6. は 2019 年 10 月に一般公開予定)
 7. de Vet HCW, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Chapter 7: Searching for studies. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 0.4 [updated September 2008]. The Cochrane Collaboration. 2008. <http://srdta.cochrane.org/>. Accessed May 7, 2019.
 8. Bossuyt PM, Leeflang MM. Chapter 6: Developing Criteria for Including Studies. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 0.4 [updated September 2008]. The Cochrane Collaboration. 2008. <http://srdta.cochrane.org/>. Accessed May 7, 2019.
 9. Editorial and Publishing Policy Resource: PROSPERO. <https://community.cochrane.org/editorial-and-publishing-policy-resource/overview-cochrane-library-and-related-content/prospéro>. Accessed May 7, 2019.
 10. 岩下愛, 山下ユミ, 阿部信一, 奥出麻里. 図解 PubMed の使い方:インターネットで医学文献を探す. 第 7 版. 東京, 日本医学図書館協会, 紀伊國屋書店;2016:72. (PubMed は 2020 年 1 月にリニューアルを予定しており、PubMedLabs にて試験公開中. <https://www.ncbi.nlm.nih.gov/labs/pubmed>. Accessed July 31, 2019)
 11. Embase 基本検索ガイド. エルゼビア・ジャパン株式会社. 2017. https://www.elsevier.com/_data/assets/pdf_file/0010/806923/Embase_searchguide.pdf. Accessed May 7, 2019.
 12. 諏訪部直子, 平紀子. わかりやすい医中誌 Web 検索ガイド: 検索事例付. 東京, 日本医学図書館協会;2013:84–85.
 13. 田窪直規（編著）, 岸田和明, 小林康隆, 時実象一, 山崎久道, 渡邊隆弘. 情報資源組織論（現代図書館情報学シリーズ）. 東京, 樹村房; 2016:106–107.
 14. Ovid Expert Searches - Health Science. (Randomised Controlled Trials – MEDLINE), (Randomised Controlled Trials – EMBASE), (Ovid MEDLINE - Limit to CANCER) etc. http://resourcecenter.ovid.com/site/resources/expert_search/healthexp.html#OvidFilters. Accessed May 7, 2019.

7) 文献レビューの進め方

文献レビューには、シングルチェックとダブルチェックがある。どちらを選択するかは、検討課題の重要性、抽出された候補文献数、文献レビュー担当者数、エビデンスレポート作成の進捗状況等により決定する。

シングルチェック、ダブルチェックともに、第1段階として題名と抄録によるチェックを行い、候補文献を抽出し、第2段階では研究論文全文をチェックし、最終候補を絞り込む。ダブルチェックは原則として2人の評価者が独立して行い、結果が一致しない場合には、会議で採否を検討する。シングルチェックの場合には、保留例について会議で検討する。採否の基準は、各検討課題の適応・除外基準に従う。

最終採用文献について、研究デザイン別の質評価やメタアナリシスを行う場合がある。この過程で、さらに科学的根拠が集約され、最終的に研究課題の **Body of Evidence** を構成する研究が決定する。構成される各研究についてはサマリーテーブルや構造化要約を作成し、エビデンスレポートに添付する。無作為化比較試験のように長期間にわたり、複数の研究が公表されている研究シリーズでは、研究課題の目的別に、必要であれば複数の研究からの情報をまとめた形で構造化要約やサマリーテーブルを作成する。

8) 研究の質評価

いずれの研究デザインにおいても計画、実施、データ解析、結果の解釈、報告のあらゆる段階において真の結果（効果推定値）に影響をおよぼす要因が存在する。本レビュー委員会が実施する文献レビューでは、コクラングループが開発した標準的な研究の質評価ツール（**Risk of Bias** 評価ツール）最新版を利用し、各レビューのテーマに合わせた調整を行い、研究の質の評価を実施する（Appendix として付記）。**Risk of Bias** 評価ツールは世界各国が実施するシステマティックレビューで頻用されており、その妥当性が検証されている。実際の評価は2名の評価者が独立して行い、不一致となった評価項目については評価者間で再検討をする。再検討の段階で結論が一致しない場合は、担当した評価者2名を含むレビュー委員会全体で討議し、最終的なコンセンサスを得る。

①無作為化比較試験

がん検診（介入）の有効性評価では、無症状の健常者を対象とし、異なる2つ（あるいは2つ以上）の検診（あるいは検診を実施しない比較対照）を無作為に割り付け、割り付け群間の結果を比較する無作為化比較試験が最も信頼レベルが高い証拠とみなされる。しかし、無作為化比較試験にも他の研究デザイン同様、真の結果に影響をおよぼす要因が依然として存在しうる。研究の質の評価には、改訂版 **Cochrane** の **Risk of Bias (RoB 2.0)**¹⁾ を利用する。本ツールを使用し、5つの検討項目（無作為化割り付け、各検診群の規定された介入内容からの逸脱、未把握のアウトカムデータ、アウトカムの測定方法、選択的な報告）の評価を行い、最終的に研究全体の **Risk of Bias** として結果の信頼性について検討する。

②コホート研究および類似するデザインの研究

無作為化比較試験による証拠が得られない場合、比較すべき介入がすでに何らかの経過で実施された集団を観察して評価したコホート研究（介入についての非無作為化比較試験）を採用する必要がある。この研究デザインの質評価を目的としたツールの暫定版が 2014 年コクラングループから報告された（ACROBAT-NRSI）²⁾。この暫定版をもとに、主にコホート研究を対象として改訂し、2016 年に発表されたものが ROBINS-I ツールである³⁾。本ツールは無作為化比較試験のための Risk of Bias ツールと同様の検討項目を有し、介入開始前の要因（交絡、対象者の選択）、介入自体（介入内容をグループ化する判定方法）、介入開始後の要因（意図される介入内容の逸脱、未把握のアウトカムデータ、アウトカムの測定方法、選択的な報告）の 3 大分類 7 項目から成る。これら個別の評価を第一に行い、最後に研究全体の Risk of Bias として結果の信頼性について検討する。

③症例対照研究

介入についての非無作為化比較試験（コホート研究）に用いる質評価と検討項目は共通しており、症例対照研究を評価する目的に特化したツールは ACROBAT-NRSI²⁾ で暫定版として公開されたが、2019 年 9 月現在まだ改訂中である。新たな改訂版が発表されるまでは ACROBAT-NRSI の暫定版に従い、コホート研究同様に評価する。介入開始前の要因（交絡、対象者の選択）、介入自体（介入内容をグループ化する判定方法）、介入開始後の要因（意図される介入内容の逸脱、未把握のアウトカムデータ、アウトカムの測定方法、選択的な報告）について個別に評価を実施する。

④診断およびスクリーニング精度研究

スクリーニング検査および精密検査の検査精度については 2011 年に発表された改訂版 QUADAS ツール（QUADAS-2）⁴⁾ を使用する。検診対象者の選択（連続サンプリング、ターゲットデザイン）、対象検査（参照基準検査との盲検化での評価、診断閾値の妥当性と明確さ）、参照基準検査（標的状态を確実に診断できる信頼性、対象検査との盲検化）、両検査実施のフローとタイミング（フォローアップ、検証バイアス、脱落、解析からの除外）の 4 項目について Risk of Bias と結果を適応するにあたっての懸念事項を評価し、最終的に研究全体として結果（診断精度推定値）の信頼性を検討する。

文献

1. Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, Reeves B, Eldridge S. A revised tool for assessing risk of bias in randomized trials In: Chandler J, McKenzie J, Boutron I, Welch V (editors). *Cochrane Methods. Cochrane Database of Systematic Reviews* 2016, Issue 10 (Suppl 1):29–31.
2. Sterne JAC, Higgins JPT, Reeves BC on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0. 2014. <http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/robins-i/acrobat-nrsi/>. Accessed May 7. 2019.
3. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR, Chan AW, Churchill R, Deeks JJ, Hróbjartsson A, Kirkham J, Jüni P, Loke YK, Pigott TD, Ramsay CR, Regidor D, Rothstein HR, Sandhu L, Santaguida PL, Schünemann HJ, Shea B, Shrier I, Tugwell P, Turner L, Valentine JC, Waddington H, Waters E, Wells GA, Whiting PF, Higgins JPT. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ*. 2016;355:i4919.
4. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JAC, Bossuyt PM, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–536.

9) 結果の統合

はじめに、採用した研究とその結果がレビューの目的に設定した検討課題にうまくあてはまるか（適用性）を評価する。下記に述べる PICO の枠組みに従い、質的・記述的な統合（qualitative synthesis）を実施する。無作為化比較試験、コホート研究、症例対照研究、スクリーニング精度研究等、研究デザインごとに図表等を使用し、各 1 次研究の特性を提示し、臨床的な異質性の有無を検討するのが目的となる^{1,2)}。

①Population（検診対象者）

年齢、性、許容される既往歴や合併症等検診対象者の特性を「研究の参加基準（方法）」と「実際に観察された検診参加者の特性（結果）」の両者を吟味し、質的・記述的に統合する。介入（群）の上乗せ検診効果を評価するために設定した比較対照（群）（例：従来法の検診、あるいは検診なし（比較対照群））のアウトカム発生率等、ベースラインリスクを評価し、研究が実施された集団自体の検討課題への合致状況（適用性）を検討する。

②Intervention（検診内容）

新規介入として使用したスクリーニング検査の特性を評価する。検討すべき項目にはスクリーニングに使用するアッセイやデバイス名（商品名、バージョン等）、実施方法と実施回数、実施者の特性（実施者の職種や専門性、経験等）、結果の判定基準とその評価者の特性、検診アルゴリズム（検診検査結果による精密検査の実施法や経過観察方法等）、共介入（検診結果や精密検査の結果に基づく治療等 2 次的な介入とその特性）を含める。さらに、スクリーニング、精密検査、治療については参加コンプライアンス（アルゴリズムの順守率）や検査・介入の実施自体の成功・不成功に関するデータも提示する。これらから、実際に実施された介入内容の適正度や検討課題との合致状況（適用性）を検討する。

③Comparator（s）（比較対照となるスクリーニング内容、あるいは検診なし）

従来型のスクリーニング検査が対照群となる場合は前述の新規介入として使用した検診内容の評価と同様の吟味を行う。従来型スクリーニングの内容（あるいは検診を実施しない比較対照）が検討課題で想定した現在の標準的な状況と隔たりがないか検討する。

④Outcome

アウトカムの定義、アウトカムの把握方法（研究固有の目的として実施された介入による積極的なアウトカムの把握、研究以外の目的で現行システムより構築されたデータ（routinely collected data）³⁾、アウトカムの内容の把握実施者、評価判定基準とアウトカム判定者について特性を吟味し、異質性を評価する。通常検討されるアウトカムには全死亡、全がん死亡、対象がん特異的死亡、対象がんの発症、対象がんの前がん病変発症、スクリーニング精度等がある。

⑤各アウトカムの効果推定値

各研究の数量的結果はアウトカム別に分類し、効果推定値を信頼区間とともにフォレストプロット等のグラフで提示し、統計学的な異質性を検討する。全体的な傾向を質的・記述的に評価する。

本項で述べられた質的・記述的なデータ統合は、採用した研究を臨床的な観点から吟味し、さらに研究デザイン、実施状況を系統的に評価する準備段階となる。この準備段階が終了後、次項にて記載する第2段階であるメタアナリシス（定量的統合）実施の判断に移行する⁴⁾。

文献

1. Institute of Medicine; Board on Health Care Services; Committee on Standards for Systematic Reviews of Comparative Effectiveness Research; Editors: Eden J, Levit L, Berg A, Morton S. Finding what works in health care: Standards for systematic reviews. Washington, D.C., The National Academies Press; 2011.
2. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD, Agency for Healthcare Research and Quality; 2014. <https://effectivehealthcare.ahrq.gov/>. Accessed May 7, 2019.
3. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Routinely collected data and comparative effectiveness evidence: promises and limitations. CMAJ. 2016;188:E158–E164.
4. Morton SC, Murad MH, O'Connor E, Lee CS, Booth M, Vandermeer BW, Snowden JM, D'Anci KE, Fu R, Gartlehner G, Wang Z, Steele DW. Quantitative Synthesis – An Update: Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Scientific Resource Center under Contract No. 290-2012-0004-C). AHRQ Publication No. 18-EHC007-EF. Rockville, MD, Agency for Healthcare Research and Quality; 2018.

10) メタアナリシス

①定量的統合実施の判断

海外のグループからメタアナリシス（定量的統合）を実施すべきかどうかの判断に役立つ系統的なアルゴリズムが複数提唱されている^{1,2)}。当レビュー委員会が行う文献レビューもこれを参考に判断を行う。以下、検診を実施して得られる臨床アウトカム（死亡やがん発症等の前項 8）で提示したスクリーニング精度よりも上位レベルのアウトカム）の比較を検討する場合について述べる。

前項 8）にて吟味した質的・記述的統合の結果を参考に、採用した研究の「臨床的な類似性」を検討し、メタアナリシスが実施可能で十分類似な「研究群」があるか検討する。この「研究群」は採用した研究全体にあたる場合のほか、類似と評価できる一部のサブグループのみの場合が考えられる。前項 8）の検討事項のうち、ここで優先すべき項目の順位には、介入内容（新規スクリーニング法）、対照（比較対照となるスクリーニング内容、あるいは比較対照）、研究の検診参加者特性、研究のデザインと実施状況、アウトカム内容、経過観察期間が提案されている¹⁾。

「臨床的な類似性」が担保される場合にも、フォレストプロットに提示した個々の研究の結果（効果推定値）から「統計学的な異質性」が観察される場合には注意を要する²⁾。研究の質（Risk of Bias）、小サイズ研究効果（小サンプルサイズの研究が大サンプルサイズの研究より大きな効果推定値を報告する傾向が観察される現象、報告バイアスや選択報告バイアス等の可能性も含まれる）³⁾の可能性も考慮し、メタアナリシスを実施すべきか判断する²⁾。採用した研究数が 10 件以上の場合には小サイズ研究効果について統計学的に検討すべきか考慮する³⁾。

②メタアナリシスのモデル選択

採用した研究には、臨床的あるいは研究のデザインや実施上の異質性が観察されると思ってよい。メタアナリシスを実施することが妥当であると判断した場合も、これらの類似性が極めて強く認識される場合を除き、研究間の異質性を考慮したランダム効果モデルを使用するのが一般的である²⁾。

固定効果モデルでは、統合するすべての研究において真の効果はまったく同じ（つまり唯一）と仮定し、各研究に共通した統合点推定値とその不確実性を表す信頼区間を計算する（図 3）。これに対し、ランダム効果モデルでは、統合する各研究の効果はそれぞれの研究で固有であり、研究の数の分だけ異なる真の効果が存在することを仮定する。この異なる効果が「平均効果」を中心とし、異質性を表すパラメータである τ を標準偏差とした正規分布をとると仮定したものがランダム効果モデルである⁴⁾。算定される点推定値とその信頼区間はあくまで効果の傾向を代表する「平均効果」とその不確実性を観察しているに過ぎない。このため、固定効果モデルとは結果の解釈も異なることに注意が必要である。

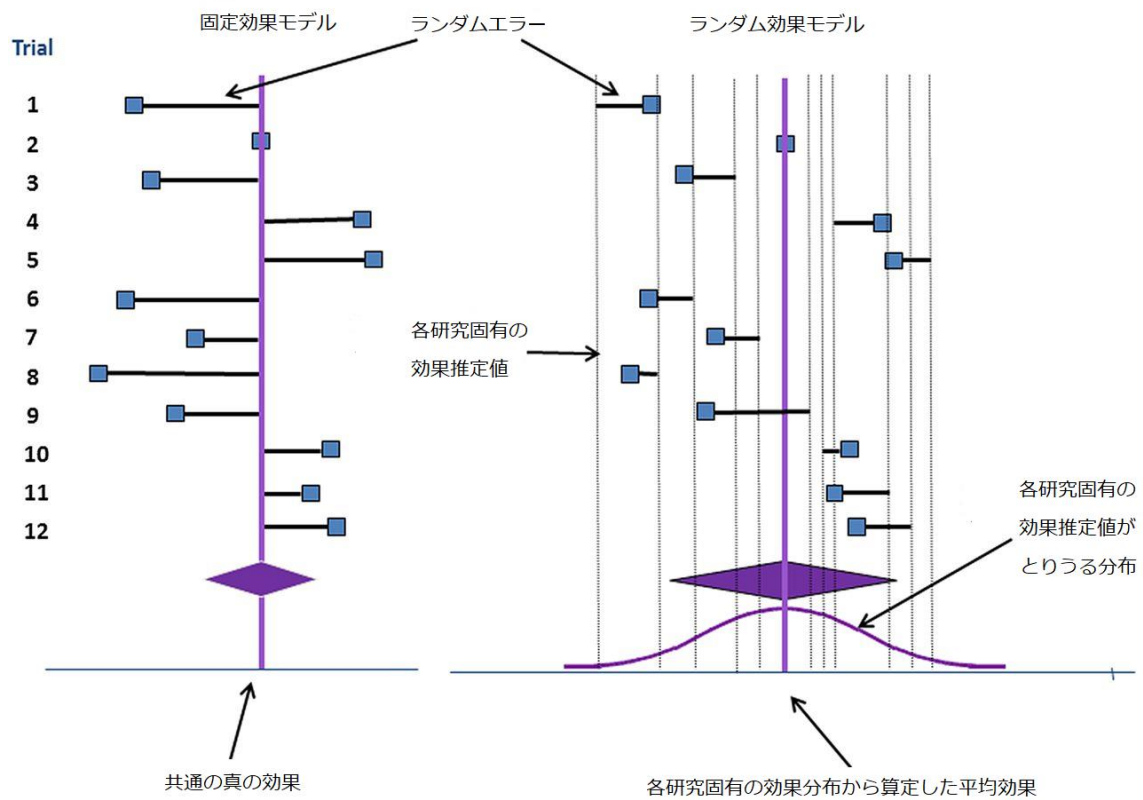


図 3. 固定効果モデルとランダム効果モデルの違い (Nikolakopoulou A, et al. Evidence-Based Mental Health 2014. BMJ Publishing Group Ltd.から許諾を得て掲載)

研究間の統計学的異質性が大きい場合（つまり、結果の推定値が広い範囲にわたる場合）、平均効果の信頼区間は個々の研究に固有の効果をすべて含む結果とはならない。このため、ランダム効果モデルでは、個別の研究にも適用可能な効果の範囲を示すため、予測区間を提示する（図 4）^{5,6)}。

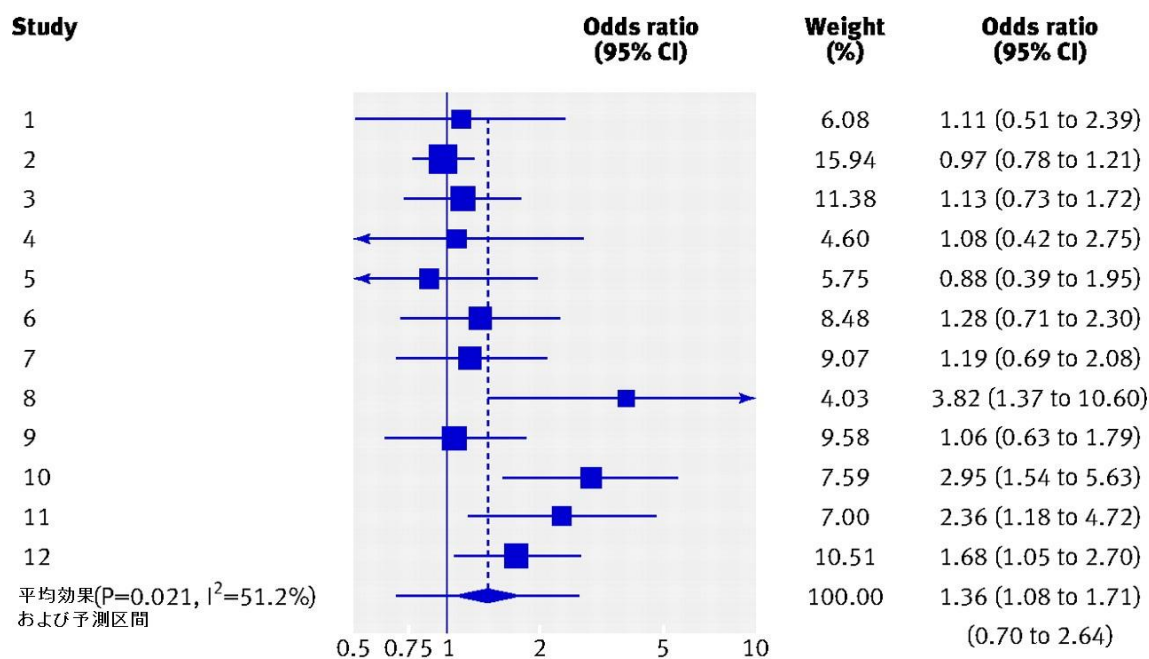


図 4. 平均効果の信頼区間（青色で最下部に提示されたダイヤモンド）とともに予測区間（最下部の青色ダイヤモンドから延長する青色水平線で示される）が提示されている（Riley RD, et al. BMJ 2011. BMJ Publishing Group Ltd.から許諾を得て掲載）

ランダム効果モデルでは 1980 年代に発表された DerSimonian-Laird 法⁷⁾ が頻用されている。この理由は計算方法が単純で多くの統計ソフトに標準装備されていることによる。しかし、前述の異質性を表すパラメータである τ が過小評価されやすく、信頼区間が過剰に狭く算定される難点がある。メタアナリシスが実施される多くの典型的な状況ではこれらの問題が起こりやすく、近年 DerSimonian-Laird 法以外のランダム効果モデルが推奨されている (図 5)^{2,4,6)}。モデルの具体的な選定方法は推奨に従うが、近年新たに提唱されたモデルである Hartung-Knapp 法および Sidik-Jonkman 法 (HKSJ 法)^{8,9)} およびその HKSJ 変法¹⁰⁾、Skovgaard 補正プロファイル尤度法¹¹⁾、階層 Bayesian モデル等を DerSimonian-Laird 法より優先的に考慮し、特定のモデルを唯一の解析方法とせず、感度分析として複数の推奨されるモデルを実施し、結果の安定性を確認する。

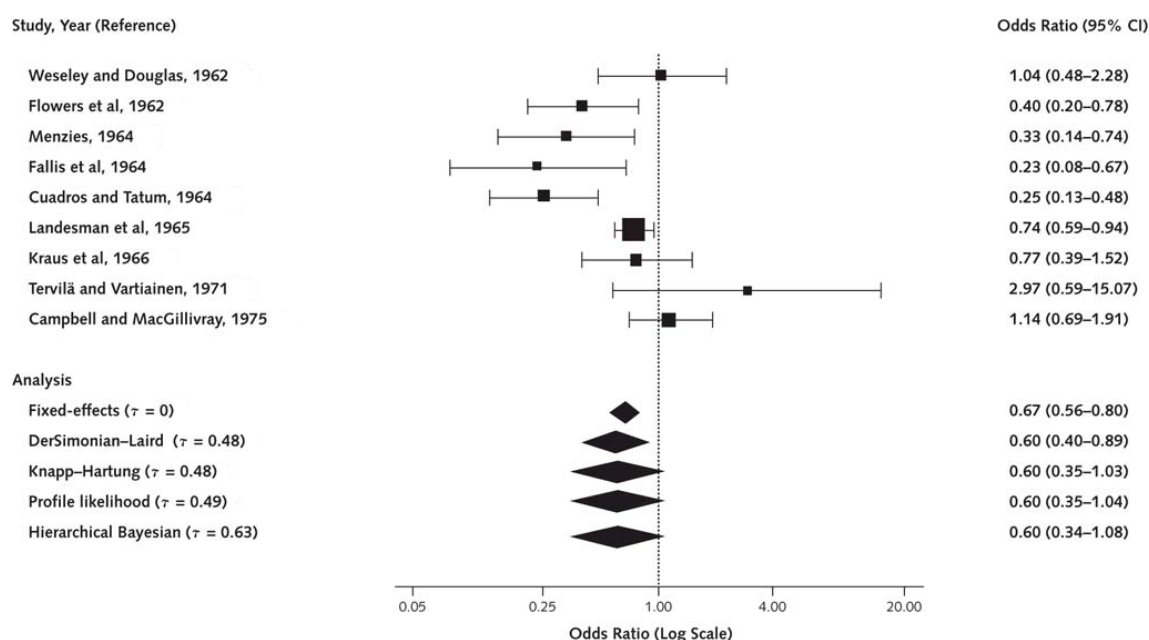


図 5. 固定効果モデルと各ランダム効果モデルによる結果の違い。解析結果は上から固定効果モデル、DerSimonian-Laird 法、Hartung-Knapp 法、Skovgaard 補正プロファイル尤度法、階層 Bayesian モデルを算定されるランダム効果パラメータ (τ) とともに示す。図のダイヤモンドは信頼区間 (Bayesian モデルのみ信用区間) を表し、予測区間ではないことに注意 (Cornell JE, et al. Ann Intern Med 2014. American College of Physicians から許諾を得て掲載)

③統計学的異質性の定量

異質性の絶対的な大きさをアウトカム指標のスケール値でそのまま表すパラメータである τ に加え、異質性がメタアナリシスにおよぼす影響をスケール値にかかわらず 0～100% の範囲で示す I^2 (取り得る範囲)^{12,13)}を算定する。 I^2 の評価にはコクラングループが提唱する、0～40% (not important)、30～60% (moderate)、50～90% (substantial)、75～100% (considerable)

の尺度¹⁴⁾を参考にするが、必ず信頼区間¹²⁾を併記し不確実性を検討する。

④階層 Bayesian モデル

ベイズ統計によるメタアナリシスは、既存の証拠（事前情報）に新たなデータを組み込みながら尤度関数を介して更新し、事後情報として新たな証拠を獲得する方法である¹⁵⁾。不確実性をより適切に計算でき、ネットワーク・メタアナリシスや多変量メタアナリシス等、複雑かつ発展的な定量的統合法への応用に優れるとして、米国 Agency for Healthcare Research and Quality (AHRQ)、英国 The National Institute for Health and Clinical Excellence (NICE)、ドイツ Institute for Quality and Efficiency in Healthcare (IQWiG) 等、先進国の公的な医療技術評価機構に積極的に採用されてきている^{2,15,16)}。当文献レビュー委員会も推奨^{2,4,6)}に従い、適切と判断されるケースには積極的に階層 Bayesian モデルを実施する。事前情報の設定法、モデル収束の判定は既報のガイドラインに従う^{2,15,17)}。

⑤ネットワーク・メタアナリシスと多変量メタアナリシス

がん検診の介入では検診効果を比較したい 3 つ以上の競合するスクリーニング方法が存在する場合も想定される。特定の 1 組の介入比較を従来型の「ペアワイズ」のメタアナリシスで 1 組ずつ評価する方法はエビデンスの全体性を適切に考慮できない可能性がある。競合する全介入方法から理論的に可能なすべての比較の組み合わせを同時に単一のモデルで解析評価するネットワーク・メタアナリシスが近年、医療技術評価にも積極的に採用されてきている^{2,15,17,18)}。また、互いに関連しあうアウトカム（例：前がん病変の発症→がんの発症→がん死亡）を個別に検討するのも同様であり、個々のアウトカムの相互関係を表すパラメータを適切に設定しながら、互いに関連しあう複数のアウトカムを同時に単一モデルで解析する多変量メタアナリシスも応用されつつある^{15,18)}。本委員会も適切と判断されるケースには積極的にこれらのメタアナリシスを実施する。

⑥スクリーニング精度のメタアナリシス

感度と特異度は診断閾値を変化させると負の相関関係が観察されることが多い。従って、精度のメタアナリシスは互いに関連しあう 2 つのアウトカムの相互関係を同時に扱う多変量（典型的には 2 変量として扱う）ランダム効果メタアナリシスを実施する^{19,20)}。検査のパフォーマンスと診断閾値をパラメータとした階層統合 ROC モデル²¹⁾は感度と特異度の 2 変量ランダム効果メタアナリシスと同様の統合が可能であり、共変数を追加しない場合には容易にパラメータ変換が可能である²²⁾。従って、どちらのモデルからも統合感度・統合特異度、階層統合 ROC 曲線の推定が可能である。

精度比較は 2 つ（以上）の競合するスクリーニング方法を全対象者に実施して直接比較を行う paired design 研究に限定することが推奨されている²³⁾。両者を通常の 2 変量モデルに

あてはめ、メタ回帰分析を応用した単純な方法²⁴⁾ から、発展的な方法に基づき複数検査の直接比較（およびネットワーク・メタアナリシス）が同時に可能な方法もある²⁵⁻³¹⁾。

文献

1. Verbeek J, Ruotsalainen J, Hoving JL. Synthesizing study results in a systematic review. *Scand J Work Environ Health*. 2012;38:282–290.
2. Morton SC, Murad MH, O'Connor E, Lee CS, Booth M, Vandermeer BW, Snowden JM, D'Anci KE, Fu R, Gartlehner G, Wang Z, Steele DW. Quantitative Synthesis – An Update: Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Scientific Resource Center under Contract No. 290-2012-0004-C). AHRQ Publication No. 18-EHC007-EF. Rockville, MD, Agency for Healthcare Research and Quality; 2018.
3. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rücker G, Harbord RM, Schmid CH, Tetzlaff J, Deeks JJ, Peters J, Macaskill P, Schwarzer G, Duval S, Altman DG, Moher D, Higgins JP. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343.d4002.
4. Cornell JE, Mulrow CD, Localio R, Stack CB, Meibohm AR, Guallar E, Goodman SN. Random-Effects Meta-analysis of Inconsistent Effects: A Time for Change. *Ann Intern Med*. 2014;160:267–270.
5. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342:d549.
6. Veroniki AA, Jackson D, Bender R, Kuss O, Langan D, Higgins JPT, Knapp G, Salanti G. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res Synth Methods*. 2019;10:23–43.
7. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177–188.
8. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med*. 2001;20:3875–3889.
9. Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med*. 2002;21:3153–3159.
10. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med*. 2003;22:2693–2710.
11. Guolo A. Higher-order likelihood inference in meta-analysis and meta-regression. *Stat Med*. 2012;31:313–327.
12. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539–1558.
13. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557–560.

14. Editors: Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2. [updated September 2009]. The Cochrane Collaboration. 2008. <https://handbook-5-1.cochrane.org/v5.0.2/>. Accessed May 7, 2019.
15. Welton NJ, Sutton AJ, Cooper NJ, Abrams KR, Ades AE. Introduction to decision models. In: Evidence Synthesis for Decision Making in Healthcare. Chichester, Wiley; 2012.
16. Bender R: Working Group "Bayes Methods". Applications of Bayesian methods in health technology assessment. Göttingen, Institute for Quality and Efficiency in Health Care; 2018.
17. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials (last updated April 2014). National Institute for Health and Clinical Excellence; 2011. <http://www.nicesdsu.org.uk>. Accessed August 5, 2017.
18. Riley RD, Jackson D, Salanti G, Burke DL, Price M, Kirkham J, White IR. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ*. 2017;358:j3932.
19. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwindermana AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58:982–990.
20. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol*. 2006;59:1331–1332.
21. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20:2865–2884.
22. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*. 2007;8:239–251.
23. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Ann Intern Med*. 2013;158:544–554.
24. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0. The Cochrane Collaboration, 2010. <http://srdta.cochrane.org/>. Accessed May 7, 2019.
25. Menten J, Lesaffre E. A general framework for comparative Bayesian meta-analysis of diagnostic studies. *BMC Med Res Methodol*. 2015;15:70.
26. Trikalinos TA, Hoaglin DC, Small KM, Terrin N, Schmid CH. Methods for the joint meta-analysis of multiple tests. *Res Synth Methods*. 2014;5:294–312.
27. Hoyer A, Kuss O. Meta-analysis for the comparison of two diagnostic tests to a common gold standard: A generalized linear mixed model approach. *Stat Methods Med Res*. 2018;27:1410–1421.

28. Nyaga VN, Aerts M, Arbyn M. ANOVA model for network meta-analysis of diagnostic test accuracy data. *Stat Methods Med Res.* 2016;27:1766–1784.
29. Nyaga VN, Arbyn M, Aerts M. Beta-binomial analysis of variance model for network meta-analysis of diagnostic test accuracy data. *Stat Methods Med Res.* 2016;27:2554–2566.
30. Dimou NL, Adam M, Bagos PG. A multivariate method for meta-analysis and comparison of diagnostic tests. *Stat Med.* 2016;35:3509–3523.
31. Cheng W. *Network Meta-Analysis of Diagnostic Accuracy Studies*. Providence, RI: Brown University; 2016.

11) サマリーテーブルの作成

がん検診の有効性評価に用いるアウトカムには、本章の冒頭で述べた評価の基本方針に示したように、中間アウトカムである検診検査の感度・特異度から、個人にとって重要で中間的なアウトカムであるがん罹患、さらに究極的なアウトカムであるがん死亡まで段階的である。GRADE グループはシステマティックレビューの結果から得られる証拠の質と効果に関する定量的な結果を多角的に考察し、エビデンスの全体性を効率よく提示する方法をアウトカム¹⁻⁵⁾と検査の感度・特異度⁶⁻⁹⁾について提唱している。当レビュー委員会が行うエビデンスレビューもこれに準じてサマリーテーブルを作成し、検診による効果を系統的に提示し、証拠のまとめを判断する基礎データとする。

11-1) エビデンスの質の評価¹⁰⁾

システマティックレビューの結果から、各アウトカムについて研究の質、結果の一貫性、結果の一般化、結果の不確実性、出版バイアスの5項目について評価を行い¹⁾、アウトカムごとに全体としての証拠の質を4段階（高い、中等度、低い、非常に低い）で評価する¹⁰⁾。以下にまとめる内容はアウトカムについての一般的な評価方法であり、検診検査の感度・特異度について（特に結果の不確実性と出版バイアスについて）は最新のGRADE シリーズ原著を参照のこと。

11-1-1) 研究の質¹¹⁾

本章8) で扱ったコクラングループが推奨している標準的な研究の質評価ツール（Risk of Bias 評価ツール）を使用し、評価する。研究デザインに応じて特異的なツールを利用する。

11-1-2) 結果の一貫性¹²⁾

本章9) で扱った臨床的な異質性と本章10) で扱った統計学的な異質性を十分吟味し、結果の一貫性について評価する。統計学的な異質性がある場合には、それを説明する要素の検討が必要である。

11-1-3) 結果の一般化¹³⁾

第一に、比較したい検診検査（介入）がネットワーク・メタアナリシス等の間接比較に由来する証拠の場合、結果の適応には直接の証拠ではない点を十分認識する必要がある。第二に、研究結果を直接利用できない可能性が懸念される状況には、①検診検査（介入）の内容が臨床課題と異なる場合、②究極のアウトカムであるがん死亡が評価されておらず中間アウトカムの評価が必要になる場合、③臨床設定が明らかに異なる場合（患者や明らかな高リスク者に対象が限定されるなど、健常な一般住民の状況から逸脱が懸念される場合、あるいは低所得国での研究等は評価アウトカムによっては背景となる医療レベルが異なる可能性がある場合）などがある。これらの要素について十分検討し、研究結果の一般化が可能かどうかを判断する必要がある。

11-1-4) 結果の不確実性¹⁴⁾

定量的結果の不確実性をメタアナリシスで算定された信頼区間にて検討する。がん検診の研究では、重要なアウトカムのイベント発生数は一般的に少ない。平均効果について有意な結論が示唆される場合も累積メタアナリシスの方法を利用した **trial sequential analysis**^{15,16)} を実施し、結果が十分なイベント数のもとで有意な結果となっている（つまり「真の陽性結果」）か、慎重な解釈を要する場合がある。

11-1-5) 出版バイアス¹⁷⁾

現時点ではコクラングループからの推奨¹⁸⁾に従い、統計学的テストを実施し評価する¹⁹⁾。

11-2) 臨床的なアウトカムの結果のサマリー¹⁻⁵⁾

メタアナリシスで算定された定量的アウトカム（通常は相対指標）を利用する。相対指標のみでなく、臨床課題にあてはまるベースラインリスクから単位人数当たりの絶対指標の結果も算定して提示する。例を表 6 に示す²⁰⁾。

11-3) 検診検査の感度・特異度のサマリー⁶⁻⁹⁾

メタアナリシスで算定された統合感度・特異度を利用する。臨床課題にあてはまるターゲット病変の有病率から単位人数当たり（1,000 人対等）に検診を行った場合の真陽性数、偽陽性数、偽陰性数、真陰性数を算定し、絶対数として検診結果を示す。例を表 7 に示す²⁰⁾。臨床的なアウトカムの証拠がない場合、さらに決断分析モデルを構築し、臨床アウトカムを仮想的に評価する方法も提唱されている⁹⁾。

表 6. 臨床的なアウトカムの結果のサマリー（子宮頸がん検診エビデンスレポート 2018 年度より引用）

一般集団における細胞診スクリーニングと比較したHPV検査によるスクリーニングの効果

（Human Development Index 2017でvery highに相当する先進国に限定）

アウトカム	研究の質					結果のまとめ		相対リスク (95% 信用区 間)	絶対イベント数		level of evidence
	参加者数 (RCT 数), 人	経過観察中央 値 (範囲), 年	研 究 の 質 (Quality)	一 貫 性 (Consistency)	一 般 化 (Generalizability)	参加者数 (イベント率, %)			絶対イベント数		
						細胞診	HPV		対照 (細胞診), 1,000,000 人年	リスク差, <i>per</i> 1,000,000 人年	
浸潤がん発症											
全研究	374,874 (6)	6 (3.6–12.0)	High	Low	High	94/180,992 (0.052)	69/193,882 (0.036)	HR, 0.69 (0.42–1.08)	112 ^{¶¶}	35人減少 (65人 減少から9人増 加)	MODERATE
HPV細胞診 併用	122,718 (4)	6.8 (6–12.0)	High	Low	High	56/55,152 (0.102)	38/125,840 (0.030)	HR, 0.57 (0.27–1.11)	112 ^{¶¶}	48人減少 (82人 減少から12人増 加)	MODERATE
HPV単独	252,156 (2)	4.5 (3.6–7.8)	High	Low	High	37/67,566 (0.055)	32/126,316 (0.025)	HR, 0.86 (0.38–2.00)	112 ^{¶¶}	16人減少 (70人 減少から112人 増加)	MODERATE
CIN3+検出											
全研究	391,311 (7) ^{¶¶¶¶}	6.4 (3–11.4)	Moderate	Low	High	994/189,326 (0.525)	1,225/201,985 (0.606)	HR, 1.10 (0.89–1.36)	1,266 ^{¶¶¶}	127人増加 (139人 減少から456人 増加)	LOW
HPV細胞診 併用	122,316 (4) ^{¶¶¶¶}	7.5 (4–11.4)	Moderate	Low	High	573/54,966 (1.042)	759/67,350 (1.127)	HR, 0.99 (0.78–1.27)	1,266 ^{¶¶¶}	13人減少 (279人 減少から342人 増加)	LOW
HPV単独	268,995 (3) ^{¶¶¶¶}	3.6 (3–6.4)	Moderate	Low	High	371/134,360 (0.276)	466/134,635 (0.346)	HR, 1.26 (0.76–2.12)	1,266 ^{¶¶¶}	329人増加 (304人 減少から1,418人 増加)	LOW

HR = hazard ratio, RCT = randomized controlled trial

^{¶¶}欧州で実施された4件の無作為化比較試験のプール解析に報告された細胞診群のイベント率 (Ronco G, et al. Lancet 2014) による

^{¶¶¶}採用7研究の累積イベントから推定

^{¶¶¶¶}無作為化を受けた全参加者から研究本来の採用基準に該当する健常対象者について行われた "modified intention-to-analysis" 集団のうち該当アウトカムについて検討された総参加者数

表 7. 感度・特異度の結果のサマリー（子宮頸がん検診エビデンスレポート 2018 年度より引用）²⁰⁾

検診方法	HC2	全 HPV 検査	HC2・細胞診併用	細胞診	備考
コホート	19 コホート	23 コホート	6 コホート	16 コホート	
対象数	107,580 人	164,577 人	19,113 人	125,532 人	
エビデンス全体の質	MODERATE	MODERATE	LOW	MODERATE	
統合感度 (95%信頼区間)	0.886 (0.786–0.945)	0.885 (0.802–0.938)	0.985 (0.780–0.999)	0.635 (0.492–0.760)	
統合特異度 (95%信頼区間)	0.902 (0.872–0.927)	0.904 (0.879–0.924)	0.844 (0.684–0.932)	0.947 (0.915–0.967)	
絶対リスク（検診 1,000 人当たりの検査結果）					CIN2+有病率 2.0%
真陽性 (95%信用区間), 人	18 (16–19)	18 (16–19)	20 (16–20)	13 (10–15)	適切な CIN2+の診断と治療機会の提供
細胞診との差	5 人増加	5 人増加	7 人増加	—	
偽陰性 (95%信用区間), 人	2 (1–4)	2 (1–4)	0 (0–4)	7 (5–10)	CIN2+にもかかわらず誤った安心と適切な検査と治療機会の喪失
細胞診との差	5 人減少	5 人減少	7 人減少	—	
偽陽性 (95%信用区間), 人	96 (72–125)	94 (74–119)	153 (67–310)	52 (32–83)	健常にもかかわらず誤った不安と不要な侵襲的検査と治療の提供
細胞診との差	44 人増加	42 人増加	101 人増加	—	
真陰性 (95%信用区間), 人	884 (855–908)	886 (861–906)	827 (670–913)	948 (897–948)	適切な健常の確認と安心の提供
細胞診との差	44 人減少	42 人減少	101 人減少	—	

文献

1. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008;336:995–998.
2. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64:383–394.
3. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, Atkins D, Kunz R, Montori V, Jaeschke R, Rind D, Dahm P, Akl EA, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Schünemann HJ. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol*. 2013;66:151–157.
4. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol*. 2013;66:158–172.
5. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, Johnston BC, Karanickolas P, Akl EA, Vist G, Kunz R, Brozek J, Kupper LL, Martin SL, Meerpohl JJ, Alonso-Coello P, Christensen R, Schunemann HJ. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. *J Clin Epidemiol*. 2013;66:173–183.
6. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams JW Jr, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt GH; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336:1106–1110.
7. Hsu J, Brozek JL, Terracciano L, Kreis J, Compalati E, Stein AT, Fiocchi A, Schünemann HJ. Application of GRADE: Making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implement Sci*. 2011;6:62.
8. Schünemann HJ, Mustafa R, Brozek J, Santesso N, Alonso-Coello P, Guyatt G, Scholten R, Langendam M, Leeftang MM, Akl EA, Singh JA, Meerpohl J, Hultcrantz M, Bossuyt P, Oxman AD; GRADE Working Group. GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *J Clin Epidemiol*. 2016;76:89–98.
9. Schünemann HJ, Mustafa RA, Brozek J, Santesso N, Bossuyt PM, Steingart KR, Leeftang M, Lange S, Trenti T, Langendam M, Scholten R, Hooft L, Murad MH, Jaeschke R, Rutjes A, Singh J, Helfand M, Glasziou P, Arevalo-Rodriguez I, Akl EA, Deeks JJ, Guyatt GH; GRADE Working Group. GRADE guidelines: 22. The GRADE approach for tests and strategies-from test accuracy to patient-important outcomes and recommendations. *J Clin Epidemiol*. 2019;111:69–82.
10. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines: 3. Rating the quality of evidence. *J Clin*

Epidemiol. 2011;64:401–406.

11. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J, Schünemann HJ. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol*. 2011;64:407–415.
12. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *J Clin Epidemiol*. 2011;64:1294–1302.
13. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *J Clin Epidemiol*. 2011;64:1303–1310.
14. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol*. 2011;64:1283–1293.
15. Miladinovic B, Hozo I, Djulbegovic B. Trial sequential boundaries for cumulative meta-analyses. *Stata Journal*. 2013;13:77–91.
16. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. *BMC Med Res Methodol*. 2017;17:39.
17. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, Williams JW Jr, Meerpohl J, Norris SL, Akl EA, Schünemann HJ. GRADE guidelines: 5. Rating the quality of evidence--publication bias. *J Clin Epidemiol*. 2011;64:1277–1282.
18. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rücker G, Harbord RM, Schmid CH, Tetzlaff J, Deeks JJ, Peters J, Macaskill P, Schwarzer G, Duval S, Altman DG, Moher D, Higgins JP. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002.
19. Murad MH, Chu H, Lin L, Wang Z. The effect of publication bias magnitude and direction on the certainty in evidence. *BMJ Evid Based Med*. 2018;23:84–86.
20. 子宮頸がん検診文献レビュー委員会. 子宮頸がん検診エビデンスレポート 2018 年度版. <http://canscreen.ncc.go.jp/>. Accessed May 7, 2019.

VII. 証拠のまとめ

がん検診の科学的根拠の証拠のまとめとして、方法別に **Body of Evidence** および利益と不利益の対比について評価を行い、それぞれについて証拠のレベルを決定する。利益の証拠のレベルおよび利益と不利益の対比が、システマティックレビューの最終的な結論となる。

1) 利益の評価

がん検診の利益は、対象となるがんの死亡率減少効果を証明することである。従って、研究のアウトカム指標はがん死亡率だが、子宮頸がん検診では浸潤がん罹患率を用いる。証拠のレベルは、個別研究の評価ではなく、死亡率減少効果に関する研究の集合体である **Body of Evidence** として評価する。科学的根拠の信頼性は、研究デザインに依存するが、無作為化比較試験の実施が困難な状況下では、観察研究の結果のみの判断もありうる。証拠のレベルは、研究デザイン、研究数、研究の質 (**Risk of Bias**)、メタアナリシスの結果から判断する (表 8)。これらの結果を総合し、信頼性が高い (**HIGH**)、中等 (**MODERATE**)、低い (**LOW**) の 3 段階に分類する。最終アウトカム指標を用いた無作為化比較試験は 1 件であっても研究の質が高い場合には、信頼性が高い (**HIGH**) と評価できる。また **Risk of Bias** のすべての項目で研究のバイアスが小さいと判断された場合は信頼性が高い (**HIGH**) と判断できる。すべての項目で研究のバイアスが大きいと判断された場合は信頼性が低い (**LOW**) と判断する。検診方法は科学的根拠の集合体である **Body of Evidence** として評価され、一元的な評価基準の設定はできない。従って、研究の質は対象となる検診方法間での相対的評価であり、絶対的な基準は存在しない。また、異なるがん検診間での比較はできない。

表 8. **Body of Evidence** の評価

Body of Evidence : 証拠のレベル		研究デザイン	研究数	Body of Evidence : 研究の質 (Risk of Bias)
HIGH	証拠の信頼性が高い	無作為化比較試験	1件以上	研究の質が高い (high)
MODERATE	証拠の信頼性が中等度			研究の質が中等度 (moderate)
LOW	証拠の信頼性が低い			研究の質が低い (low)
MODERATE	証拠の信頼性が中等度	観察研究 (コホート研究・症例対照研究)	3件以上	研究の質が中等度 (moderate)
LOW	証拠の信頼性が低い			研究の質が低い (low)

2) 利益と不利益の対比

抽出された研究の死亡率減少効果について、あり (Positive)、なし (Negative)、不明 (Insufficient) を判断する (表 9)。死亡率減少効果に関する研究が存在しない場合は、不明 (Insufficient) に相当する。

利益の有無にかかわらず、すべてのがん検診に不利益は存在する。このため、利益が認められない場合には「利益はなく、不利益あり (NEGATIVE)」、利益が不明な場合には「利益は不明だが、不利益あり (NEGATIVE)」と判断する。両者はともに、がん検診として実施する科学的根拠に欠け、不利益のみをもたらす可能性が高い。

がん検診としての利益が確認された場合には、利益と不利益の対比を検討する。すべての不利益を包含することは不可能である。従って、偽陽性率を代替指標とするが、場合によっては要精検率を用いる場合もありうる。利益については、Body of Evidence の統合結果としてメタアナリシスの結果を用いるか、あるいは代表的研究の結果を用いる。証拠のレベル (利益) の信頼性が低い場合には、「利益は希薄だが、不利益あり (BORDERLINE)」とする。証拠のレベル (利益) の信頼性が中等度以上の場合には、一定数の受診者が特定の検診を受けた場合の利益と不利益を比較する。ただし、両者の比較に絶対的な基準はなく、対象となる検診方法間での相対的な評価となる。また、利益と不利益の大きさを同等に評価することは困難なことから、最終的に不利益の大きさに着目し、「利益はあるが、不利益小 (HIGH)」「利益はあるが、不利益中等度 (MODERATE)」「利益はあるが、不利益大 (LOW)」の3段階に分類する。

表 9. エビデンスレベルの定義

進行がん罹患率減少効果	Body of Evidence	利益と不利益の対比		
あり (Positive)	証拠の信頼性が高い (HIGH)	利益はあるが、 不利益小 (HIGH)	利益はあるが、 不利益中等度 (MODERATE)	利益はあるが、 不利益大 (LOW)
	証拠の信頼性が中等度 (MODERATE)			
	証拠の信頼性が低い (LOW)			
		利益は希薄だが、不利益あり (BORDERLINE)		
なし (Negative)	証拠の信頼性が高い (HIGH)	利益はなく、不利益あり (NEGATIVE)		
	証拠の信頼性が中等度 (MODERATE)			
	証拠の信頼性が低い (LOW)			
不明 (Insufficient)	証拠の信頼性が低い (LOW)	利益は不明だが、不利益あり (NEGATIVE)		

利益と不利益の対比 (図 6)、1 万人の女性が子宮頸がん検診を受診した場合を示したフロー (図 7) である。フロー図より、要精検者数と救命数 (図 7 では浸潤がん発症数) とが比較できる。

図 6. 利益・不利益の対比 (子宮頸がんの例)

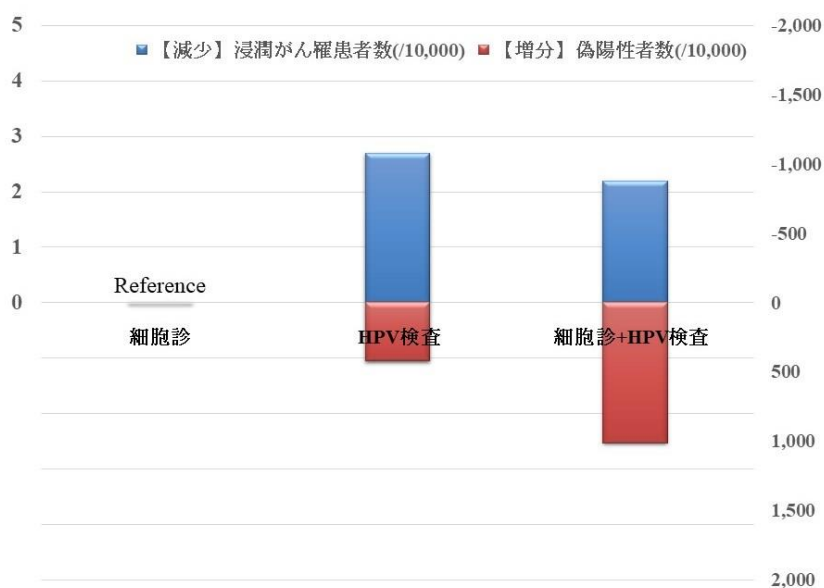
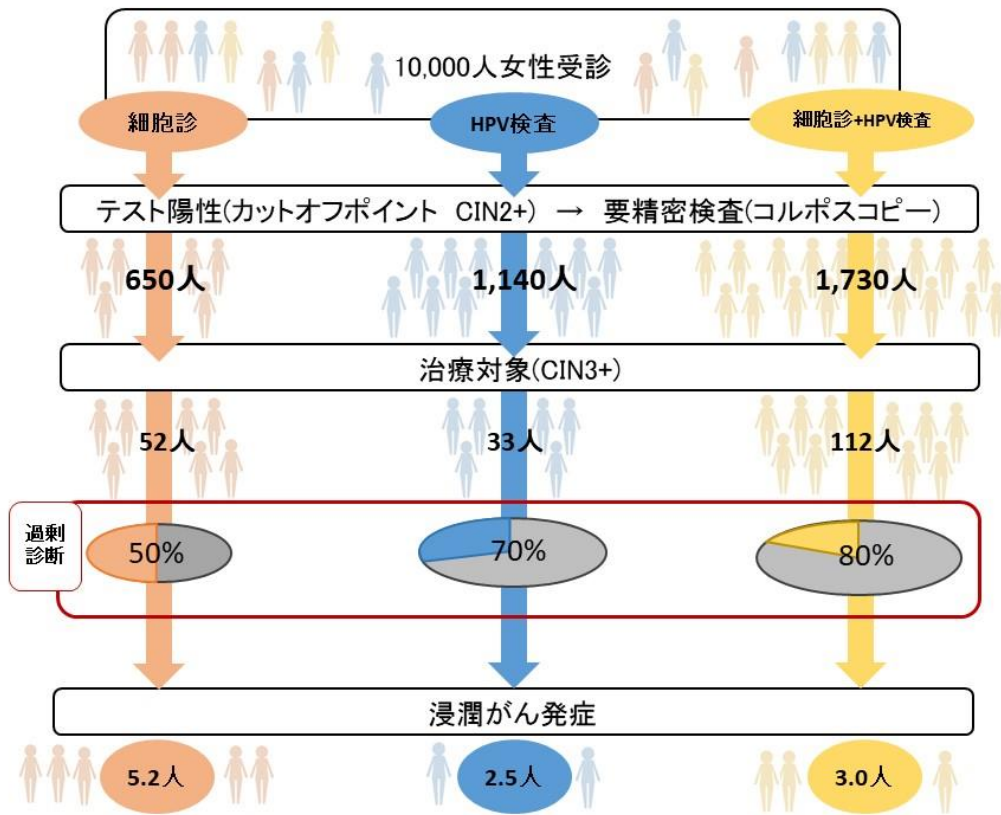


図 7. 子宮頸がん検診の方法別フロー（子宮頸がんの例）



VIII. 対象年齢と検診間隔

1) 対象年齢

がん検診の対象年齢の設定は、疾病負担を示す対象となるがんの罹患率・死亡率、有効性、不利益が参照される。U.S. Preventive Services Task Force ではモデル分析や NNT(Number Needed to Treat)等を用いて対象年齢が決定されており、これらは必ずしも標準的な方法ではないが、一部ガイドライン作成機関では採用している。このほか、対象年齢の決定には医療資源や社会的文化的背景の影響もある。

本稿においては、科学的見地から、対象年齢を検討する。有効性、不利益については、最終的に両者のバランスを検討する。不利益については、がん検診により収集できる情報は異なるが、偽陽性、過剰診断、偶発症等を考慮する。

がん検診の有効性は、死亡率減少効果（子宮頸がん検診では浸潤がん罹患率減少効果）を検討した研究から、その対象年齢をもとに判断する。ただし、複数の研究がカバーする年代であること、さらには有効性の有無について研究結果の一貫性があることをもって、至適対象年齢を検討する。

- ① 複数の無作為化比較試験があり、それらの研究の多くがカバーする年代は、死亡率減少効果（子宮頸がん検診では浸潤がん罹患率減少効果）を認めると判断する。
- ② 症例対照研究およびコホート研究について、特定の年代の死亡率減少効果（子宮頸がん検診では浸潤がん罹患率減少効果）を検討した研究を抽出し、複数の研究で効果が確認できる年代を対象年齢に取り込む。

不利益について年代別のデータが得られるものは偽陽性率の代替指標である要精検率を用いて検討する。要精検率の年代間格差を確認する。

対象年齢は、死亡率減少効果が確認された年代について、要精検率の年代間格差を比較したうえで、合議により適切な対象年齢を決定する。両者の比較についてはNNTを用いることもあるが、判断基準の閾値は定めない。対象年齢の決定には、また、年齢別のがんの罹患率・死亡率も参照する。

2) 検診間隔

対象年齢とともに、検診間隔も死亡率減少効果（子宮頸がん検診では浸潤がん罹患率減少効果）とともに検討されている場合が多い。検診間隔を延長することは、検診回数を減少させることに繋がり、不利益を減少させることができる。

死亡率減少効果と合わせて検診間隔を検討している場合には、両者を検討する。

- ① 検診間隔を延長しても死亡率減少効果が同等の場合、検診間隔を延長する。
- ② 検診間隔を延長することにより、死亡率が減少する場合、その大きさと不利益（特に要精検率）を比較する。両者の比較にはNNTを用いることもあるが、判断基準の閾値は定めない。

一方、現在評価の確立した方法と比較したうえで、検診間隔の延長の可能性は、子宮頸が

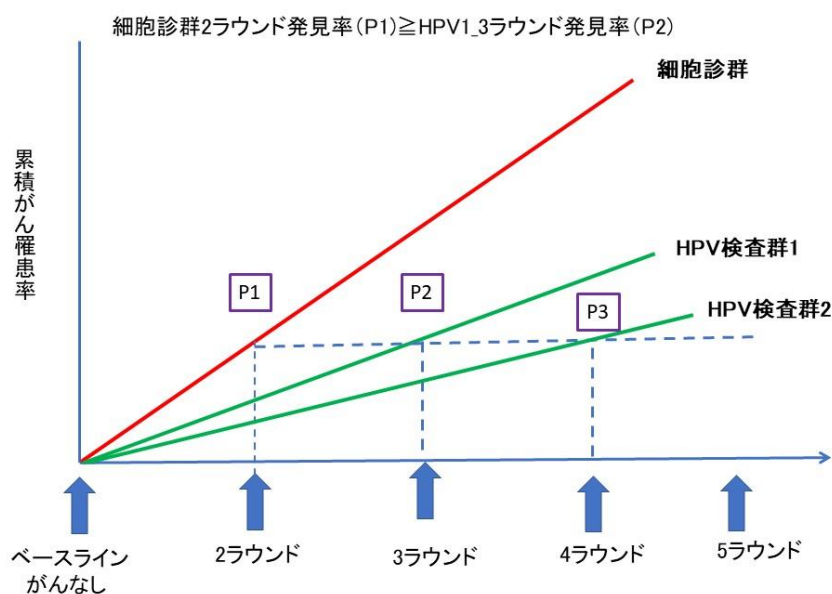
ん検診の場合、無作為化比較試験あるいはコホート研究において、初回陰性者を追跡調査し、2ラウンド以降の累積罹患率をもとに、検診間隔の延長を検討している^{1,2)}。

HPV 検査による子宮頸がん検診の検診間隔延長検討のイメージを図 8 に示した。細胞診の検診間隔は通常 1～3 年で設定されている。ベースラインでがんなしと判定された HPV 検査、細胞診を定期的に行い、2 ラウンド以降の累積発見率を比較検討する。

- ① HPV 検査群、細胞診群の 2 ラウンドの累積がん罹患率を比較する。細胞診群の累積発見率は HPV 検査群に比べて高い。細胞診に比べ、HPV 検査の感度は高く、先取りしている可能性が高いからである。
- ② 3 ラウンドの HPV 検査群の累積がん罹患率と 2 ラウンドの細胞診群の累積発見率を比較する。細胞診群 2 ラウンド発見率 (P1) と HPV1_3 ラウンド発見率 (P2) が同等であれば、HPV 検査群の検診間隔は細胞診の通常の検診間隔の 2 倍 (2～6 年) に延長できる。
- ③ 同様に、4 ラウンドの HPV 検査群の累積がん罹患率と 2 ラウンドの細胞診群の累積発見率を比較する。細胞診群 2 ラウンド発見率 (P1) と HPV1_4 ラウンド発見率 (P3) が同等であれば、HPV 検査群の検診間隔は細胞診の通常の検診間隔の 3 倍 (3～9 年) に延長できる。

検診間隔を延長しても同等の効果が確認できる場合、現在の検診間隔を延長できる。累積罹患率の比較方法は、前がん病変も含めて検診の標的病変とする子宮頸がん検診や大腸がん検診で利用できるが、他のがん検診については検討が必要である。また、モデル解析の検討も求められる。

図 8. 検診間隔延長検討のイメージ図



文献

1. Katki HA, Kinney WK, Fetterman B, Lorey T, Poitras NE, Cheung L, Demuth F, Schiffman M, Wacholder S, Castle PE. Cervical cancer risk for women undergoing concurrent testing for human papillomavirus and cervical cytology: a population-based study in routine clinical practice. *Lancet Oncol.* 2011;12:663–672.
2. Dijkstra MG, van Zummeren M, Rozendaal L, van Kemenade FJ, Helmerhorst TJ, Snijders PJ, Meijer CJ, Berkhof J. Safety of extending screening intervals beyond five years in cervical screening programmes with testing for high risk human papillomavirus: 14 year follow-up of population based randomised cohort in the Netherlands. *BMJ.* 2016;355:i4924.

IX. ガイドラインのための付加情報

エビデンスレポート本体の課題である「利益（死亡率減少効果）」「利益と不利益の対比」のほか、ガイドライン作成の推奨に必要な情報として「医療資源・費用」「受診者の価値観」を補足的に検討する。情報収集は、これまでの文献レビューと同様の方法をとるか、主たる研究やレビューを主体としたサイテーションリサーチのいずれかを用いる。ただし、「医療資源・費用」は該当する検診方法に関する費用効果分析の文献レビューを行い、必要に応じて CHEERS¹⁾ を用いた公表論文の妥当性の評価も行う。

1) 医療資源・費用

新技術を用いたがん検診の実施に影響する、医療機関、人材、医療機器等の医療資源の実態を把握するとともに、該当するがん検診の費用効果分析を行い、医療資源配分を検討する。エビデンスレポート作成過程で、追加的に費用効果分析を行うことはないが、先進国で行われた費用効果分析でおおむね費用効果的と判断された場合には、わが国でも同様の成果が期待できると判断する。

①医療資源

- 官庁統計による医療資源の把握（医療従事者数、医療機器、施設数等）。
- 文献レビューによるがん検診に資する医療資源を確認する。
- 文献レビューによる医療資源とがん検診受診率との関連を確認する。

②費用効果分析

- 文献により、がん検診の費用効果分析を抽出する。
- 抽出された研究について、CHEERS¹⁾ を用いて報告の妥当性を評価する。
- 研究結果を比較し、わが国におけるがん検診実施の可能性を検討する。

2) 受診者の価値観

新技術をがん検診に導入する場合には、受容の可能性や受診率への影響を検討する必要がある。受容に関する質的研究については、新技術への positive な面と negative な面の両者ともに把握する必要がある。

- 新技術の導入について、受容の可能性を検討した QOL 研究や質的研究の文献レビューを行う。
- 新技術と受診率の変動を検討した研究について文献レビューを行う。

文献

1. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E; CHEERS Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Value Health. 2013;16:e1–5.

X. エビデンスレポートの作成と公開

エビデンスレポートは、AF と検討課題、文献抽出過程と検討課題ごとの科学的根拠のまとめ、証拠のまとめを含む。ただし、対象年齢や検診間隔等、がん検診固有の課題についても同様の検討を行う場合がある。補足情報として、文献検索式および主要な研究に関する構造化要約を添付する。

ガイドライン作成委員会は、エビデンスレポート・ドラフトをもとに会議を運営し、最終的な推奨グレードを決定する。推奨グレード決定に必要となる科学的根拠や追加情報を求める場合があるが、文献レビュー委員会はその要求に応じ、追加的な検討を行い、ガイドライン作成を支援する。また、ガイドライン作成委員会の開催には、文献レビュー委員会のメンバーが臨席し、科学的根拠の評価過程や内容の説明を行う。

公開フォーラムに先立ち、国立がん研究センターの「科学的根拠に基づくがん検診推進のページ」(<http://canscreen.ncc.go.jp/>) にガイドライン・ドラフトおよびエビデンスレポート・ドラフトを公開する。公開フォーラムでは、エビデンスレポートの説明を行い、科学的根拠に関する質疑には回答する。また、エビデンスレポートにおける科学的根拠の評価に関する指摘について、適切な追加修正を行う。

最終的に、ガイドラインが国立がん研究センター社会と健康研究センター運営会議、国立がん研究センター執行役員会の承認を得て公開される同時期に、エビデンスレポートも上記ホームページと印刷物の両者を公開する。なお、個別課題の評価は、一部は英文論文として投稿するとともに、エビデンスレポートの英文版も作成し、公表する。

ガイドラインの更新は5年を目標としていることから、更新時期に合わせ、再度システマティックレビューを行う。

文献レビュー委員会 名簿

委員長

濱島 ちさと （医療技術評価、がん検診）

帝京大学 医療技術学部 看護学科 保健医療政策分野 教授

委員

寺澤 晃彦 （総合内科診療、臨床疫学）

藤田医科大学 医学部 救急総合内科学 准教授

細野 寛代 （産婦人科、がん疫学）

一般財団法人 全日本労働福祉協会 東海診療所

名古屋市立大学 大学院医学研究科 公衆衛生学分野 研究員

片山 貴文 （医療技術評価）

兵庫県立大学 看護学部 統計・情報系 教授

佐々木 清寿 （消化器内科）

聖路加国際病院附属クリニック 予防医療センター 一般内科 医長

がん検診ガイドライン事務局

中山 富雄 （疫学、がん検診）

国立がん研究センター 社会と健康研究センター 検診研究部 部長

＊（ ）内は主たる専門分野

文献レビュー委員会メンバーの利益相反

ガイドラインの作成に関わるいずれの委員は、特定の企業からの研究費等の支援を受けていない。

本レポート作成のための研究費

平成 27–28 年度 国立がん研究センター研究開発費 (26-A-30)「検診ガイドライン作成と検診提供体制の政策提言のための研究」班および平成 29–31 年度 国立がん研究センター研究開発費 (29-A-16)「検診ガイドライン作成と科学的根拠に基づくがん検診推進のための研究」班に基づく。

謝辞

以下の方々には、「有効性評価に基づくがん検診ガイドライン」のためのエビデンスレポート作成方法の作成に関わる関連業務等を担当していただきましたことに感謝いたします。
(敬称略)

星 佳芳 (衛生学・公衆衛生学)

国立保健医療科学院 研究情報支援研究センター 上席主任研究官

浅井 淳子 帝京大学 医療技術学部 看護学科 保健医療政策研究室

富永 育子 帝京大学 医療技術学部 看護学科 保健医療政策研究室

中川 由美子 帝京大学 医療技術学部 看護学科 保健医療政策研究室

樋田 記子 帝京大学 医療技術学部 看護学科 保健医療政策研究室

松島 佳乃子 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 検診研究部

Appendix : Risk of Bias 評価ツール

Appendix : Risk of Bias 評価ツール

検診無作為化比較試験

大項目	小項目	判定基準	コメント	小項目別： (おそらく) 良/(おそらく) 不良/不明	Risk of Bias (低/高/懸 念あり) と 方向 (介入 群/比較群/0 効果/予測不 可能)
1. 無作為化	無作為化割り付け	・ 良：①乱数表②コンピューターによる乱数③コイントスの裏表④シャッフルカード⑤封筒法⑥サイコロ⑦抽選/くじ⑧最小化法 ・ 不良：①交互選択②誕生日・入院日・カルテ番号による方法③臨床医の判断や参加者の好み④実施できる介入⑤系統的あるいは思いつきなど ・ 不明：上記の記載がないか判断がつかない場合			
	割り付けの隠匿	・ 良：①中央割り付け②適切な封筒法（封筒は連続番号付記＋不透明＋封印） ・ 不良：参加者が無作為化の前に割り付け群が予想されうる状況が疑われる			
	ベースラインデータの不均衡	・ 良：不均衡がない場合、あるいは不均衡があっても統計学的な偶然の範囲内の場合 ・ 不良：無作為化に問題があり、①群間のサイズに異常な違いがある場合②偶然を逸脱した数の統計学的な有意差のある不均衡③不均衡が明らかに観察された場合（特にアウトカム発症に重要な予後因子） ・ 不明：記載がないか判断がつかない場合			
	参加者のマスク化	両方のスクリーニングの実施やシャムスクリーニング等の使用による介入のマスク化が可能			

2. 意図された検診からの逸脱	介入担当者のマスク化	両方のスクリーニングの実施による介入のマスク化が可能			
	意図された介入からの逸脱（検診への参加の事実のみ）	スクリーニングおよび結果による追加介入の通常で予想される範囲を超えるコンプライアンス不良や中止、割り付けられていないスクリーニング法への移行や併用について評価			
	共介入（検診への参加の事実と継続）	追加介入（精密検査や治療）の方法に重要な差はないか			
	介入の実施成功率（検診への参加の事実と継続）	スクリーニングは意図された通りに成功して実施されているか			
	介入のアドヒアランス（検診への参加の事実と継続）	スクリーニング、精密検査、フォローアップ検査等は予定された時期・回数が順守されているか			
	逸脱から生じた結果（検診への参加の事実と継続）	上項で検討した逸脱内容がアウトカムに影響するかどうかを評価			
	適切な解析法の判定（検診への参加の事実と継続）	ITT 解析、per-protocol 解析、per-治療解析のどれがデザインにとって最適か検討			
	ITT 解析が適切な場合	ITT（実際の介入内容に関係なく、無作為化割り付け群のままでの結果解析）を評価			
	ITT 解析を実施しないことによる結果	5%を超える ITT 逸脱を重要とするが、稀なイベントの場合は逸脱の影響は大きい			

3. 結果データの逸失	全参加者（ほぼ全参加者）にデータがあるか	連続変数なら 90–95%；二区分変数の場合はイベント率による（稀なイベントの場合は影響は大きい）			
	データの逸失率と理由は両群間で同様か	統計学的な偶然を超えるものは理由を併記			
	データの逸失に対処できる信頼性の高い結果	1 次研究実施者の対処法による。さらに、1 次研究実施者の感度分析、エビデンスレポートグループの追加解析にもよる			
4. アウトカムの判定	アウトカム評価者のマスク化	不良：マスク化がない場合			
	アウトカムのバイアスを伴う評価からの保護	個々のアウトカムごとに評価が必要だが、全死亡（あらゆる原因による死亡）の場合には通常マスク化は不要			
5. 選択的な報告	マルチプル・アウトカム（結果の測定法）	プロトコールとの確認が必要 ・ 良：①計画された測定法の結果がすべて記載されている②唯一の測定法で結果も唯一③測定方法に計画との齟齬はあるが、明らかに結果と関連しない適切な理由が記載されている ・ 不良：1 つの結果に多数の測定方法が実施され、統計学的有意性等により選択された結果が報告されている可能性が高い			
	マルチプル解析（結果の解析法）	プロトコールとの確認が必要 ・ 良：①計画された解析法の結果がすべて記載されている②唯一の解析法で結果も唯一③解析方法に計画との齟齬はあるが、明らかに結果と関連しない適切な理由が記載されている ・ 不良：1 つの結果に多数の解析がされ、統計学的有意性等により選択された結果が報告されている可能性が高い			

6. その他	アウトカムの判定	アウトカムの評価：方法と質の両面から考慮①研究独自の調査方法②研究とは関係なく日常的に収集されるデータ（死亡小票、がん登録、医療記録）			
	アウトカムの重要性	アウトカムは重要か（前がん病変が発症＜浸潤がんの発症＜がん死亡＜全死亡）			
	持ち越し効果	研究開始前の持ち越し効果の可能性（例：当該検診受診歴や受診回数、別の予防・治療・検診介入）			

検診コホート研究

大項目	小項目	コメント・内容 の記述	小項目： （おそらく） はい/（おそらく）いいえ /不明	Risk of Bias（高/ 低/不明）と方向 （過大評価/過小 評価/不明）
1. 交絡因子	1.1 交絡は無視できる			
	・上問「はい」なら交絡因子全体については低リスクとして終了 ・上問「いいえ」なら検診が途中で中止や別の検診に変更になっていないか、中止・変更により生じた検診期間別に解析がされていないか検討する →検診期間に区切って解析がされていないなら 1.3 へ →検診期間に区切って解析がされていれば 1.4-1.6 へ			
	1.2 検診の中止または別検診への変更がある場合、その理由は（経時的に）変化する（アウトカムの）予後因子と関係している			
	上問について →「いいえ（変化する予後因子と無関係）」であれば 1.3、1.4、1.5 のみ検討して終了 →「はい（変化する予後因子に関連）」であれば 1.6、1.7 のみ検討して終了			
	1.3 重要な予後因子が（層別化、回帰、マッチング、標準化、逆確率重み付け推定法（IPW）等で）適切に補正されている			
	1.4 確立された交絡因子が的確な方法で測定されている			
	1.5 検診開始後に検診に影響を受ける因子も交絡補正に使用している			
	1.6 重要な予後因子を補正するだけでなく（検診の継続・変更に関わる）変化する予後因子にも補正（例：逆確率重み付け推定法（IPW）等）している			
	1.7 1.6 で交絡を検討した各因子は的確な方法で測定されている			

2. 対象者の選択	2.1 研究（あるいは解析）の対象者の選択は「検診開始後」に測定された特定の因子と関係がある			
	上問について →「いいえ（開始後因子と無関係）」であれば（2.2-2.3をとばして）2.4へ →「はい（開始後因子に関連）」であれば2.2へ			
	2.2 検診の参加に影響した開始後因子は検診自体の参加に関連がある			
	2.3 （2.2が「はい（開始後因子は検診に関連）」の場合）検診の参加に影響した開始後因子はアウトカム自体あるいはアウトカムの原因と関連がある			
	2.4 「検診の開始」と「経過観察の開始」は大多数の参加者で同時である			
	2.5 （2.2と2.3に「はい」または2.4に「いいえ」）選択バイアスが（逆確率重み付け推定法（IPW）やアウトカム欠損データを補完するモデル分析等で）適切に補正されている			
3. 検診介入の有無の判断	3.1 検診（を受けたと判断する）群の定義は検査方法、検査状況、回数、実施タイミングなど明確にしてある			
	3.2 検診を受けたと判断する情報は検診を受けたまさにその時点で記録されている			
	3.3 アウトカムの有無あるいはアウトカムの予後因子を知らない評価者が検診介入の有無を判断している			
4. 検診介入方法の逸脱	4.1 （検診への参加の事実のみを評価する場合）スクリーニングおよび結果による追加介入に通常で予想される範囲を超えるコンプライアンス不良や中止、割り付けられていないスクリーニング法への移行や併用			
	4.2 （検診への参加の事実のみを評価する場合）上記のコンプライアンス不良・中止や別の検診方法の移行・併用が比較する群間で不均衡がある			
	4.3 （検診への参加自体と継続を評価する場合）アウトカムに関わる共介入（別検診、精密検査・診断、治療の機会均等性）は比較する群間で同様である			
	4.4 （検診への参加自体と継続を評価する場合）検診は大多数の患者に適切に実施されている			
	4.5 （検診への参加自体と継続を評価する場合）検診参加の検診アルゴリズムのコンプライアンスは高く、高い継続率で、かつ比較対照の検診への移行率は低い			

	4.6 (検診への参加自体と継続を評価する場合) 4.3–4.5 のどれかにでも問題がある場合、逆確率重み付け推定法 (IPW) や操作変数法等で適切に補正がされている			
5. 結果データの逸失	5.1 アウトカムデータは (ほぼ) 全員分が記録されている			
	5.2 アウトカムデータあるいは検診介入の有無が不明のため研究あるいは解析から除外者がいる			
	5.3 補正解析に用いる因子が欠損したための除外者がいる			
	5.4 5.1–5.3 どれかに問題がある場合、除外者の理由と分布は比較する両群で同等か			
	5.5 5.1–5.3 どれかに問題がある場合、1 次研究の実施者あるいはレビュー者によって信頼のおける解析 (complete-case 解析および慎重に仮定を検討された多重代入法 (multiple imputation 等)) を含めたデータ補完による感度分析) が実施されているか			
6. アウトカム評価	6.1 アウトカムは検診参加・内容を知っている解析者に影響を受けるか (例: 全死亡、機械のみで測定する結果を事後的に変更できない場合は影響が低い)			
	6.2 アウトカム評価者は参加者の当該検診の受診・参加状態を知っている			
	6.3 アウトカムの評価方法は比較される群間で同様			
	6.4 当該検診の有無に基づく何らかの系統的エラーがありうる			
7. 選択的な結果報告	7.1 同じアウトカムで複数の測定方法が検討されている			
	7.2 同じアウトカムで複数の解析方法が実施されている			
	7.3 同じアウトカムで複数のサブグループ解析や一部の参加者の除外のもとで実施されている			
8. その他	8.1 不適切な除外: がん死亡から直前期 (例: 1 年以内) の検診を解析から除外している (除外は不適切)			
	8.2 検診回数ごとにサブグループ解析や回帰分析をして検診の効果を検討している (健康な人に検診回数が多いという evidence があり、不適切)			

検診症例対照研究

大項目	小項目	コメント・内容の記述	小項目： （おそらく）はい/（おそらく）いいえ/不明	Risk of Bias（高/低/不明）と方向 （過大評価/過小評価/不明）
1. 交絡因子	1.1 交絡は無視できる（「はい」なら交絡については低リスクとして終了）			
	1.2 交絡因子が（層別化、回帰、マッチング、標準化、逆確率重み付け推定法（IPW）のいずれかで）適切に補正されている			
	1.3 確立された交絡因子が的確な方法で測定されている			
	1.4 補正に用いた交絡因子は検診介入開始前に測定したベースライン因子のみ（検診開始後に測定した因子では補正していない）			
2. 対象者の選択	2.1 対照群は症例群を選択したのと同じ集団（あるいは同じプライマリケア医や近隣者）から選ばれた。あるいは、別の選択バイアスを回避する方法がとられた			
3. 検診内容の判定	3.1 検診介入の有無を判断する定義（実施・判定方法、地域性・医療機関、頻度、タイミングについて）は明確である			
	3.2 検診を受けたかどうかは検診を受けたときに記録されている（思い出しバイアスの懸念がありインタビューやアンケートのみの場合は「いいえ」とする）			
	3.3 アウトカムの有無（あるいはアウトカムの予後因子）を知らない評価者が検診介入の有無を判断している			
4. 意図された検診からの逸脱	4.1 アウトカムに関わる共介入（別検診、精密検査・診断、治療の機会均等性）は比較群間で同様である			
	4.2 症例群と対照群（検診なしまたは別検診）の変更（変更回数、変更の程度（部分～完全変更））による影響は少ない			
	4.3 検診の実施状況（検診提供者の実施内容、検診受診者のコンプライアンス、検診の提供状況（例：住民検診あるいは病院検診への移行はないか））は一貫している			
5. 結果データの逸失	5.1 アウトカムデータの判断は完璧である			
	5.2 検診介入の有無の判断は完璧である			
	5.3 補正解析に用いる交絡因子が欠損したために除外された人数は少ない			

	5.4 (5.1、5.2、5.3 のどれか1つでも「いいえ」の場合) データ欠損となった人数の割合や理由は対照群・症例群間で同様			
	5.5 (5.1、5.2、5.3 のどれか1つでも「いいえ」の場合) 欠損データについては適切な統計学的手法がとられているか (「はい」のためには多重代入法 (multiple imputation) のみでは十分でなく、complete-case も含めての慎重なデータ補完による感度分析が必要)			
6. アウトカムの判定	6.1 症例群 (対照群も) かどうかの判断は客観的な定義による			
	6.2 症例群 (対照群も) かどうかの判断は検診の有無がわからない状況でされている			
7. 選択的な報告	7.1 当該検診が複数の定義・分類で検討されている			
	7.2 同じアウトカムで複数の解析方法が実施されている			
	7.3 同じアウトカムで複数のサブグループ解析や一部の参加者が除外されている			
8. その他	8.1 不適切な除外: がん死亡から直前期 (例: 1 年以内) の検診を解析から除外している (除外は不適切)			
	8.2 検診回数ごとにサブグループ解析や回帰分析をして検診の効果を検討している (健康な人に検診回数が多いという evidence があり、不適切)			

検診スクリーニング精度研究

大項目	小項目	コメント・内容の記述	小項目 (はい/いいえ/不明)	バイアスのリスク (高/低/不明)	適用可能性の懸念 (高/低/不明)
1. 対象者の選択(対象の設定・研究デザイン)	1.1 連続した対象者サンプルか、または、無作為な対象者サンプルを組み入れたか				
	1.2 症例対照研究 (2 ゲート・サンプリング) デザインの使用を避けているか				
	1.3 一般的な健常者集団から「不適切な除外」を避けているか (何らかの基準で一般的な対象者の枠組みを逸脱する厳選をしている場合は「いいえ」)				
2. スクリーニング検査	2.1 スクリーニング検査の結果の解釈は、標準的な精密検査 (参照基準) の結果とマスク化で行われているか (スクリーニング検査前を精密検査の前に実施しているなら「はい」)				
	2.2 スクリーニング検査の診断基準を明確にしているか (検査のカットオフ値が事前に定義されているなら「はい」。後から最適化している場合は「いいえ」)				
3. 精密検査 (参照基準)	3.1 精密検査 (参照基準) は、標的疾患の状態を正しく分類できる (信頼性の高いもの) と考えられるか				
	3.2 精密検査 (参照基準) の結果の解釈は、スクリーニング検査の結果をマスク化して行っているか				
4. フローとタイミング (フォローアップ、検証バイアス、脱落、解析からの除外)	4.1 スクリーニング検査と精密検査 (参照基準) の実施の間隔は適切か (できるだけ近い時期に行っているか。精密検査前に治療が開始されていないか。フォローアップ検査が含まれる場合は検査内容・期間は適切か)				
	4.2 「すべての対象者」が精密検査 (参照基準) を受けているか (全員が受けているなら「はい」。受けていない者がいて、インデックス検査の結果で参照基準を受けるか分かれる場合は「いいえ」)				
	4.3 対象者は同一の精密検査 (参照基準) を受けているか (同一でない場合は「いいえ」)				

	4.4 すべての対象者が解析に含まれているか（追跡不能例、対象者数 ≠解析者数の場合は「いいえ」）				
--	--	--	--	--	--