「日本における傷病名を中心とするレセプト情報から得られる指標のバリデーションに関するタスクフォース」報告書

岩上将夫[1] 青木事成[2] 赤沢学[3] 石黒智恵子[4] 今井志乃ぶ[5] 大場延浩[6] 草間真紀子[7†] 小出大介[8] 後藤温[9] 小林典弘[10] 佐藤泉美[11] 中根早百合[2] 宮崎真[12] 久保田潔*[13]

*日本における傷病名を中心とするレセプト情報から得られる指標のバリデーションに関するタスクフォース委員長

1 London School of Hygiene and Tropical Medicine

2 中外製薬株式会社 安全性リアルワールドデータサイエンス部

3 明治薬科大学薬学部

4 医薬品医療機器総合機構 医療情報活用部 疫学課

5 国立病院機構本部総合研究センター診療情報分析部

6 日本大学薬学部

7 東京大学大学院薬学系研究科

8 東京大学大学院医学系研究科

9 国立がん研究センター社会と健康研究センター 疫学研究部

10 塩野義製薬株式会社 経営戦略本部 デジタルインテリジェンス部

11 京都大学大学院医学研究科

12 MSD 株式会社 グローバル研究開発本部 薬剤疫学グループ

13 NPO 日本医薬品安全性研究ユニット

†現所属:国立研究開発法人日本医療研究開発機構

キーワード:バリデーション、アウトカム、感度、特異度、陽性的中度、陰性的中度

連絡先: NPO 日本医薬品安全性研究ユニット 〒113-0034 東京都文京区湯島 1-2-13 御茶ノ水明神ビル 4F 久保田潔

E-mail: kubotape-tky@umin.net

<抄録>

【目的】医療情報データベース(DB)の活用が薬剤疫学、その他の分野で進んでおり、2017 年 10 月 26 日に公布され 2018 年 4 月 1 日に実施される GPSP 省令の改正でも医療情報 DB を用いた調査が、医薬品の製造販売後データベース調査として具体的に明記された。医療情報 DB のデータは通常、調査研究以外の目的で収集されている。したがって、医療情報 DB を用いる薬剤疫学研究や製造販売後調査の結果が信頼に足るかを判断する上で、医療情報 DB から得られる情報、特に傷病名の正確性をより適切に評価するためのバリデーション研究の実施が必要となる場合も想定される。日本においては、これまでバリデーション研究の報告例は限られており、その経験は不足している。そのため、日本薬剤疫学会の「日本における傷病名を中心とするレセプト情報から得られる指標のバリデーションに関するタスクフォース」(以下「本 TF」)は日本の環境下でいかにバリデーション研究を実施するかについての指針を与えることを目的に 2016 年 7 月に設立された。対象とするのは、より真に近いと考えられる情報(ゴールドスタンダード)と比較して傷病名などの正確性を評価する研究であり、システムバリデーションや疾患予後予測モデルの妥当性研究は含まれない。

【方法】バリデーション研究に言及した教科書やガイドライン、国内外で実施されたバリデーション研究を検討した。過去の研究の検討では、網羅的な系統的レビューは意図せず、タイトルに"Valid*"が含まれる論文を中心に、これまでにどのような研究が実施されていたのかを俯瞰できる程度にとどめ、(1)利用するDB の種類、(2)バリデーション研究のセッティング、(3)アウトカムの定義、(4)リンケージの有無・方法、(5)ゴールドスタンダードの定義、(6)サンプリング方法とサンプルサイズ、(7)妥当性を測定する指標、(8)指標の閾値・利用法の 8 項目を検討した。これらに基づき、主に上記 8 項目に関してバリデーション研究でとるべき手順・チェック項目を検討した。さらに日本において可能なバリデーション研究の有り方を考察した。

【結果】教科書では Strom の Pharmacoepidemiology 第 5 版 41 章にバリデーション研究に関する記載が見られた。ガイドラインでは、バリデーション研究の方法が部分的に言及されているものは見出されたが、バリデーション研究それ自体を主たる対象にするものは見出されなかった。検討した国内外 100 余りのバリデーション研究については、(1)診療報酬請求データ、電子カルテ情報のほか多様な DB が対象にされていた。(2)国・地域・被保険者集団を包括する DB からランダムサンプリングした患者を対象に行った"population-based"と考えられるバリデーション研究も見られたが、少数の医療機関を受診した患者集団や研究者が独自に設定した集団において実施されたバリデーション研究も散見された。(3)ICD コード単独、または薬・検査・診療行為との組み合わせなどがアウトカム定義に用いられていた。(4)リンケージは多くの研究でアウトカム定義の評価に必須の要素であり、多くの研究で利用されていた。(5)ゴールドスタンダードにはカルテレビューのほか、疾患登録などが使われていた。(6)サンプル方法、サンプルサイズは何をゴールドスタンダードとするかに強く依存していた。(7)感度、特異度、陽性的中度、陰性的中度の全てを求めているもののほか、陽性的中度のみを評価したものも多かった。(8)指標の閾値を示したものはほとんどなく、少数、特定の研究の目的を達成する上で必要な「感度〇%以上の指標を見出すことを目標とした」などの記載が見られた。

バリデーション研究の手順としては、「〇〇のデータベースにおいて、□□の疾患を特定する△△の傷病名（またはアルゴリズム）が妥当であることを示す」などのリサーチクエスチョンを明確にすることがまず重要である。今後の研究（バリデーション研究で評価するアウトカム定義を用いて実施する研究）に利用する DB としては、日本では診療報酬明細書（レセプト）の DB のほか、Diagnosis Procedure Combination (DPC)データ、電子カルテ情報、疾患レジストリーなどが想定される。日本では DB から元データに遡ることが困難であるため、今後の研究に利用する DB のうち一部の偏ったサンプルを対象にバリデーション研究を行わなければならないことが多い。そのような状況では、バリデーション研究における対象集団の特徴を明確にする、対象集団を重症度で分類したサブグループごとに指標の値を示す、また偽陽性・偽陰性の例の特徴を明確にすることなどにより、バリデーション研究における対象集団と異なる集団においても参考情報として利用できるよう工夫することが求められる。アウトカム定義としては傷病名マスターや ICD-10 コードを中心とし、複数の定義を評価すべきである。ゴールドスタンダードをカルテレビューで実施する時は、適切な診断基準に沿って判定が可能となるような調査票を設計する必要があり、可能な限り複数の評価者が独立に評価し、その一致度を報告すべきである。求めるべき指標として、陽性的中度だけか、感度、特異度などの評価も必要かはリサーチクエスチョンに依存する。サンプリング方法、サンプルサイズは評価すべき指標に感度が含まれ、ゴールドスタンダードがカルテレビューである時に特に問題となる。対象集団全体またはそのランダムサンプルの全例をレビューすることが基本であるが、特定のアウトカム定義を満たすか否かにより異なる抽出割合でサンプリングする方法、“all possible cases”を想定したサンプリング方法（真のケースを含むと考えられる複数のサブグループからカルテレビューで真のケースを見出し、それを対象集団に含まれる全てのケースとみなす方法）も選択可能である。指標の閾値の絶対的基準を定義することは難しく、各研究の目的により、得られた指標の値を参考にアウトカム定義が適切であるかを判断すべきである。

日本においては、バリデーション研究において重要な役割を果たすリンケージ（照合）は一般に困難であるため、病院（診療所）単位で、保管されている過去の診療報酬明細書（レセプト）を院内のカルテ情報や院内疾患登録などとリンケージ（照合）する病院（診療所）ベースのバリデーション研究が中心とならざるをえない。バリデーション研究にあたっては、DPC データまたは DPC レセプトの傷病名については医科レセプトの傷病名よりも特異性が高いことが期待され、今後のバリデーション研究でその妥当性が評価されるべきであろう。また、病院（診療所）単位のバリデーション研究以外の方法として、大規模コホート研究における国保レセプトを用いたバリデーション研究も可能な方法と考えられる。

【考察・結論】バリデーション研究はデータベース研究の質を大きく高める（特に、より適切な比較を可能にすることを通じて内的妥当性を高める）ことを可能にする。質の高いデータベース研究は、医薬品などの安全性・有効性に関して、信頼性が高くかつ他では得難い知見の獲得を可能とし、その結果は国民の健康と福祉に貢献しうる。現状、国内でのバリデーション研究の実施は困難を伴う場合が多いが、今後、医療情報 DB の適切な利用を進めようとする者は、データベース研究の有用性を高めるためにも、医療情報 DB の二次利用に対する国民的理解を得ながら、バリデーション研究を実施しやすい環境の向上に寄与することが求められる。本報告書がバリデーション研

究の意義や基本的な考え方の理解を促し、与えられた条件下で目的を達成するための最適な方法を見出す手掛かりになることを期待する。

目次

1. はじめに

診療報酬請求データ(claims data)や電子カルテ(electronic health records, EHR または electronic medical records, EMR)のような日常診療における医療情報（routinely collected data）は圧倒的な情報量を持つことが最大の強みである[(1)]。こうした医療情報、医療情報データベースを用いた薬剤疫学研究は海外では 1990 年頃より盛んに行われており、昨今、日本においても徐々に見られるようになっている。

また、国内の薬事規制においても医療情報データベースの活用が進められており、2012 年 4 月 11 日付で公表された「医薬品リスク管理計画指針について（薬食安発 0411 第 1 号、薬食審査発 0411 第 2 号）」の中では、安全性監視活動において医療情報データベースの利用可能性について言及があり、また、2017 年 10 月 26 日付で交付された「医薬品の製造販売後の調査及び試験の実施の基準に関する省令（平成 16 年厚生労働省令第 171 号）」（Good Post-marketing Study Practice, GPSP）の改正によって医薬品の再審査及び再評価の申請資料として医療情報データベースの利用が認められるようになった。こういった薬事規制上の変化に伴い、医薬品の製造販売後調査における医療情報データベースの利活用が現実的課題として上りつつある。

薬剤疫学の分野において医療情報データベースを用いた観察研究は、ランダム化比較試験のサンプルサイズ・観察期間では検出されないレベルの稀なアウトカムと薬剤の関連の検討や日常診療下における薬物使用実態の調査に適している。一方、明確なリサーチクエスチョンを背景に計画されたコホート研究や疾患レジストリーと比べると、情報の正確さや詳細さが不十分であったり、必要な交絡因子の情報が不足していたりすることが弱点である[(1)]。日本薬剤疫学会では、2012 年の「医薬品リスク管理計画指針」に関連し、日本語/英語版の「より良い Pharmacovigilance Plan 策定に向けての提言」[(2)]を発表し、日本の市販後の調査の在り方について一石を投じてきた。その提言の中で「データベース研究は安全対策の意思決定につながるエビデンスを与えることもありうるが、アウトカムなどに関する指標のバリデーションが不十分であるか、他のデータ源との連結が困難で、医療機関内の元データにもどっての確認ができない場合には次のアクション（Primary Data Collection を含む）を行うべきかを判断するためのスクリーニングの役割にとどまるだろう。」と言及している通り、データベース研究の信頼性を高める上でバリデーション研究の実施は重要なステップである。しかし、我々の知る限り、バリデーション研究についての方法論・マニュアルは国内外ともに存在しない。

そこで日本薬剤疫学会では「日本における傷病名を中心とするレセプト情報から得られる指標のバリデーションに関するタスクフォース」を立ち上げ、日本・海外において実施されてきたアウトカムなどに関するバリデーション研究及び関連するガイドライン・教科書をレビューし、日本において実施可能なバリデーション研究の手順・問題点・将来的な課題を明確にすることを目標に 2016 年 7 月から活動を行ってきた。その総括としての本報告書は、医療情報データベースを用いて薬剤疫学研究及びそのためのバリデーション研究を行う研究者を対象とするだけでなく、研究の結果を解釈し行政的な判断を行う方々が参照することも想定している。さらに本報告書に書かれたバリデーションに関する方法論・考え方は、日本のレセプト（正式名称は診療報酬明細書または調剤報酬明細書）を用いた薬剤疫学研究だけでなく、医療情報データベースを用いた臨床疫学研究に幅広く適用可能である。

2. バリデーション研究の概要

(1) バリデーション研究とは

バリデーション(validation)は「妥当性(validity)の確認」を意味する。医療情報データベースを用いて薬剤疫学研究を行う際には、研究対象集団・曝露・アウトカム・交絡因子をデータベースから得られる情報（入力された傷病名、処置名、処方内容、検査結果など）を用いて定義する必要があり、本タスクフォースではこの定義を「より真に近いと考えられる情報（ゴールドスタンダード）と比べて妥当性を確認すること」をバリデーションとし、そのために行う研究をバリデーション研究と呼ぶ。

なお、医療情報データベースなどを用いる研究においてバリデーションという言葉は、本タスクフォースの指すバリデーションとは異なる概念で使われることがあるため、注意が必要である。例えば、予後予測モデルの評価もまたバリデーションと呼ばれることがある[3, 4]。また、薬事規制下において、医薬品承認申請等の資料作成に申請者が使用するコンピューター化システムに対して期待される結果を与えることを検証することはコンピューター化システムバリデーション(CSV)と呼ばれている[5]。

(2) アウトカム定義のバリデーション研究の重要性

データベース研究の際に定義しなければならない要素の中でも、特にアウトカム定義は重要であり研究の質を左右する。アウトカムが何らかの疾病である場合、アウトカム定義には International Classification of Diseases Version 10 (ICD-10)コードなどの傷病名を用いることが多いが、その傷病名が真の疾患発症を意味しているか、また真に疾患発症している患者に対して適切な傷病名が入力されているかどうかは確認してみなければわからない。妥当性が確認されていないアウトカム定義を用いた研究結果は信頼できない可能性がある。入力された傷病名の妥当性は疾患毎並びに医療機関毎に異なる可能性があり、また国の医療制度やデータベース本来の目的（例：診療報酬請求を目的としているか否か）によっても影響を受ける[6]。このため、一概に医療情報データベースの傷病名を妥当と考えることは適切ではないし、逆に日常診療データの傷病名は全て研究で使用できないと決めつけることも適切ではない。より高いエビデンスレベルを求めるならば、利用するデータベースの中でアウトカム定義の妥当性を疾患毎に適切な方法で得られたサンプルを用いて検討し、妥当性が高いことが確認できたものに絞って薬剤疫学研究を行うことが望ましい。

なお、アウトカム定義だけでなく交絡因子(confounder)や曝露(exposure)の定義の妥当性も重要である。交絡因子については、最終的な研究結果（例：曝露とアウトカムの関連を示す相対リスク）に決定的な影響を与える交絡因子が存在すると考えられるのなら、その因子についてはバリデーション研究の対象となりうる。しかし、交絡因子に固有の研究方法があるわけではなく、通常、アウトカム定義のバリデーション研究と同様に行う。また、曝露については、データベースを利用する薬剤疫学研究において、特定の薬剤の処方または調剤の電子記録によって定義される場合がほとんどであり、電子記録が標準化され同じようにデータに格納されている限りは、高い妥当性を持つと考えられる。処方または調剤の記録が正しくとも、患者が実際に薬を服用しなければバイアスを起こしうるが、コンプライアンス（服薬遵守）はデータベース研究以外の観察研究や介入研

究でも起こりえる問題であり、本タスクフォースが扱うバリデーション研究の対象外である。

(3) 代表的な教科書におけるバリデーション研究に関する記載の要約

バリデーション研究に特化した具体的な方法論・マニュアルではないものの、薬剤疫学の代表的な教科書である Pharmacoepidemiology, 5th Edition（Brian L. Strom ら）の中の、 Part IV Chapter 41 Validity of Pharmacoepidemiologic Drug and Diagnosis Data には、妥当性を論じるにあたっての留意事項が多く記載されている。以下は、その重要な点についてまとめたものである。

① バリデーション研究について

(i) バリデーションはデータベースから得られる情報とゴールドスタンダードを比べることにより行うものである。つまり、ゴールドスタンダードではない情報源と比較した結果を「バリデーション（研究）」と呼ばないよう明確に区別し、研究者は用語の使用に関し細心の注意を払うべきである。

(ii) ゴールドスタンダードは、（バリデーション研究の先にある最終目的となる研究の）リサーチクエスチョン、アウトカム及びゴールドスタンダードの設定に利用できるデータ、他のデータソースの利用可能性等で選択される。

(iii) バリデーション研究のベストプラクティスには、診療記録番号のような 2 つのデータセットを正確にリンケージ（照合）する情報が必要不可欠である。

② 妥当性を測定する指標について

妥当性(validity)を測定する指標には、感度・特異度・陽性的中度(positive predictive value, PPV)・陰性的中度（negative predictive value, NPV）がある（表 1）。感度と特異度は一般的にトレードオフの関係（感度の高い指標では特異度が低く、特異度の高い指標では感度は低くなる関係）があり、また的中度(PPV と NPV)は感度・特異度に依存する。感度と特異度を重要視しつつ、バリデーション研究においては、感度・特異度・陽性的中度・陰性的中度の 4 指標を算出できるようデザインすることが理想的である。また、これらの指標のうち、いずれの指標が最も重要な指標であるかは、研究のセッティングにより変わる。さらに、同じ感度・特異度であっても、それらが研究結果に与えるバイアスの程度は、研究集団における曝露の割合等の影響を受けるため、指標の絶対値にはあまり意味がない（具体例は付録1を参照）。

表 1.妥当性を測定する 4 つの指標

		ゴールドスタンダードの定義を満たすか（真のケースか）			
		Yes	No	合計	
アウトカムの定義を満たすか（陽性のケースか）	Yes	a（真の陽性）	b（偽陽性）	a+b	**陽性的中度(PPV) =a/a+b**
	No	c（偽陰性）	d（真の陰性）	c+d	**陰性的中度(NPV) =d/c+d**
	合計	a+c	b + d	a+b+c+d	
		感度=a/a+c	**特異度=d/b+d**		

③ 誤分類に関する考慮事項

曝露（医薬品）に関する administrative data（管理データ）や medical record data（医療記録データ）の利用は、未請求処方、アドヒアランス、OTC 医薬品の併用などの情報を得ることが出来ない等、考慮すべき点もあるものの、（思い出しバイアスが懸念事項となる）質問票等での情報収集と比較しアドバンテージがあり、誤分類については一般に大きな問題とは考えられていない。これに対し、診断に関する administrative data の利用においてはよく懸念となる。特に、外来患者の診断データは、確からしさが低いため注意が必要である。医学的に定義が不十分な疾患や、保険者からの医療費償還時のルールや保険の範囲等、様々な要因が ICD-10 コード化時のバイアス（系統誤差）を起こしている可能性もある。さらに、既に評価されたデータであっても、医療の進歩等や医療環境の変化に伴い、医療行為やコード分類が変化することもあり、再度の妥当性評価が必要となる事がある。

なお本タスクフォースでは、データベース研究に関連する国内外の代表的なガイドライン・ガイダンスにおいても、アウトカムのバリデーション研究に関する言及の有無及び内容の確認を行ったが、研究方法や結果の解釈の仕方についての包括的な記述は確認できなかった（詳細は付録 2 を参照）。

3. バリデーション研究実施の手順・チェック項目

まず本タスクフォースでは、過去のバリデーション研究のレビューを行った（詳細は付録 3 を参照）。その結果を踏まえて、本タスクフォースから日本で行うバリデーション研究の手順及びチェック項目について提案する。

まず研究の前提事項として、バリデーション研究も数ある研究の 1 つであり、明確なリサーチクエスチョンを立てて研究を始める必要がある[7]。その際、「〇〇のデータベースにおいて、□□の疾患を特定する△△の傷病名（またはアルゴリズム）が妥当であることを示す」（仮説検証的研究）あるいは「〇〇のデータベースにおいて、□□の疾患を妥当に特定できるアルゴリズムを探す」（探索的研究）旨の記載が基本となるが、さらに一歩踏み込んで、その研究結果が今後どのような研究で使われるのか、最終目的を明確にすることが望ましい。何故ならば、バリデーション研究後の使用目的によって、求めるべき指標やサンプリング方法が異なってくるからである。例えば、後述する通り、薬剤Ａと薬剤Ｂのあるアウトカムに対する相対リスクを正しく求めることが最終目的であれば、アウトカム定義を満たす患者群からのみランダムサンプリングを行い高い陽性的中度を示せることが出きれば十分かもしれない。一方で、アウトカムの発生率を検討する研究で使われるのなら、感度も求める必要がある。また、他の臨床研究同様、ガイドライン（Guidelines for Good Pharmacoepidemiology Practices など）に沿ったプロトコールの作成と倫理審査のステップを経ることが望ましい。また近年は観察研究であっても、研究プロトコールを UMIN 臨床試験登録システムなどに事前登録することが求められるようになってきている[8]。さらに可能であれば、少数を対象としたパイロット研究により、研究の実施可能性や必要な調査項目の記載漏れ等について確認することが望ましい。

ここから以下の(1)～(8)の項目は、過去のバリデーション研究のレビューの際に用いた項目（付録 3）と対応している。

(1)利用する医療情報データベースの種類の理解

今後の研究、すなわちバリデーション研究で評価する（データベースから得られる情報による）アウトカム定義を用いて実施する研究に利用する医療情報データベースが、入院・外来の診療報酬請求データ・退院時記録・電子カルテ・疾患レジストリー、またはそれらの組み合わせ、のどれに当たるか良く理解しておく。日本のレセプトデータ（DPC レセプト含む）は診療報酬請求データに当たり、Diagnosis Procedure Combination (DPC)データは主に厚生労働省による DPC/PDPS（Diagnosis Procedure Combination / Per-Diem Payment System：診断群分類包括評価制度）の評価のための退院時記録からなる administrative data である。また厚生労働省とPMDA が構築中の医療情報データベース（Medical Information Database Network、以下 MID-NET®）は当該病院における診療報酬請求データ及び電子カルテを組み合わせたデータベースである。

また、近年多くの日本の学会が疾患レジストリーを開設しているが、海外のそれと異なり、疾患レジストリーを他のデータベースとリンケージ（照合）させて研究に使用することは現状では困難である。このため、疾患レジストリー自体に曝露やアウトカムの情報を蓄積して研究を行う、自己完結型のコホートの構築が志向されている。疾患レジストリーについてもバリデーション研究の実施は重要であり、倫理上の問題の整理を含めて将来的な課題と考えられる。

バリデーション研究を行う前にまず、今後の研究に利用する医療情報データベースがどのような集団を対象に実施され、また"population-based"と言えるのか、について理解しておくことが重要である。この報告書では、「居住地域や加入する保険などが共通する健康人を含む集団」を対象とする研究を"population-based"の研究と呼ぶ。日本において利用可能なデータベースのうち、レセプト情報・特定健診等情報データベース(いわゆる National Database、NDB)のレセプトデータは日本の全国民から得られたデータであり、"population-based"と言える。また、複数の健康保険組合が保有するレセプトデータを蓄積したデータベースは、特定の種類の保険者のデータであり、性・年齢分布は日本の全国民を代表してはいないが、特定の集団を代表する"population-based"のデータベースと考えることができる。一方、特定の医療機関のデータを蓄積したデータベースについては、日本では患者が受診する医療機関を居住地域に関係なく自由に選択できるため、"population-based"とみなすことはできない。また、過去のバリデーション研究の中には、一定期間にある病院に入院した、またはクリニックで診療を受けた全員を対象として実施された研究[9-14]や、あるデータベースの中で特定の疾患をもつ患者集団を対象とする研究[15-19]もみられた。

(2)バリデーション研究のセッティングの理解

理想的なバリデーション研究は、今後の研究に利用する医療情報データベースそのもの(からランダムサンプリングを行った集団)を対象とした研究である。この場合、バリデーション研究で求めた妥当性の指標(感度・特異度・陽性的中度・陰性的中度)を、利用する医療情報データベースに直接外挿可能である。しかし現実的には、今後の研究に利用する医療情報データベースのうち一部の偏ったサンプルを対象にバリデーション研究を行わなければならないことが多い。

例えば、上述の通り、NDB のレセプトデータは日本の全国民から得られたデータであり、"population-based"である。しかし日本の現状では、レセプト側から患者または医療機関をランダムサンプリングして元データに遡ってバリデーション研究を実施することは困難である。このため、まずバリデーション研究が実行可能な医療機関を任意に決めるところから開始しなければならないが、その(単数または複数の)医療機関が NDB などレセプトデータを得た集団を代表し、"population-based"と呼べるバリデーション研究になっていることはまずありえない。

このような条件下で実施されるバリデーション研究、特にバリデーション研究で評価したアウトカム定義をレセプトデータで用いることを意図している場合には、バリデーション研究のセッティング・対象集団の特徴の記述がより重要となってくる。例えば、バリデーション研究が実施された地域と期間、医療機関の数と特徴(例:大学病院、救急指定病院、がん拠点病院、受診患者数、病床数、等)および対象集団の特徴(年齢・性別構成、重症度の分布、等)を記述して、これを利用するレセプトデータにおける平均的な特徴と比べることなどにより、どれだけバリデーション研究の結果がレセプトデータ全体に外挿可能かを議論することが望ましい。

また、レセプトデータベース研究での利用を意図して行う、日本における病院単位のバリデーション研究では、バリデーション研究における対象集団の定義への配慮が必要である。例えば、特定の病院の整形外科で診療を受けている患者が糖尿病にも罹患しているものの、糖尿病については他の医療機関で診療を受けているのなら、当該病院に保管されているレセプトデータには糖尿病の傷病名が含まれていない可能性がある。しかし、この患者を糖尿病の傷病名を含むアウト

カム定義の偽陰性例に分類することには慎重でなければならない。他の医療機関で糖尿病の診療を受けていれば、その傷病コードはレセプトデータベースには存在するからである。一般に他の医療機関で診療を受けていることが明らかな患者は病院単位のバリデーション研究では対象集団から除外するか、除外する／しない場合の両方の結果を示すことが適切である。たとえば、先天異常に関するバリデーション研究では、出産後、他の医療機関に移った症例を除くと陽性的中度が高くなることが報告されている[20]。

一方、特定の医療機関のデータを蓄積したデータベースを用いた研究を意図している場合には、データベースそのもの（からランダムサンプリングを行った集団）を対象にバリデーション研究が実施可能である場合が多い。しかし、データベースを構成している医療機関の一部の医療機関のみでバリデーション研究を行わなければならない場合は、バリデーション研究が実施された医療機関・対象集団の特徴とデータベース全体の平均的な特徴の対比により、結果の外挿性の議論が必要である。

なお、”population-based”ではない医療情報データベースを用いてバリデーション研究を行った場合、または、ある医療情報データベースのうち特定の疾患をもつ患者集団を対象としてバリデーション研究を行った場合には、求めた感度や特異度の解釈の仕方には注意が必要である。例えば、一定期間にある病院に入院した、またはクリニックで診療を受けた全員を対象とする研究[9-14]では、感度・特異度は「その病院に受診したという条件付きの感度・特異度」を意味する。ある疾患をもつ患者集団を対象とする研究[15-19]では、感度・特異度は「その疾患を持つという条件付きの感度・特異度」を意味し、その疾患を持つ患者集団における（その疾患の続発症などの）アウトカムの有病割合(prevalence)または発生割合／率(incidence proportion/rate)を推定するのに使うことができる。また、後述するように、感度、陽性的中度などの妥当性の指標の値が異なるサブグループを見出すこと、偽陽性、偽陰性の患者の特徴を明らかにすることなどが、バリデーション研究で評価したアウトカム定義が、それを用いて実施するデータベース研究においてどの程度有用かの情報を与えることにつながると考えられる[13, 20-29]。

一方で、研究者がアウトカム定義を満たす集団と満たさない集団から異なる割合でサンプリングしたものを単純に足し合わせることによって作り出した集団において感度・特異度を評価した研究も見られた[26, 30-34]が、このような恣意的な集団はその後の研究で再現できないため、求めた感度・特異度に意味を見出すことは難しい。また、あるアウトカム定義を満たす（したがって本来、陽性的中度しか求めることができない）集団において感度、特異度などを求めている研究[35, 36]も見られたが、本タスクフォースとしては、アウトカム定義を満たす集団と満たさない集団から異なる割合でサンプリングした場合には、後述する適切な補正を行うなどにより、サンプルを得た集団全体における感度や陽性的中度などを示すことを推奨する。

(3)アウトカムの定義

医療情報データベースにおけるアウトカムは、1 つの要素（傷病名のみ、薬剤名のみ、処置名のみ、検査結果のみ、など）、または 2 つ以上の要素を組み合わせたアルゴリズムによって定義されるが、傷病名、または傷病名とそれ以外の情報の組み合わせが用いられることが多い。本タスクフォースは必要に応じて複数のアルゴリズムについて陽性的中度などの指標を求め、比較を行う

ことを推奨する。傷病名はコード化されたものを使うことが多く、日本のレセプトでは独自の（ただし、ICD-10 と対応可能な）傷病名マスターが使用され、DPC データベースでは ICD-10 コードも使用される。傷病名によるアウトカム定義では、傷病名コードのリスト作り、および傷病名コードが入力された期間と期間内の入力頻度の選択、が必要である。まず傷病名コードのリスト作りについて、同じ疾患の定義を意図していても、異なる研究者が選んだ傷病名コードのリストは異なる可能性がある（例：上部消化管出血を定義するための傷病名コードのリストに吐血や黒色便を含めるかどうか）。よって、その疾患の専門家も含めた 2 人以上の研究者が議論してリストを作ることが好ましい[37]。また、ICD-10 コードのように海外でも使われている傷病名コードを用いる場合は、過去の研究で使われた傷病名コードのリストを各文献から参考にできることがある。作成または参照した傷病名コードのリストは論文内または Appendix に明記すべきである。

次に傷病名コードが入力された期間と期間内の入力頻度の選択について考察する。日常診療、特に外来では一度だけの入力病名は検査目的で入力されている可能性があり、あるいは医師が疑い病名を意図して入力した際に「疑い」を付け忘れた可能性も否定できない。そこで、カナダのバリデーション研究で見られるように、外来で一定期間内に 2 回以上の傷病名が入力された時に初めてアウトカムと定義するなど、一定期間内の受診回数をアウトカム定義に含める場合がある。ただし、カナダでは外来診療であっても primary diagnosis の指定が医療費償還の条件になるのに対し、日本の外来診療は出来高払いであり、「主傷病」の指定は必須ではないなど違いがあることを認識しておくことは重要である。また、日本の診療報酬請求制度を考慮すると、「異なる月に」2 回以上の傷病名入力という条件を付けたほうが良いかもしれない。一方、カナダのバリデーション研究では、入院した場合は 1 回の主診断（primary diagnosis）でアウトカム定義が満たされるとしていることが多い[38-42]。日本のレセプトデータでも入院レセプトと外来（入院外）レセプトは判別できるため同様の定義は可能である。ただし、どのような定義が適切かは疾患に大きく依存する（例：慢性疾患や癌は 2 回以上の傷病名入力が好ましいかもしれないが、急性肝炎や転倒・骨折などの一過性のアウトカムについては 1 回の傷病名でも十分かもしれないし、心筋梗塞など入院することが確実なアウトカムでは入院病名のみを対象としたほうがいいかもしれない）。そのため、その疾患の専門家も含めた 2 人以上の研究者が議論し、複数のパターンが提案された場合はバリデーション研究で比較することが望ましい。

2 つ以上の要素を用いてアルゴリズムを作成する場合[11, 12, 37, 43, 44]も、複数のパターンを作ってバリデーション研究で比較し、最終的にどのアルゴリズムを用いることが適切か論文の中で議論するべきである。なお、傷病名以外の情報を使う場合、医療機関によって処置や検査の方針が異なる可能性がある（例：上部消化管内視鏡処置を行う患者に対して全例生検を行う病院もあれば、癌が疑われた場合にのみ生検を行う病院もある）ため、他施設でも処置・検査の方針が同様か注意しながらアルゴリズムを作成することが、結果の一般化可能性を高めることにつながる。

(4)リンケージの有無・方法の確認

海外のバリデーション研究では、評価されるデータベース（全体、またはサンプリングされた一部の患者）をその他のデータベースまたは電子カルテとリンケージ（照合）させて、ゴールドスタンダードの情報を得ることが多い。またカナダの（入院の）Discharge Abstract Database と（外来の）

claims data などを組み合わせた administrative data のように評価されるデータ自体が複数のデータベースのリンケージ（照合）によって成り立っている場合もある[39, 40]。その際、リンケージ（照合）させるデータベース同士に共通の個人レベルの識別子(social security number など)がある場合は 1:1 対応させることができ、通常 deterministic（一方のデータベースに含まれる特定の 1 個人が他方のデータベースに含まれる 1 個人と同一であることが確実である時に行う）リンケージ（照合）が可能となる[23, 25]。それができない場合は患者の氏名・生年月日・居住地域などの情報を用いて deterministic または probabilistic（一致する確率が一定程度以上であれば行う）リンケージ（照合）を行う[45]。Deterministic リンケージ（照合）では入力データにエラーが多いとリンケージ（照合）ができないため除外される人の割合が高くなり、Probabilistic リンケージの場合は異なる 2 人の個人が同一とみなされる可能性が排除できないため、リンケージ（照合）の詳細について記述することが望ましい[46]。

しかし日本の法律・指針および制度では現在、医療情報データベース同士を照合することやレセプトや DPC データベースから各医療機関の電子カルテに遡ることは困難である。特に改正個人情報保護法における「匿名加工情報」および次世代医療基盤法における「匿名加工医療情報」に該当するデータについては、カルテレビューにより行うバリデーション研究で必要となる特定の個人を識別リンケージ（照合）することが法的に禁止されている。匿名加工情報や匿名加工医療情報でないデータについても（例：NDB や疾患レジストリー）、個人レベルでのリンケージ（照合）は大きく制約されていることが多い。よって当面、日本で実施されるバリデーション研究の多くでは、各医療機関の中で保存しているレセプト・DPC データと電子カルテや院内の疾患レジストリーへの登録情報を比べる作業に頼らざるを得ない。なお、各医療機関の中では同じ患者の情報が元々まとめて保存されてあるという前提があるため、この作業はリンケージ（照合）には当たらない。

(5)ゴールドスタンダードの定義

ゴールドスタンダードの設定は、まずカルテレビューをするか否かに大別される。カルテレビューをしない状況は比較的稀で、疾患レジストリーが存在する場合か、臨床検査（例：低ナトリウム血症[47]）や病理（例：癌の確定診断[44]）など検査結果のみでゴールドスタンダードが設定できる場合が含まれる。疾患レジストリーをゴールドスタンダードとする場合[25, 29, 45, 48, 49]、疾患レジストリーがそもそも高い妥当性を持ち合わせているかを確認しておく必要がある。まだ開設されたばかりで網羅的に症例が登録されていない、あるいは疾患の正確性が十分確認されていない症例が多く混ざっているような疾患レジストリーはゴールドスタンダードとして使うことはできない。疾患レジストリーの妥当性に関する研究（疾患レジストリーのバリデーション研究）が過去に実施されている場合はその結果を引用するべきであるし、実施されていない場合は疾患レジストリーがゴールドスタンダードとして妥当であるとの判断が合理的である旨記述することが望ましい。検査結果をゴールドスタンダードにする場合は、検体が正確に測定・標準化（例：尿蛋白の±、+、2+の分類）・記録されているか、また病理診断や画像診断の場合は診断結果が標準化・電子化されているかを確認しておく必要がある。もし標準化・電子化されていない場合は、標準化・電子化の作業をまず実施するか、実質的に医療機関に戻って行うカルテレビューと同じ方法でバリデーション研究を実施しなければならない。

カルテレビューをする場合、その質を可能な限り担保する努力が必要である。過去の文献の中には「カルテレビューにより目的の疾患の有無を判断しゴールドスタンダードとした」と一文のみ記述しているものが見られたが[50]、これは好ましくない。まず、対象疾患の判断基準について、世界的に確立した診断基準がある場合には、その診断基準にそった情報収集が可能になるような調査票を作成・利用し、評価することが望ましく、報告時には用いた基準を明記する。明確な診断基準がない場合には、その研究でどのような基準を用いたかを明示するべきである。いずれの場合も使用した調査票については、論文内またはAppendixなどの形で示すことが望ましい[51, 52]。そして、その診断基準または調査票に沿って、1 人の患者につき 2 人（以上の少人数）の評価者が別々に判断することが望ましく、さらに 1 人の評価者が同一の患者について異なる時点で判断することも考慮する。研究計画書と報告書の中では、評価者間（内）の一致度を表すκ（カッパ）係数を信頼性(reliability)の指標として算出する[41, 53, 54]。レビューするカルテの数が多い場合には、その一部についてのみ評価者間（内）の一致度を算出することも認められる[40]。なお、評価者の専門性（例：対象疾患の専門医、訓練を受けた医療関係者、等）も記載しておくことが望ましい。評価者の中に対象疾患の専門家がいる場合は、（複数の）スクリーニング担当と（1 人の）最終判断担当に分ける方法[41, 51]も許容されるかもしれない。その場合も、スクリーニング担当者間（内）の一致度は求めることを推奨する。κ（カッパ）係数の他に、一致しなかった評価結果の取り扱い（例：2 人で話し合って最終的に決めるか、3 人目の判断を仰ぐかなど）についても記述する。

(6)サンプリング方法の選択とサンプルサイズの設定

サンプリング方法は、以下(A)～(E)の方法に大別される。それぞれの方法に長所・短所があり、目的と状況に応じて適切に選択することが望ましい。なお、この分類は本タスクフォースで行った過去の約 100 文献のレビューにもとづいたものであり、(A)～(E)と異なる方法でサンプリングを行うことは必ずしも誤りとは言えない。

(A) 全患者を対象にする場合

まず、カルテレビューをしない場合、すなわち疾患レジストリーや検査結果を用いてゴールドスタンダードを定義する場合、データベースまたは対象の医療機関に登録されている全患者を検討することが多い。つまり、データベースの中でアウトカムの定義を満たす患者を全て特定し、一方でゴールドスタンダードの定義を満たす患者も全て特定し、全患者を対象に 2×2 テーブルを作成する。調査症例数を制限なく検討できる理由は、ひとえにアウトカムとゴールドスタンダードの定義を満たす患者を、それぞれコンピューターを用いて簡単に特定できるからである。もし疾患レジストリーや検査結果が標準化・電子化されておらず手作業で判断する必要がある場合は、まず標準化・電子化による作業の効率化が必要である。

(B)全患者からランダムサンプリングを行う場合

カルテレビューをする場合は、大きな労力・時間がかかるため全例調査は現実的でなく、通常調査症例数を制限せざるを得ない。その際、利用するデータベースまたは対象の医療機関の全患者の中から調査症例をランダムサンプリングすることは可能であり、実際、数千から数万例のランダムサンプルのカルテレビューによるバリデーション研究がこれまでにいくつか実施されている[40, 41, 55]。しかし、頻度の低い疾患では、例え数千例のランダムサンプリングをしたとしてもアウトカ

ムとゴールドスタンダードの定義を満たす患者がそれぞれ数例しか含まれず、妥当性の指標を十分に計算できない[38]。望ましい精度で感度・特異度・陽性的中度・陰性的中度の全てを求めるためには、全例を代表する①アウトカム定義をみたす集団（陽性のケース）、②アウトカム定義をみたさない集団（陰性のケース）、③ゴールドスタンダードの定義を満たす集団（真のケース）、④ゴールドスタンダードの定義を満たさない集団（偽のケース）を一定数（例：最低 100 人、根拠は後述の段落を参照）特定しなければならない。アウトカムが稀な疾患の発生である場合に、このうち最も困難なのが③にあたる真のケースを一定数以上特定することである。例えば、研究対象集団に含まれる真のケースの期待される割合が 1%であるのなら、100 人の真のケースを特定するためには 10,000 人(100 人÷0.01)のランダムサンプルが必要である。一方、アウトカムの頻度が比較的高いセッティングにおけるバリデーション研究（例：集中治療室でしばしば起こるアウトカム[56]）では、比較的小さなサンプル数でも十分な数の真のケースが含まれる可能性がある。

(C)アウトカム定義を満たす患者群からのみランダムサンプリングを行う場合

データベースの中でアウトカム定義を満たす患者群のみからランダムサンプリングして、その症例に限ってカルテレビューをする方法である[37, 53, 57, 58]。ただしこの方法では、アウトカム定義を満たさない患者（陰性のケース）が含まれないため、陽性的中度しか求められないことに留意する。その際サンプルサイズの設定としては、「求めたい指標の 95%信頼区間が最大±10%を目標とするならば 100 例、±5%を目標とするならば 400 例必要」という計算式があり(詳細は付録 4 を参照)、これを目安として米国の Mini Sentinel 調査では同サンプリング方法で陽性的中度を求める際には 100 例あれば十分と考えている[52]（前述した最低 100 人の根拠）。ただし、この見積りは、得られる陽性的中度の大きさが不明であり、得られる陽性的中度の大きさに関わらず±10%または±5%以内になるように陽性的中度 50%（最も信頼区間の幅が広くなる状況）を仮定した時のものである。より高い陽性的中度が得られることが相当確実に期待できる場合は、より少ない症例数を設定することも許されるかも知れない。

なお、アウトカムの定義が複数パターン考慮された場合は、それぞれのパターンを満たす患者から 100 例ずつランダムサンプリングをすることが確実ではあるが、各パターンに重複があるとすれば、その分だけ調査症例数を多少減らせるかもしれない。

(D)アウトカム定義を満たす患者群と満たさない群からランダムサンプリングを行う場合

アウトカム定義を満たす患者群からだけではなく、満たさない患者群からも異なる抽出割合でランダムサンプルを実施し、陽性的中度のみならず感度（および特異度・陰性的中度)の推定につなげる方法である[59]。この方法は、アウトカム定義を満たす患者群を予め定めた一定数サンプリングすることにより、得たい精度で陽性的中度を求めた上で、さらに精度は低くなっても感度についてもある程度の情報を得ようとする時に適している。ただし、アウトカムが稀な疾患の場合には、アウトカム定義を満たさない患者群に必要とされるサンプルサイズは相当大きくなる。どの程度のサンプル数とするかは、研究の目的により異なりうる。この方法を用いる場合には、アウトカム定義を満たす患者群からは最低 100 例、アウトカム定義を満たさない患者群のサンプルサイズについては付録 4 を参照。なお母集団が大きい場合には、アウトカム定義を満たす患者群と満たさない群からそれぞれランダムサンプリングを行う前に、まず全集団からのランダムサンプリングを行う（または適切と考えられる範囲で、特定の期間など全集団の一部の情報を使う）ことで取り扱う母

集団のサイズを減らしデータのハンドリングを良くしておく作業も考慮される。

(E)“all possible cases”を想定したサンプリングを行う場合

研究対象集団（全体あるいはそのランダムサンプリング）の中で真のケースを多く含むことが期待されるサブグループを 2 種類以上（例：「レセプトに該当病名がある」サブグループと「関連する検査結果が正常値の〇倍以上」のサブグループ）特定し、これらのサブグループの集合体を対象にカルテレビューで真のケースを見出し、これを「全ての真のケース」と仮定して、陽性的中度のみならず感度（および特異度・陰性的中度）も推定する方法である。この方法において、真のケースを全て含んでいると考えられるサブグループの集合体を Widdifield は″all possible cases″と呼んでいる[38]。同様の方法で真のケースを全て(または、その大半を)特定しようとする試みはこれまでのバリデーション研究でもみられる[13, 25, 60-64]。この方法では、複数のサブグループの一つに「アウトカム定義を満たす」グループを通常含めるが、まず、集団全体に含まれる真のケース（と考えられる）全例を特定し、その上でアウトカムの定義を満たすか否かで全患者を対象に 2×2 テーブルを作成するという点では（A)または（B)と共通する。”all possible cases”の方法では、いずれかのサブグループに属する全例についてカルテレビューを実施する。母集団が大きい場合、適切と判断されれば、まず母集団を代表するより小さな集団を（調査期間を短くするか、ランダムサンプリングによって）設定し、その集団に含まれる該当者全員に関してカルテレビューを実施し、最終的に 100 例以上の真のケースを特定することを目標とする。

なお、”all possible cases”の方法は、集団全体またはそのランダムサンプルの調査ではないので、「検討の対象となったサブグループの集合体に含まれない真のケース」の存在を否定することは通常できない。特に、設定可能なサブグループに含まれない真のケースが相当数存在することが確実であるような場合には、″all possible cases″の方法は、感度が過大評価され、アウトカム定義が「実際よりもよい」との誤った印象を与えることにつながるため推奨できない。設定されたサブグループから漏れる真のケースがそれほど多いとは考えられない場合にも、″all possible cases″の方法を用いた場合には、「検討の対象となったサブグループの集合体に含まれない真のケース」の程度が、推定された指標にどのような影響を与えるかについての感度解析をしておくことが望ましい[62]。

(7)妥当性を測定する指標の計算

妥当性(validity)を測定する指標には、感度・特異度・陽性的中度・陰性的中度が含まれる（前述の表 1 を参照）。臨床疫学の教科書でとりあげられる医学検査の妥当性の検討では、疾患のスクリーニングから確定診断へのプロセスなどを念頭に、感度と特異度が重要視されるが、医療情報データベースにおいて相対リスクの推定を主な目的として実施される薬剤疫学研究においてアウトカム定義の妥当性を検討する際には、まず高い陽性的中度を証明することを最優先にしている研究が多い。その理由は、陽性的中度が高いアウトカム定義の特異度は通常高く（理由：表 1 において陽性的中度[a/a+b]が高い＝偽陽性[b]が少ない＝特異度[d/b+d]が高い）、また「比較される 2 群（例：糖尿病薬 A 群 vs.糖尿病薬 B 群）間のアウトカムに対する相対リスクは、アウトカム定義の特異度（陽性的中度）が高く、感度が（その大小に関わらず）2 群間で同じ(非差異誤分類の)場合に正しく求まる」(“with perfect specificity, nondifferential sensitivity of disease

misclassification will not bias the risk ratio”(Modern epidemiology 3rd edition)[65]からである。アウトカム定義の感度が 2 群間で大きく異なることが考えにくいのであれば、陽性的中度が高いアウトカム定義を用いることが正しい結果を得るための必要条件となる。一方、陽性的中度が高いのみならず感度も高い指標が見いだされれば、研究対象集団における有病割合(prevalence)または発生割合/率(incidence proportion/rate)が正しく推定でき、2 群間の全体リスク差/率差を求めることができる。なお、一般的に感度と特異度・陽性的中度はトレードオフの関係にあり、高い陽性的中度を目指してアウトカム定義を厳しく設定する場合には、偽陰性が増え感度が下がる可能性があるため、感度の評価は可能な限り同時に行うことが好ましい。また、陽性的中度は対象集団における有病率の影響を受けるため、特異度の評価も行うことが理想である。

また、本タスクフォースでは、レビューを行った過去の約 100 文献にもとづき、妥当性の指標として感度・特異度・陽性的中度・陰性的中度を求めることを基本と捉えたが、診断疫学の分野では併せて診断オッズ比(＝陽性尤度比/陰性尤度比＝(感度×特異度)/((1−感度)×(1−特異度)))や ROC 曲線下面積も計算されることが多い。バリデーション研究においてもこれらの指標は有用である可能性があり[66]、実際に計算している研究も存在する[67]。よって、これらの指標も感度・特異度・陽性的中度・陰性的中度と同時に計算し、その有用性について検討することは、今後のバリデーション研究の課題の 1 つと考えられる。

以下に「(6)サンプリング方法の選択とサンプルサイズの設定」の項で示した各サンプリング方法(A)～(E)に対応した各指標の計算方法を示す。

(A) 全患者を対象にした場合、または(B)全患者からランダムサンプリングを行った場合

全患者またはランダムランプリングされた患者を対象に表 1 のような 2×2 テーブルを作成することにより指標の全てを求めることができる。

(C)アウトカム定義を満たす患者群からのみランダムサンプリングを行った場合

2×2 テーブルは作成できず、陽性的中度(＝アウトカム定義及びゴールドスタンダードの定義を満たす患者数/アウトカム定義を満たす患者数)のみ求めることができる。

(D)アウトカム定義を満たす患者群と満たさない群からランダムサンプリングを行った場合

サンプリングのウエイトを考慮した 2×2 テーブルを作成することにより感度・特異度を正しく計算できる。つまり、サンプリングした患者のカルテレビュー結果をそのまま 2×2 テーブルにするのではなく、アウトカム定義を満たす患者群からのサンプリング割合を逆数にして(例:20%のサンプリングをしていたとしたら 5 倍)掛けた患者数、アウトカム定義を満たさない患者群からのサンプリング割合を逆数にして(例:5%のサンプリングをしていたとしたら 20 倍)掛けた患者数から 2×2 テーブルを作成することにより、正しい感度・特異度が求まる(詳細は表 2 を参照)。

表 2. アウトカム定義を満たす患者群からのサンプリング割合と満たさない患者群からのサンプリング割合が異なる場合の感度・特異度の求め方

シナリオ:全 5,000 症例のうち、アウトカム定義を満たす患者 500 人から 20%(100 人)のサンプリングを行い、満たさない患者 4500 人から 5%(225 人)のサンプリングを行い、カルテレビューを行った。その結果、アウトカム定義を満たす患者 100 人のうち 90 人、満たさない患者 225 人のうち 45 人がそれぞれ真のケースであった。

1） サンプリングのウエイトを考慮していない誤った感度・特異度の求め方

		ゴールドスタンダードの定義を満たすか（真のケースか）			
		Yes	No	合計	
アウトカムの定義を満たすか（陽性のケースか）	Yes	90 (a)	10 (b)	100	陽性的中度(PPV) = 90/100 = 90%（95% CI:82.4-95.1%）
	No	45 (c)	180 (d)	225	陰性的中度(NPV) = 180/225 = 80%（95% CI:74.2-85.0%）
	合計	135	190	325	
		感度 = 90/135 = 66.7%	特異度 = 180/190 = 94.7%		

2） サンプリングのウエイトを考慮した正しい感度・特異度の求め方

		ゴールドスタンダードの定義を満たすか（真のケースか）			
		Yes	No	合計	
アウトカムの定義を満たすか（陽性のケースか）	Yes	450	50	500	陽性的中度(PPV) = 450/500 = 90%（95% CI : 82.4-95.1%）**
	No	900	3,600	4,500	陰性的中度(NPV) = 900/4,500 = 80%（95% CI : 74.2-85.0%）***
	合計	1,350	3,650	5,000	
		感度 = 450/1350 = 33.3%（95% CI:26.3-42.3%)*（95%CI:26.2-42.2%)¶	特異度 = 3,600/3,650 = 98.6%（95% CI:97.8-99.5%)*（95%CI:97.7-99.4%)¶		

CI = 信頼区間

*以下の近似式による推定

感度（sen）の 95%信頼区間：sen*exp[±1.96(1-sen)√(1/a+1/c)]

特異度（spe）の 95%信頼区間：spe*exp[1.96(1-spe)√(1/b+1/d)]

詳細については付録 4 参照

¶Bootstrap 法による推定（10,000bootstrap sample から求めた percentiles）

**実際の検討数 100 例(全 500 例の 20%)から計算

***実際の検討数 225 例(全 4,500 例の 5%)から計算

(E)“all possible cases”を想定したサンプリングを行った場合

最終的に全患者を想定した 2×2 テーブルを作成することを念頭に、求められるところから順に各マスを埋めていく作業が必要である(詳細は表 3 を参照)。

表 3. “all possible cases”を想定したサンプリングを行った場合の 2×2 テーブルの作り方

シナリオ:全 10,000 症例から 2000 例をランダムサンプリングした。真のケースを多く含むことが期待されるサブグループの集合体として 300 人特定された(その中には、アウトカム定義を満たす患者群をサブグループの1つとして含めており、200 人いた)。300 人をカルテレビューした結果、240 人が真のケースであり、このうち 180 人がアウトカム定義を満たす患者であった。

		ゴールドスタンダードの定義を満たすか (真のケースか)			
		Yes	No	合計	
アウトカムの定義を満たすか (陽性のケースか)	Yes	180***	20	200*	陽性的中度(PPV) = 180/200= 90% (95% CI:85.0-93.8%)
	No	60***	1,740	1800*	陰性的中度(NPV) = 1,740/1,800 = 96.7% (95% CI:95.7-97.4%)
	合計	240**	1,760**	2,000	
		感度 = 180/240 = 75.0% (95% CI:69.0-80.3%)	特異度 = 1,740/1,760 = 98.9% (95% CI:98.3-99.3%)		

*～***各マスを埋めていく順序:

*2,000 症例のうち、アウトカム定義を満たす患者が 200 人。よって満たさない患者は 1,800 人と計算される。

**300 人につきカルテレビューをしたところ真のケースは 240 人。よって真でない(偽の)ケースは 1,760 人と計算される。

***カルテレビューで真のケースであった 240 人のうち 180 人が陽性のケースであったことから、60 人が偽陰性のケースと計算。(その結果、残りのマスも自動的に埋まる。)

なお、本項目に関する注意点として、すでに何度か言及してきたことではあるが、現状の日本で可能な病院単位のバリデーション研究の患者集団における指標の値は、利用するデータベース(例:レセプトデータ)にそのまま当てはまらない可能性が高いことに留意すべきである。一般に、集団に含まれる特定の疾患の割合(有病割合)が異なれば、感度、特異度が同一であっても陽性的中度や陰性的中度が異なることが知られているが、感度、特異度自体も集団によって異なりうる[68]。たとえば疾患の重症度が偽陰性の割合に影響するのなら、感度(および陽性的中度と陰性的中度)は重症度の分布が異なる二つの集団では異なりうる。また、当該疾患の診断コードを誤っ

て与えられる患者（例：類似する疾患、疾患の初期で診断基準を満たさないが将来当該疾患の患者として診断される可能性が否定できない患者）の分布が集団によって異なるのなら、特異度（および陽性的中度と陰性的中度）も集団によって異なりうる。したがって、病院単位のバリデーション研究では、単に感度や特異度を求めるにとどまらず、感度などに影響を与えうる疾患の重症度などの因子を明らかにし、その分布と重症度ごとの指標の値を示すべきである。同様に、特異度に影響を与えうる類似疾患などの存在を明らかにし、その分布と特異度などへの影響を示すべきである。また、重症度については、データベース研究でも利用可能な、治療薬、合併症などで区分した重症度を示し、データベース研究における感度などが、バリデーション研究において求められた感度などとどのように異なるかを予測できるようにすべきである。同様に偽陽性となりやすい疾患などについても、データベース研究においても特定可能な特徴を示すことにより、データベース研究における指標の値の予測に利用可能とすべきである。

理解促進のために架空の例を以下に示す。ある 1 病院で、レセプトの関節リウマチ病名の陽性的中度を求めるバリデーション研究を行ったと仮定する。カルテレビューをゴールドスタンダードとし、アウトカム定義を満たす群（関節リウマチ病名が付いている患者）から 100 例のランダムサンプリングを行った。その結果、100 人中 70 人が真のケースと特定され、陽性的中度は 70%と計算された。しかし、当病院における関節リウマチ患者の重症度の分布は、レセプトデータベースにおける関節リウマチ患者の重症度の分布と異なる可能性が懸念された。そこでサブグループ解析として、(i)生物学的製剤使用患者、(ii)（生物学的製剤以外の）抗リウマチ薬使用患者、(iii)抗リウマチ薬非使用患者の 3 群に分けた。その結果、サブグループの構成は(i)群：(ii)群：(iii)群 = 20:60:20 であり、陽性的中度は(i)群で 90%、(ii)群で 70%、(iii)群で 50%であった。一方、レセプトデータベースにおけるサブグループの構成は(i)群：(ii)群：(iii)群 = 5：50：45 と判明した。つまり、陽性的中度の高かった(i)群よりも、陽性的中度の低かった(iii)群の比率がレセプト全体において高かった。ここで、各サブグループの陽性的中度がバリデーション研究を行った病院とレセプト全体の間で等しかったという（大きな）仮定を置くと、レセプトデータベースでの陽性的中度＝(5 × 90 + 50 × 70 + 45 × 50)/100 = 62%と、当病院で求めた陽性的中度 70%よりも低いことが予測できる。

このようなサブグループ解析は有用ではあるが、特にカルテレビューの結果をゴールドスタンダードとするバリデーション研究では、サンプル数は通常多くなく、サブグループの数も 2 つまたは 3 つ程度にとどめることが現実的であり、1 グループに相当数（最低 10 例以上程度）含まれていることが必要である。したがって、このようなサブグループ解析とともに、偽陽性、偽陰性の患者について詳細に検討し、結果を報告することを本タスクフォースとしては推奨する[13, 20-29]。偽陽性、偽陰性の患者がバリデーション研究を実施した施設に特異的なのか、あるいは、どのような医療機関でも起こりうると考えられるかなどを考察することにより、特定の医療機関において実施されたバリデーション研究の結果をその医療機関の患者が代表しているとは考えられないデータベース研究において利用する上で有用な示唆を与えうると考えられる。

(8)指標の閾値・利用法に関する考察

最後に「求めた陽性的中度や感度・特異度がいくつ以上であれば、そのアウトカムの定義が今後の薬剤疫学研究に使えるか?」という疑問について絶対的な基準の設定は困難である。主な理由として、各研究の目的により閾値は異なるからである。例えば、仮説の検証のための研究では

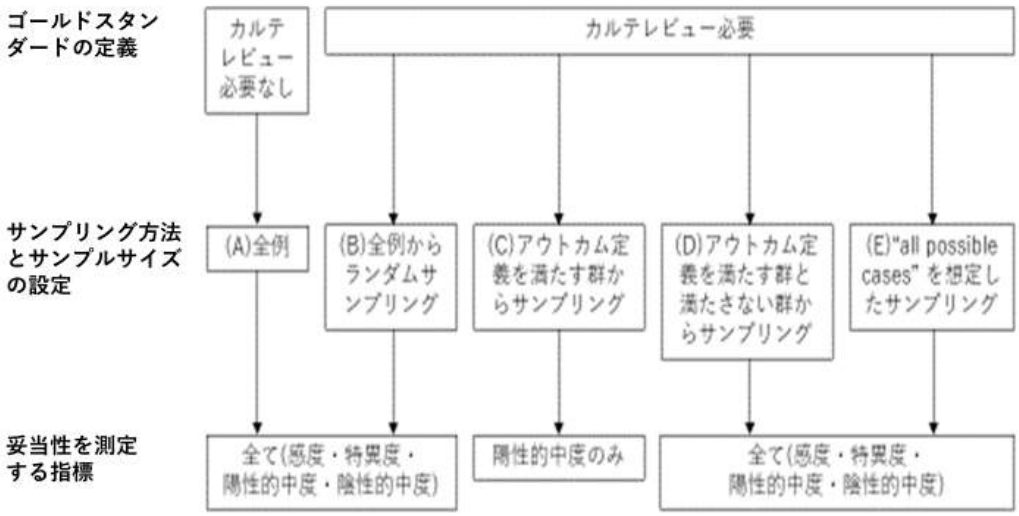
正しい相対リスクが求まるよう高い陽性的中度が要求される一方で、仮説の生成のための（探索的な）研究では多くのアウトカムを把握できるよう高い感度が望まれるかもしれない。また、2 群間の絶対リスク差やアウトカムの有病割合(prevalence)または発生割合／率(incidence proportion/rate)を求めたい場合も高い感度が要求される。

海外のバリデーション研究では、末期肝疾患の患者を見出すためのアルゴリズムの陽性的中度が 80%以上であることが望ましいとする論文[69]、COPD の患者を見出すためのアルゴリズムの陽性的中度が 85%以上であることが望ましいとする論文[54]が見られた。ただし、前者の「80%」は末期肝疾患の患者における薬の効果を評価するために患者群を特定するという目的に沿って設定された値であり、後者の「85%」は COPD 患者の治療状況や治療の質、その他 COPD 患者に関連する臨床的なクエスチョンを検討するために患者群を特定するという目的に沿って設定された値であることに留意する必要がある。一方、ある値以下であれば、指標として使うことはできないとの言及も見られた。例えば、アナフィラキシーのバリデーション研究において、陽性的中度 70%以下の結果に対して「今回のアウトカムの定義は、あくまで候補を絞りこむ手段に留まり、真のケースの確定にはその後カルテレビューを要する」と考察しているものが見られた[37]。

一つのバリデーション研究の中で複数のアウトカム定義のパターンまたはアルゴリズムを検討する際は、陽性的中度と感度の相対的なバランスが、最適なアルゴリズムを決定する一助となる。一般的には、アルゴリズムの条件を厳しくするほど陽性的中度（または特異度）が上がるが、感度（あるいはデータベースの中で特定できる患者数）が下がるという、トレードオフ関係が認められる。過去のバリデーション研究では陽性的中度を下げない範囲で感度ができるだけ高くなるアルゴリズムを推奨していることが多いが[39, 40]、研究者らが考える最適なアルゴリズムについての考察を論文内に記載しておくことが望ましい。

まとめとして、以上で述べたバリデーション研究の手順・チェック項目を一覧表の形にした（付録5）。バリデーション研究のプロトコール作成・研究の遂行・論文執筆の際に参照されたい。また、上記(1)～(8)の項目のうち、「(5)ゴールドスタンダードの定義」、「(6)サンプリング方法とサンプルサイズの設定」、「(7)妥当性を測定する指標の計算」には対応関係があることはあらためて意識しておきたい（図 1）。

図 1.ゴールドスタンダードの定義、サンプリング方法とサンプルサイズの設定、妥当性を測定する指標、の対応関係。3(バリデーション研究実施の手順・チェック項目)の(5)、(6)、(7)の相互関係を示す。


ゴールドスタンダードの定義
カルテレビュー必要なし
カルテレビュー必要
サンプリング方法とサンプルサイズの設定
(A)全例
(B)全例からランダムサンプリング
(C)アウトカム定義を満たす群からサンプリング
(D)アウトカム定義を満たす群と満たさない群からサンプリング
(E)"all possible cases" を想定したサンプリング
妥当性を測定する指標
全て(感度・特異度・陽性的中度・陰性的中度)
陽性的中度のみ
全て(感度・特異度・陽性的中度・陰性的中度)

4. 今後日本で、より望ましいバリデーション研究を実施するために

最後に、日本におけるバリデーション研究の現状を総括し、今後より望ましいバリデーション研究を実施するために必要な知識・事項・提言についてまとめた。

(1) 日本におけるバリデーション研究の現状

日本のバリデーション研究は 2010 年頃から実施されるようになり、2013 年以後に論文がいくつか発表されている。

2013 年の大場らの論文[70]は、(株)日本医療データセンターのレセプトデータと資格喪失理由を含む被保険者のデータを用い、レセプト上の転帰「死亡」とレセプト情報から求めた Charlson Comorbidity Index を組み合わせ、保険者の保有する資格喪失理由「死亡」をゴールドスタンダードとするバリデーション研究であり、レセプト情報から得られる情報による「死亡」の定義が感度 60%程度、陽性的中度 95%をもつことなどを明らかにした。

2015 年の佐藤らの論文[11]は、単一の病院におけるレセプトデータから新規に発生した乳がんを定義し、院内がん登録をゴールドスタンダードとしてその妥当性を検討するバリデーション研究であり、傷病名と治療に関する情報を組み合わせた定義が感度 90%、陽性的中度 87%などをもつことを明らかにした。

2015 年の山口らの論文[71]では、DPC 参加病院で EBM Provider((株)メディカル・データ・ビジョン)のうち、検査値の得られる 16 病院のデータを用い、データベースから利用可能な患者情報、病名一覧、薬剤、処置、検査による専門医の判断をゴールドスタンダードとして、ICD10 コードと薬、診療行為の組み合わせによる静脈血栓塞栓症と出血性イベントのアウトカム定義の陽性的中度がそれぞれ 75.0%、73.3%であることを明らかにした。

2016 年の田中ら[72]の論文は 3 つの病院でレセプト情報から得られる非定型大腿骨折の診断コードを電子カルテから得られる情報をゴールドスタンダードとして評価するバリデーション研究であり、感度は 82%前後、特異度は 100%であったことを明らかにした。

2017 年の山名ら[73]の論文は、4 つの DPC 病院でランダムサンプルした 315 人の患者の DPC データの 16 の疾患、10 の診療行為、SS-MIX データの 13 の臨床検査結果を、カルテレビューをゴールドスタンダードとして評価するバリデーション研究であり、DPC データ、SS-MIX のデータの妥当性は将来の研究に使う上で十分な妥当性を有していることを明らかにした。

(2) 行政が主導するバリデーション研究

MID-NET®は、医薬品の安全対策を目的として、厚生労働省と PMDA の医療情報データベース基盤整備事業において構築されたデータベースシステム及び関連ネットワークの総称である。厚生労働省が公募により選定した協力医療機関 10 拠点 23 病院における、2009 年以降のレセプトデータ、DPC データ、及び検査結果値を含む電子カルテデータ(以下、あわせて「HIS(Hospital Information System)データ」という。)が含まれており、データベース規模としては 2018 年までに約 400 万人になる見込みである。MID-NET®は分散型データベース構造であり、各拠点にデータベース(以下、「統合 DS」(統合データソース))が設置されている。各拠点の統合 DS には、SS-MIX2 の規約に基づき送信された HIS データが移行されており、この移行の過程で、傷病名、検体検査名、

薬剤名等に対してマッピング・標準化処理（医療機関で利用されている院内コード及び名称から標準コード及び名称への変換、臨床検査値と単位等の標準化など）が行われている。利用者がMID-NET®を解析する場合、作成したプログラムをオンサイトセンターから専用回線を通じて対象医療機関に送付すると、医療機関の担当者が依頼内容を承認することで対象となるデータが連結不可能匿名化処理を経て抽出され、解析処理が実行される仕組みとなっている。この抽出データ及び解析結果は医療機関の担当者による承認を経て複数施設統合処理センター（以下、「センター」）に送付される。利用者はセンターに送られてきたデータをシステム上で更に解析することが可能であるが、結果の持ち出しについては統計処方を施した集計結果についてのみ可能である。

2017年度より、日本医療研究開発機構の医薬品等規制調和・評価研究事業として、MID-NET®協力医療機関におけるアウトカムバリデーション研究が実施されている。各医療機関内において、候補とするアウトカム定義に基づき抽出された集団からランダムサンプリングした患者について、MID-NET®には含まれていない診療録情報を確認し、各拠点共通の判定基準に基づいて、その領域の専門医が真偽を判定する。対象となるアウトカムは、医薬品のリスク管理計画で記載されている安全性検討事項の上位のものから選定され、3年間で約20のアウトカムに対して実施される予定である。

(3) DPCレセプトのアウトカム定義への利用

日本のレセプトデータの利用を意図して行うバリデーション研究において、DPC レセプトの傷病名は医科レセプトの傷病名よりも特異性が高いことが期待され、今後のバリデーション研究でその妥当性が評価されるべきである。その際、DPC レセプトの構造についてよく理解しておくことが重要と考えられる。

本邦では2003年よりDPCに基づく、包括支払制度（DPC/PDPS）が導入された[74,75]。DPC参加病院では DPC 導入の影響評価に関わる調査のために3ヵ月に1度、厚生労働省に提出されるDPC データおよび医療費償還のための DPC レセプトと DPC に包括されなかった患者の医科レセプトが作成される。DPC 準備病院では DPC データおよび医科レセプトが作成され、それ以外の病院ではそれまで通り医科レセプトのみ作成される。DPCレセプトはDPCデータのDファイルに相当し、傷病コード（7桁）、ICD-10コード、傷病名区分コードが記録され、DPC データの様式1と同等の疾病情報が得られる。さらに、医科レセプトと同様、主傷病フラグ、診療開始日が記録される。但し、主傷病フラグは任意の入力項目であることから、バリデーション研究のアウトカム定義では、傷病名区分がより有用と考えられる。DPC レセプトにおける傷病名区分は、「傷病名（医療資源を最も投入した傷病名）」「副傷病名」「主傷病名」「入院の契機となった傷病名」「医療資源を2番目に投入した傷病名」「入院時併存傷病名」「入院後発症傷病名」の7項目で構成されるため、バリデーション研究のアウトカム定義で有用と考えられる。また、DPC レセプト、医科レセプトにはその病院における当該疾病に係る診療開始日が記録されるため、曝露との時間的前後関係を特定できる。但し、複数の病院の電子カルテやその病院で発行されたレセプト情報を集積したデータベースでは、病院が変わると患者の追跡ができないため、当該病院における診療開始日より前に疾患が発生している可能性があり、疾患の発生日の推定には注意が必要である。さらに、DPC/PDPSは急性期の疾患を対象とし、退院時に医療資源病名が確定し、適応される入院期間に上限があ

るという特徴がある。よって、長期の入院を伴う疾患など慢性期の傷病名のバリデーションにはDPCレセプトは不向きであり、医科レセプト情報との併用が必要である。一方、バリデーション研究においてアウトカムに対するアルゴリズムが組まれる場合には、手術、検査、処置の情報も重要となる。DPC レセプトにはこれら診療行為の情報が含まれるので、それぞれの情報を DPC コードから特定することができる。また、DPC コード自体をアウトカムや研究対象集団を決めるアルゴリズムとして利用することも可能かも知れない。DPC コードは、疾病と治療内容の組み合わせで患者を分類したもので、14 桁のコードで成り立っており、同一のコードは同一の疾患であるだけではなく、治療内容も同一であることを意味するからである[76]。

(4) 大規模コホート研究における国保レセプトを用いたバリデーション研究の可能性

バリデーション研究は、単施設もしくは複数施設で実施されることもあるが、外的妥当性の観点からは、研究の母集団を代表する集団でバリデーション研究を行うことが望ましい。国内の複数の地域住民を対象としたpopulation-basedコホート研究で、国民健康保険(国保)レセプト、後期高齢者医療制度(後期高齢医療)レセプトなどのレセプト情報を収集し、かつ、悉皆的な疾病登録が実施されている場合には、外的妥当性の高いバリデーション研究が実施可能である。

2010 年以降に開始された、第 3 世代の大規模コホート研究の次世代多目的コホート研究(JPHC-NEXT)や東北メディカル・メガバンク機構(ToMMo)では、研究参加者のレセプト情報も収集している[77]。このようなコホート研究では、レセプト情報と全国がん登録などの悉皆的な疾病登録とリンケージ(照合)させることにより、バリデーション研究が可能となる。全国がん登録の研究利用では、原則本人からの同意が必要であるため、全国がん登録を用いたアウトカム定義のバリデーション研究実施は困難だが、JPHC-NEXT や ToMMo ではこの同意の有無に関する情報も含まれている。代表性のある大規模なコホート研究でバリデーション研究が実施されれば、レセプト情報でがんなどのアウトカム定義の評価において有用な情報となることが期待される。悉皆的な疾病登録がない場合であっても、基幹病院への入院を要するような重篤疾患であれば、コホート調査対象地域の基幹病院で入退院サマリ、病歴(紙カルテや電子カルテ)を参照することにより、バリデーション研究が実施可能と考えられる。また、血液検査値や血圧値などで疾病の有無が判別可能な糖尿病、高血圧、脂質異常症、慢性腎機能障害などの慢性疾患では、健康診断結果や病院受診時の臨床検査値を参照すればバリデーション研究ができると考えられる

大規模コホート研究におけるバリデーション研究にも限界がある。第 1 に、現時点では、国保レセプトや後期高齢者医療レセプト以外のレセプト情報の収集は困難であり、その他(例:協会けんぽ)のレセプト情報を用いたバリデーション研究は実施できない。国保加入者や後期高齢者医療受給者とその他の保険加入者では、背景因子が大きく異なる可能性がある点に注意が必要である。第 2 に、上記のようなコホート研究の対象者数は約 10 万人で「大規模」であるが、希少疾患のバリデーション研究は困難である。第 3 に、背景因子やレセプト病名登録に地域差があることが予想される場合には、コホート対象地域のバリデーション研究結果をその他の集団への一般化可能性には限界がある。

このような限界に留意しつつも、大規模コホート研究におけるバリデーション研究が実施され、レセプト情報の妥当性に関するエビデンスが集積することに期待したい。

(5) 産官学・医療機関における理解の促進

2017 年の GPSP 省令の改正に伴い、医療情報データベースの活用が製造販売後調査の一つの方法として認められるに至った。バリデーション研究の必要性については、PMDA によって平成 26 年 3 月 31 日に発表された「医療情報のデータベース等を用いた医薬品の安全性評価における薬剤疫学研究の実施に関するガイドライン」や、平成 29 年 6 月 9 日に厚生労働省の 2 課長通知として発出された「製造販売後の医薬品安全性監視における医療情報データベースの利用に関する基本的考え方について」に一部盛り込まれている。しかし、より高いエビデンスレベルが求められる場合にはバリデーション研究の実施が必要であるという認識はおろか、バリデーション研究という用語自体、馴染みがない産官の関係者が多い。本報告書が、バリデーション研究の理解の促進の一助となることを期待したい。製薬産業をはじめデータベース研究を実施しうる民間企業においても、バリデーション研究の意義を理解し、その推進に協力するべく、業界内での啓発活動が求められる。医療機関において、バリデーション研究の意義、特に診療報酬請求のために発行されるレセプトのバリデーション研究がデータベース研究の質向上に必要であることへの理解促進も重要である。レセプトは本来医療費償還のための請求書であり、「請求書として正確」であることは「医学的に正確」であることとは合致しない。たとえば claims database を含む administrative database では低ナトリウム血症[47]や血清クレアチニン値によって定義される急性腎障害[18]の感度が低いことが知られているが、これは低ナトリウムや血清クレアチニン上昇自体が主たる診断(したがって医療費償還の対象)となることが多くないことを反映している。たとえ「請求書として正確」であっても、バリデーション研究における特異度や感度は 100%にはなりえないことへの理解促進が必要である。学(アカデミア)にはバリデーション研究への医療機関の協力を得つつ、社会的重要性も考慮しながらバリデーション研究を現場で推し進める役割を果たすことが求められる。

(6) 個人を識別するためのリンケージ(照合)

米国における Medicare、Medicaid、民間保険由来の医療費償還請求書(claims)や Health maintenance organization (HMO)などが有する Electronic Health Record のデータベースのデータは、(研究のために提供されるデータセットに含まれるかどうかは別として)Social Security Number などの個人ごとに付与された ID(個人 ID)とともに管理されている[20, 53, 78]。北欧[47]やカナダ[49, 54]における各種レジストリーや退院時サマリのデータベースのデータも個人 ID とともに管理されており、バリデーション研究を目的に個人 ID を用いたリンケージ(照合)が行われることがある。特に北欧では、バリデーション研究に限らず、個別の(薬剤疫学)研究ごとに、複数のデータベースを個人 ID でリンケージ(照合)し、研究用の新たなデータセットを作成して研究者に提供することが恒常的に実施されている。これに対し、日本ではデータベースのリンケージ(照合)にもとづくバリデーション研究の実施は困難である。理由の一つとして、研究者等が利用可能なデータベースの多くはそれぞれで匿名化されたデータであり、また日本には、異なるデータベース間で同一人物を紐付けるために利用可能な個人 ID が存在しないことが上げられる。また、2017 年 5 月 30 日に全面施行された改正個人情報保護法にて新たに導入された匿名加工情報の利用にあたっては、同法第 36 条および第 38 条にて「本人を識別するために、当該匿名加工情報を他の情報と照合してはならな

い」旨が定められている。

2017 年に公布された「医療分野の研究開発に資するための匿名加工医療情報に関する法律」(次世代医療基盤法)は通常では許されない個人情報の連結を可能とするものであり、国の認定を受けた「認定匿名加工医療情報作成事業者」が異なるタイプのデータベース同士をリンケージ(照合)できる可能性がある。たとえばレセプトデータと疾患レジストリーをリンケージ(照合)することが可能なら、研究の質の大幅な向上が期待できる。特にレセプトデータと複数の疾患レジストリーをリンケージ(照合)できれば、疾患レジストリーA から特定された疾患 A をもつ患者集団において(レセプトデータで特定可能な)薬の曝露がどのように疾患 B の発生というアウトカムを起こすかを疾患レジストリーB で確認する、などの研究が可能になるかもしれない。

しかし、次世代医療基盤法の下においても、バリデーション研究、特にカルテレビューによるバリデーション研究は必ずしも容易ではないと考えられる。たとえば、その第 18 条では、複数の医療機関などから個人情報を得て匿名加工医療情報を作成する「認定匿名加工医療情報作成事業者」について、「本人を識別するために、当該匿名加工医療情報を他の情報と照合してはならない」旨が定められている。ただし、運用のあり方によっては、次世代医療基盤法はカルテレビューによるバリデーション研究を含む様々な形態のレコードリンケージの根拠となりうるとも考えられる。今後、次世代医療基盤法の下でどのようなリンケージ(照合)が行われていくのかに注目したい。

その他、「人を対象とする医学系研究に関する倫理指針」の「第 17 匿名加工情報の取り扱い」では匿名加工情報を作成またはその提供を受けた研究者は「本人を識別するために、当該匿名加工情報を他の情報と照合してはならない」旨が定められている。ただし、「人を対象とする医学系研究に関する倫理指針」で言及されている「匿名加工情報」は改正個人情報保護法における匿名加工情報取扱事業者による商業的データベースなどの「匿名加工情報」に限定されることに留意する必要がある。しかし、現状に日本では匿名加工情報取扱事業者による「匿名加工情報」以外のデータベースについても、リンケージ(照合)は強く制限されている。たとえば「レセプト情報・特定健診等情報(NDB)の提供に関するガイドライン」でも、その「第 5 レセプト情報等の提供依頼申出手続」において「有識者会議が特に認めた場合を除き、提供されたその他の個体識別が可能となる可能性があるデータ(別の利用目的で提供されたその他のレセプト情報等を含む)とのリンケージ(照合)を行わないこと」とされており、「第 6 提供依頼申出に対する審査」の「4 審査基準」でも「特定個人を識別することを内容とする分析方法、手法も認めない」とされている。

バリデーション研究、特にカルテ情報をゴールドスタンダードとして実施するバリデーション研究は、匿名化されたデータベースに含まれる情報の正確性を評価するために、データベース内の個人の特定を不可欠のプロセスとする研究である。しかし現在の日本では、データベース内の個人の特定を不可欠とする研究が存在すること、さらに、その研究がデータベース研究そのものの質の向上に必須であるとの認識が十分とは言い難い。我が国のデータベース研究の質を担保するためには、本人を識別するために他の情報とリンケージ(照合)することを認め、適切なバリデーション研究を実施できる環境の整備は必要不可欠である。

5. 結論

バリデーション研究その他により、データベース研究を十分質の高い研究とすることは可能であ

り、そのようなデータベース研究は医薬品などの安全性・有効性に関するその他の方法では困難な信頼できる知見を得て、国民の健康と福祉に貢献しうる。また厳密なプライバシー保護と両立しうる方法でリンケージ（照合）を認める仕組み作りは可能である。国では 2017 年に厚生労働大臣を本部長とする「データヘルス改革推進本部」を立ち上げ、「健康医療介護分野の工程表」をまとめ、医療等 ID の導入を踏まえ、データの連結を推し進める方針でデータ連結の調査・研究も計画されている。その成果としてカルテなどを用いるバリデーション研究を含むデータのリンケージ（照合）の可能性が広がることにも期待したい。今後、データベースの適切な利用を進めようとする者は、これらの点に関する国民的理解を深め、バリデーション研究を実施しやすい環境を実現してデータベース研究の有用性を高めることに関しても努めるべきであろう。

バリデーション研究は、データベース研究のより適切な結果の解釈を可能にするが、その方法論はデータベース研究と同様に固定なものではない。現実的には、研究に用いたデータベースの種類、そのデータベースに課せられている法的および技術的制限、研究で利用できる人的および金銭的リソースなどの条件を考慮しつつ、最適な方法はそれぞれの研究ごとに熟慮されるべきものである。

この報告書がバリデーション研究の意義や基本的な考え方の理解を促し、与えられた条件下で目的を達成するための最適な方法を見出す手掛かりになることを期待する。

参考文献

1. Strom BL. Overview of Automated Databases in Pharmacoepidemiology. In Textbook of Pharmacoepidemiology. Eds Strol BL, Kimme SE.Jonh Wiley & Sons, Ltd. NJ, 2006, 167-71.
2. 小宮山靖、青木事成、古閑晃、久保田潔. より良い Pharmacovigilance Plan 策定に向けての提言.薬剤疫学 2015;20:73-83.
3. Uneno Y, Taneishi K, Kanai M, Okamoto K, Yamamoto Y, Yoshioka A, et al. Development and validation of a set of six adaptable prognosis prediction (SAP) models based on time-series real-world big data analysis for patients with cancer receiving chemotherapy: A multicenter case crossover study. PloS one. 2017;12(8):e0183291. Epub 2017/08/25. doi: 10.1371/journal.pone.0183291. PubMed PMID: 28837592; PubMed Central PMCID: PMCPMC5570326.
4. op den Winkel M, Nagel D, Sappl J, op den Winkel P, Lamerz R, Zech CJ, et al. Prognosis of patients with hepatocellular carcinoma. Validation and ranking of established staging-systems in a large western HCC-cohort. PloS one. 2012;7(10):e45066. Epub 2012/10/17. doi: 10.1371/journal.pone.0045066. PubMed PMID: 23071507; PubMed Central PMCID: PMCPMC3465308.
5. 厚生労働省医薬食品局監視指導・麻薬対策課長. 医薬品・医薬部外品製造販売業者等におけるコンピュータ化システム適正管理ガイドラインについて. 平成 22 年 10 月 21 日. http://ecompliance.co.jp/CSV/csguideline.pdf.
6. Strom BL. Validity of Pharmacoepidemiologic Drug and Diagnosis Data. In Textbook of Pharmacoepidemiology. Eds Strol BL, Kimme SE.Jonh Wiley & Sons, Ltd. NJ, 2006, 239-58.
7. ISPE. Guidelines for Good Pharmacoepidemiology Practices (GPP). https://www.pharmacoepi.org/resources/policies/guidelines-08027/.
8. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. Pharmacoepidemiology and drug safety. 2017;26(9):1033-9. Epub 2017/09/16. doi: 10.1002/pds.4297. PubMed PMID: 28913966; PubMed Central PMCID: PMCPMC5639372.
9. Semins MJ, Trock BJ, Matlaga BR. Validity of administrative coding in identifying patients with upper urinary tract calculi. The Journal of urology. 2010;184(1):190-2. Epub 2010/05/19. doi: 10.1016/j.juro.2010.03.011. PubMed PMID: 20478584.
10. Gonzalez-Fernandez M, Gardyn M, Wyckoff S, Ky PK, Palmer JB. Validation of ICD-9 Code 787.2 for identification of individuals with dysphagia from administrative databases. Dysphagia. 2009;24(4):398-402. Epub 2009/04/29. doi: 10.1007/s00455-009-9216-1. PubMed PMID: 19399554.
11. Sato I, Yagata H, Ohashi Y. The accuracy of Japanese claims data in identifying breast cancer cases. Biological & pharmaceutical bulletin. 2015;38(1):53-7. Epub 2015/03/07. doi: 10.1248/bpb.b14-00543. PubMed PMID: 25744458.
12. MacFarlane LA, Liu CC, Solomon DH, Kim SC. Validation of claims-based algorithms for gout flares. Pharmacoepidemiology and drug safety. 2016;25(7):820-6. Epub 2016/05/28. doi: 10.1002/pds.4044.

PubMed PMID: 27230083; PubMed Central PMCID: PMCPMC4930384.

13. Hsieh CY, Chen CH, Li CY, Lai ML. Validating the diagnosis of acute ischemic stroke in a National Health Insurance claims database. Journal of the Formosan Medical Association = Taiwan yi zhi. 2015;114(3):254-9. Epub 2013/10/22. doi: 10.1016/j.jfma.2013.09.009. PubMed PMID: 24140108.
14. de Achaval S, Feudtner C, Palla S, Suarez-Almazor ME. Validation of ICD-9-CM codes for identification of acetaminophen-related emergency department visits in a large pediatric hospital. BMC health services research. 2013;13:72. Epub 2013/02/26. doi: 10.1186/1472-6963-13-72. PubMed PMID: 23433397; PubMed Central PMCID: PMCPMC3585943.
15. Vanderloo SE, Johnson JA, Reimer K, McCrea P, Nuernberger K, Krueger H, et al. Validation of classification algorithms for childhood diabetes identified from administrative data. Pediatric diabetes. 2012;13(3):229-34. Epub 2011/07/21. doi: 10.1111/j.1399-5448.2011.00795.x. PubMed PMID: 21771232.
16. Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. Journal of clinical epidemiology. 1999;52(3):199-207. Epub 1999/04/21. PubMed PMID: 10210237.
17. Youngson E, Welsh RC, Kaul P, McAlister F, Quan H, Bakal J. Defining and validating comorbidities and procedures in ICD-10 health data in ST-elevation myocardial infarction patients. Medicine. 2016;95(32):e4554. Epub 2016/08/12. doi: 10.1097/md.0000000000004554. PubMed PMID: 27512881; PubMed Central PMCID: PMCPMC4985336.
18. Molnar AO, van Walraven C, McArthur E, Fergusson D, Garg AX, Knoll G. Validation of administrative database codes for acute kidney injury in kidney transplant recipients. Canadian journal of kidney health and disease. 2016;3:18. Epub 2016/04/09. doi: 10.1186/s40697-016-0108-7. PubMed PMID: 27057318; PubMed Central PMCID: PMCPMC4823855.
19. Ingeman A, Andersen G, Hundborg HH, Johnsen SP. Medical complications in patients with stroke: data validity in a stroke registry and a hospital discharge registry. Clinical epidemiology. 2010;2:5-13. Epub 2010/09/25. PubMed PMID: 20865097; PubMed Central PMCID: PMCPMC2943185.
20. Palmsten K, Huybrechts KF, Kowal MK, Mogun H, Hernandez-Diaz S. Validity of maternal and infant outcomes within nationwide Medicaid data. Pharmacoepidemiology and drug safety. 2014;23(6):646-55. Epub 2014/04/18. doi: 10.1002/pds.3627. PubMed PMID: 24740606; PubMed Central PMCID: PMCPMC4205050.
21. Bowker SL, Savu A, Donovan LE, Johnson JA, Kaul P. Validation of administrative and clinical case definitions for gestational diabetes mellitus against laboratory results. Diabetic medicine : a journal of the British Diabetic Association. 2017;34(6):781-5. Epub 2016/10/16. doi: 10.1111/dme.13271. PubMed PMID: 27743395.
22. Franchi C, Giussani G, Messina P, Montesano M, Romi S, Nobili A, et al. Validation of healthcare administrative data for the diagnosis of epilepsy. Journal of epidemiology and community health. 2013;67(12):1019-24. Epub 2013/09/12. doi: 10.1136/jech-2013-202528. PubMed PMID: 24022813.
23. Metcalfe A, Sibbald B, Lowry RB, Tough S, Bernier FP. Validation of congenital anomaly coding in Canada's administrative databases compared with a congenital anomaly registry. Birth defects

research Part A, Clinical and molecular teratology. 2014;100(2):59-66. Epub 2013/12/07. doi: 10.1002/bdra.23206. PubMed PMID: 24307632.

24. Stringer E, Bernatsky S. Validity of juvenile idiopathic arthritis diagnoses using administrative health data. Rheumatology international. 2015;35(3):575-9. Epub 2014/10/02. doi: 10.1007/s00296-014-3142-8. PubMed PMID: 25270916.
25. Dregan A, Moller H, Murray-Thomas T, Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. Cancer epidemiology. 2012;36(5):425-9. Epub 2012/06/26. doi: 10.1016/j.canep.2012.05.013. PubMed PMID: 22727737.
26. Furlan JC, Fehlings MG. The National Trauma Registry as a Canadian spine trauma database: a validation study using an institutional clinical database. Neuroepidemiology. 2011;37(2):96-101. Epub 2011/09/17. doi: 10.1159/000330835. PubMed PMID: 21921642.
27. Schwartz KL, Jembere N, Campitelli MA, Buchan SA, Chung H, Kwong JC. Using physician billing claims from the Ontario Health Insurance Plan to determine individual influenza vaccination status: an updated validation study. CMAJ open. 2016;4(3):E463-e70. Epub 2016/10/13. doi: 10.9778/cmajo.20160009. PubMed PMID: 27730110; PubMed Central PMCID: PMCPMC5047797.
28. Coleman N, Halas G, Peeler W, Casaclang N, Williamson T, Katz A. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. BMC family practice. 2015;16:11. Epub 2015/02/05. doi: 10.1186/s12875-015-0223-z. PubMed PMID: 25649201; PubMed Central PMCID: PMCPMC4324413.
29. Joseph KS, Fahey J. Validation of perinatal data in the Discharge Abstract Database of the Canadian Institute for Health Information. Chronic diseases in Canada. 2009;29(3):96-100. Epub 2009/06/17. PubMed PMID: 19527567.
30. Rudmik L, Xu Y, Kukec E, Liu M, Dean S, Quan H. A validated case definition for chronic rhinosinusitis in administrative data: a Canadian perspective. International forum of allergy & rhinology. 2016;6(11):1167-72. Epub 2016/11/04. doi: 10.1002/alr.21801. PubMed PMID: 27228224.
31. Ramalle-Gomara E, Ruiz E, Serrano M, Bartulos M, Gonzalez MA, Matute B. Validity of discharge diagnoses in the surveillance of stroke. Neuroepidemiology. 2013;41(3-4):185-8. Epub 2013/09/21. doi: 10.1159/000354626. PubMed PMID: 24051447.
32. Gabriel SE, Crowson CS, O'Fallon WM. A mathematical model that improves the validity of osteoarthritis diagnoses obtained from a computerized diagnostic database. Journal of clinical epidemiology. 1996;49(9):1025-9. Epub 1996/09/01. PubMed PMID: 8780612.
33. Husain N, Blais P, Kramer J, Kowalkowski M, Richardson P, El-Serag HB, et al. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. Alimentary pharmacology & therapeutics. 2014;40(8):949-54. Epub 2014/08/27. doi: 10.1111/apt.12923. PubMed PMID: 25155259; PubMed Central PMCID: PMCPMC5331854.
34. Verkooijen HM, Fioretta G, Chappuis PO, Vlastos G, Sappino AP, Benhamou S, et al. Set-up of a

population-based familial breast cancer registry in Geneva, Switzerland: validation of first results. Annals of oncology : official journal of the European Society for Medical Oncology. 2004;15(2):350-3. Epub 2004/02/05. PubMed PMID: 14760133.

35. Ogdie A, Alehashemi S, Love TJ, Jiang Y, Haynes K, Hennessy S, et al. Validity of psoriatic arthritis and capture of disease modifying antirheumatic drugs in the health improvement network. Pharmacoepidemiology and drug safety. 2014;23(9):918-22. Epub 2014/07/22. doi: 10.1002/pds.3677. PubMed PMID: 25044030; PubMed Central PMCID: PMCPMC4149813.
36. Harboe KM, Anthonsen K, Bardram L. Validation of data and indicators in the Danish Cholecystectomy Database. International journal for quality in health care : journal of the International Society for Quality in Health Care. 2009;21(3):160-8. Epub 2009/03/24. doi: 10.1093/intqhc/mzp009. PubMed PMID: 19304994.
37. Walsh KE, Cutrona SL, Foy S, Baker MA, Forrow S, Shoaibi A, et al. Validation of anaphylaxis in the Food and Drug Administration's Mini-Sentinel. Pharmacoepidemiology and drug safety. 2013;22(11):1205-13. Epub 2013/09/17. doi: 10.1002/pds.3505. PubMed PMID: 24038742; PubMed Central PMCID: PMCPMC4113322.
38. Widdifield J, Ivers NM, Young J, Green D, Jaakkimainen L, Butt DA, et al. Development and validation of an administrative data algorithm to estimate the disease burden and epidemiology of multiple sclerosis in Ontario, Canada. Multiple sclerosis (Houndmills, Basingstoke, England). 2015;21(8):1045-54. Epub 2014/11/14. doi: 10.1177/1352458514556303. PubMed PMID: 25392338.
39. Butt DA, Tu K, Young J, Green D, Wang M, Ivers N, et al. A validation study of administrative data algorithms to identify patients with Parkinsonism with prevalence and incidence trends. Neuroepidemiology. 2014;43(1):28-37. Epub 2014/10/18. doi: 10.1159/000365590. PubMed PMID: 25323155.
40. Widdifield J, Bombardier C, Bernatsky S, Paterson JM, Green D, Young J, et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. BMC musculoskeletal disorders. 2014;15:216. Epub 2014/06/25. doi: 10.1186/1471-2474-15-216. PubMed PMID: 24956925; PubMed Central PMCID: PMCPMC4078363.
41. Tu K, Wang M, Jaakkimainen RL, Butt D, Ivers NM, Young J, et al. Assessing the validity of using administrative data to identify patients with epilepsy. Epilepsia. 2014;55(2):335-43. Epub 2014/01/15. doi: 10.1111/epi.12506. PubMed PMID: 24417710.
42. Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. Chronic diseases and injuries in Canada. 2013;33(3):160-6. Epub 2013/06/06. PubMed PMID: 23735455.
43. Williamson T, Green ME, Birtwhistle R, Khan S, Garies S, Wong ST, et al. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. Annals of family medicine. 2014;12(4):367-72. Epub 2014/07/16. doi: 10.1370/afm.1644. PubMed PMID: 25024246; PubMed Central PMCID: PMCPMC4096475.

44. Kim SC, Gillet VG, Feldman S, Lii H, Toh S, Brown JS, et al. Validation of claims-based algorithms for identification of high-grade cervical dysplasia and cervical cancer. Pharmacoepidemiology and drug safety. 2013;22(11):1239-44. Epub 2013/09/13. doi: 10.1002/pds.3520. PubMed PMID: 24027140; PubMed Central PMCID: PMCPMC3855630.
45. Aboa-Eboule C, Mengue D, Benzenine E, Hommel M, Giroud M, Bejot Y, et al. How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke registry as control. Journal of neurology. 2013;260(2):605-13. Epub 2012/10/19. doi: 10.1007/s00415-012-6686-0. PubMed PMID: 23076827; PubMed Central PMCID: PMCPMC3566387.
46. Rezaie A, Quan H, Fedorak RN, Panaccione R, Hilsden RJ. Development and validation of an administrative case definition for inflammatory bowel diseases. Canadian journal of gastroenterology = Journal canadien de gastroenterologie. 2012;26(10):711-7. Epub 2012/10/13. PubMed PMID: 23061064; PubMed Central PMCID: PMCPMC3472911.
47. Holland-Bill L, Christiansen CF, Ulrichsen SP, Ring T, Jorgensen JO, Sorensen HT. Validity of the International Classification of Diseases, 10th revision discharge diagnosis codes for hyponatraemia in the Danish National Registry of Patients. BMJ open. 2014;4(4):e004956. Epub 2014/04/25. doi: 10.1136/bmjopen-2014-004956. PubMed PMID: 24760354; PubMed Central PMCID: PMCPMC4010845.
48. Vasta R, Boumediene F, Couratier P, Nicol M, Nicoletti A, Preux PM, et al. Validity of medico-administrative data related to amyotrophic lateral sclerosis in France: A population-based study. Amyotrophic lateral sclerosis & frontotemporal degeneration. 2017;18(1-2):24-31. Epub 2016/11/01. doi: 10.1080/21678421.2016.1241280. PubMed PMID: 27797285.
49. Jiang J, Southern D, Beck CA, James M, Lu M, Quan H. Validity of Canadian discharge abstract data for hypertension and diabetes from 2002 to 2013. CMAJ open. 2016;4(4):E646-e53. Epub 2016/12/27. doi: 10.9778/cmajo.20160128. PubMed PMID: 28018877; PubMed Central PMCID: PMCPMC5173472.
50. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes care. 2002;25(3):512-6. Epub 2002/03/05. PubMed PMID: 11874939.
51. Tu K, Wang M, Young J, Green D, Ivers NM, Butt D, et al. Validity of administrative data for identifying patients who have had a stroke or transient ischemic attack using EMRALD as a reference standard. The Canadian journal of cardiology. 2013;29(11):1388-94. Epub 2013/10/01. doi: 10.1016/j.cjca.2013.07.676. PubMed PMID: 24075778.
52. Cutrona SL, Toh S, Iyer A, Foy S, Cavagnaro E, Forrow S, et al. Design for validation of acute myocardial infarction cases in Mini-Sentinel. Pharmacoepidemiology and drug safety. 2012;21 Suppl 1:274-81. Epub 2012/01/25. doi: 10.1002/pds.2314. PubMed PMID: 22262617; PubMed Central PMCID: PMCPMC3679667.
53. Wahl PM, Terrell DR, George JN, Rodgers JK, Uhl L, Cataland S, et al. Validation of claims-based diagnostic codes for idiopathic thrombotic thrombocytopenic purpura in a commercially-insured population. Thrombosis and haemostasis. 2010;103(6):1203-9. Epub 2010/03/31. doi:

10.1160/th09-08-0595. PubMed PMID: 20352159.

54. Lacasse Y, Daigle JM, Martin S, Maltais F. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. Canadian respiratory journal. 2012;19(2):e5-9. Epub 2012/04/27. PubMed PMID: 22536584; PubMed Central PMCID: PMCPMC3373291.
55. Baghestan E, Bordahl PE, Rasmussen SA, Sande AK, Lyslo I, Solvang I. A validation of the diagnosis of obstetric sphincter tears in two Norwegian databases, the Medical Birth Registry and the Patient Administration System. Acta obstetricia et gynecologica Scandinavica. 2007;86(2):205-9. Epub 2007/03/17. doi: 10.1080/00016340601111364. PubMed PMID: 17364284.
56. Rishi MA, Kashyap R, Wilson G, Hocker S. Retrospective derivation and validation of a search algorithm to identify extubation failure in the intensive care unit. BMC anesthesiology. 2014;14:41. Epub 2014/06/04. doi: 10.1186/1471-2253-14-41. PubMed PMID: 24891838; PubMed Central PMCID: PMCPMC4041644.
57. Mamtani R, Haynes K, Boursi B, Scott FI, Goldberg DS, Keefe SM, et al. Validation of a coding algorithm to identify bladder cancer and distinguish stage in an electronic medical records database. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2015;24(1):303-7. Epub 2014/11/13. doi: 10.1158/1055-9965.epi-14-0677. PubMed PMID: 25389114; PubMed Central PMCID: PMCPMC4294969.
58. Fombonne E, Heavey L, Smeeth L, Rodrigues LC, Cook C, Smith PG, et al. Validation of the diagnosis of autism in general practitioner records. BMC public health. 2004;4:5. Epub 2004/04/29. doi: 10.1186/1471-2458-4-5. PubMed PMID: 15113435; PubMed Central PMCID: PMCPMC394332.
59. Klungsoyr K, Harmon QE, Skard LB, Simonsen I, Austvoll ET, Alsaker ER, et al. Validity of pre-eclampsia registration in the medical birth registry of norway for women participating in the norwegian mother and child cohort study, 1999-2010. Paediatric and perinatal epidemiology. 2014;28(5):362-71. Epub 2014/07/22. doi: 10.1111/ppe.12138. PubMed PMID: 25040774; PubMed Central PMCID: PMCPMC4167249.
60. Ducharme R, Benchimol EI, Deeks SL, Hawken S, Fergusson DA, Wilson K. Validation of diagnostic codes for intussusception and quantification of childhood intussusception incidence in Ontario, Canada: a population-based study. The Journal of pediatrics. 2013;163(4):1073-9.e3. Epub 2013/07/03. doi: 10.1016/j.jpeds.2013.05.034. PubMed PMID: 23809052.
61. Hall GC. Validation of death and suicide recording on the THIN UK primary care database. Pharmacoepidemiology and drug safety. 2009;18(2):120-31. Epub 2008/12/26. doi: 10.1002/pds.1686. PubMed PMID: 19109801.
62. Kumamaru H, Judd SE, Curtis JR, Ramachandran R, Hardy NC, Rhodes JD, et al. Validity of claims-based stroke algorithms in contemporary Medicare data: reasons for geographic and racial differences in stroke (REGARDS) study linked with medicare claims. Circulation Cardiovascular quality and outcomes. 2014;7(4):611-9. Epub 2014/06/26. doi: 10.1161/circoutcomes.113.000743. PubMed PMID: 24963021; PubMed Central PMCID: PMCPMC4109622.

63. Jorgensen LK, Dalgaard LS, Ostergaard LJ, Andersen NS, Norgaard M, Mogensen TH. Validity of the coding for herpes simplex encephalitis in the Danish National Patient Registry. Clinical epidemiology. 2016;8:133-40. Epub 2016/06/23. doi: 10.2147/clep.s104379. PubMed PMID: 27330328; PubMed Central PMCID: PMCPMC4896464.
64. Tairou F, De Wals P, Bastide A. Validity of death and stillbirth certificates and hospital discharge summaries for the identification of neural tube defects in Quebec City. Chronic diseases in Canada. 2006;27(3):120-4. Epub 2007/02/20. PubMed PMID: 17306063.
65. Greenland S, Lasth TL. Disease Misclassification. In Modern Epidemiology 3rd ed. Eds Rothman KJ, Greenland S, Lasth TL, Lippincott Williams & Wilkins PA, 2008. 358-60.
66. Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. Journal of Clinical Epidemiology 2003;56(11):1129-35.
67. Turner RM, Chen YW, Fernandes AW. Validation of a Case-Finding Algorithm for Identifying Patients with Non-small Cell Lung Cancer (NSCLC) in Administrative Claims Databases. Front Pharmacol. 2017 Nov 30;8:883.
68. Greenland S, Lasth TL. Relation of Predictive Values to Sensitivity and Specificity. In Modern Epidemiology 3rd ed. Eds Rothman KJ, Greenland S, Lasth TL, Lippincott Williams & Wilkins PA, 2008. 357-8.
69. Goldberg DS, Lewis JD, Halpern SD, Weiner MG, Lo Re V, 3rd. Validation of a coding algorithm to identify patients with hepatocellular carcinoma in an administrative database. Pharmacoepidemiology and drug safety. 2013;22(1):103-7. Epub 2012/11/06. doi: 10.1002/pds.3367. PubMed PMID: 23124932; PubMed Central PMCID: PMCPMC3540172.
70. Ooba N, Setoguchi S, Ando T, Sato T, Yamaguchi T, Mochizuki M, et al. Claims-based definition of death in Japanese claims database: validity and implications. PloS one. 2013;8(5):e66116. Epub 2013/06/07. doi: 10.1371/journal.pone.0066116. PubMed PMID: 23741526; PubMed Central PMCID: PMCPMC3669209.
71. 山口拓洋、冨士武史、赤木將男、安部靖之、中村真潮、山田典一ほか. 医療情報データベースを用いた静脈血栓塞栓症発症、出血性イベントのバリデーション研究. 医薬品情報学 2015; 17:87-93.
72. Tanaka S, Hagino H, Ishizuka A, Miyazaki T, Yamamoto T, Hosoi T. Validation Study of Claims-based Definitions of Suspected Atypical Femoral Fractures Using Clinical Information. 薬剤疫学 2016; 21:13-9.
73. Yamana H, Moriwaki M, Horiguchi H, Kodan M, Fushimi K, Yasunaga H. Validity of diagnoses, procedures, and laboratory data in Japanese administrative data. Journal of epidemiology. 2017;27(10):476-82. Epub 2017/02/01. doi: 10.1016/j.je.2016.09.009. PubMed PMID: 28142051; PubMed Central PMCID: PMCPMC5602797.
74. Ishii M. DRG/PPS and DPC/PDPS as Prospective Payment Systems. Japan Medical Association journal : JMAJ. 2012;55(4):279-91. Epub 2012/07/01. PubMed PMID: 25237234.
75. Okamura S, Kobayashi R, Sakamaki T. Case-mix payment in Japanese medical care. Health policy. 2005;74(3):282-6.

76. Matsuda S, Ishikawa KB, Kuwabara K, Fujimori K, Fushimi K, Hashimoto H. Development and use of the Japanese case-mix system. Eurohealth. 2008; 14:25-30.
77. 井上真奈美. 日本におけるがんの大規模コホート研究 : 歴史と展開. 日本保険医学会誌 2015;113:147-57.
78. Setoguchi S, Solomon DH, Glynn RJ, Cook EF, Levin R, Schneeweiss S. Agreement of diagnosis and its date for hematologic malignancies and solid tumors between medicare claims and cancer registry data. Cancer causes & control : CCC. 2007;18(5):561-9. Epub 2007/04/21. doi: 10.1007/s10552-007-0131-1. PubMed PMID: 17447148.

「日本における傷病名を中心とするレセプト情報から得られる指標のバリデーションに関するタスクフォース」報告書

付録 1:同じ感度・特異度であっても研究集団の曝露の割合によって研究結果が変わるケース・コントロール研究の例

Strom 編の Pharmacoepidemiology 第 5 版の第 41 章に示された、曝露の誤分類がケース・コントロール研究の結果に与える影響について解説する。ここで解説するのと同様の問題はアウトカムの誤分類でも起こりうる。

以下のケース・コントロール研究で示すように、同じ感度・特異度であっても、それらの誤分類が結果に与えるバイアスの程度は、研究集団における曝露の割合等の影響を受ける。従って、これらの指標の絶対値にはあまり意味が無く、究極の基準は効果の測定に影響するバイアスの程度である。

【セッティング】

- 真のオッズ比(OR):3.0、ケースとコントロール各 1,000 例
  - 曝露の測定への妥当性尺度:
    - 感度:(ケース)0.9、(コントロール)0.8
    - 特異度:(ケース)0.95、(コントロール)0.99

ソース集団(およびコントロール)の曝露割合 = 10%

真)

	曝露	非曝露
Case	250	750
Control	100	900

OR =3.0

測定値)

	曝露	非曝露
Case	262 [a]	738 [b]
Control	89 [c]	911 [d]

OR =3.6

a: 250・0.9+750・(1−0.95); b: 250・(1−0.9)+750・0.95;
c: 100・0.8+900・(1−0.99);d: 100・(1−0.8)+.900・0.99

ソース集団(およびコントロール)の曝露割合 = 90%

真)

	曝露	非曝露
Case	964	36
Control	900	100

OR =3.0

測定値)

	曝露	非曝露
Case	869 [e]	131 [f]
Control	721 [g]	279 [h]

OR =2.6

e: 964・0.9+36・(1−0.95); f: 964・(1−0.9)+36・0.95;
g: 900・0.8+100・(1−0.99); h: 900・(1−0.8)+.100・0.99

(教科書を元に著者ら作成)

「日本における傷病名を中心とするレセプト情報から得られる指標のバリデーションに関するタスクフォース」報告書

付録 2:データベース研究に関連するガイドライン・ガイダンスにおけるバリデーション研究に関する記載の要約

本タスクフォースでは、データベース研究に関連する国内外のガイドライン・ガイダンス等から、アウトカムのバリデーション研究に関する言及の有無や言及されている内容について抜き出した。以下は、その重要な点についてまとめたものである。

(ア)　医療情報のデータベース等を用いた医薬品の安全性評価における薬剤疫学研究の実施に関するガイドライン(初版 平成 26 年 3 月 31 日:PMDA)

本ガイドラインは、PMDA 及び製薬企業等が医療情報のデータベースを二次利用して医薬品の安全性評価を行う際に、適切な薬剤疫学研究が実施されるよう留意事項をまとめたものであり、「3. データソース」の「(3)バリデーション」において、バリデーション研究の目的や方法論、実施する意義、課題などについて網羅的に解説している。

① バリデーション研究の必要性について

1. バリデーション研究を実施することを基本推奨する。特に診療報酬データの傷病名コードのみでアウトカムを定義する場合には、バリデーション研究が必要となることがある。また、研究実施計画書および研究結果報告書(もしくはそれに代わる記録)では、バリデーション研究の実施について下記の点について言及する:対象集団、曝露、アウトカム、または共変量等の各種定義に対するバリデーション研究について、予定も含めた実施の有無について記載(実施する場合はその研究の計画等を添付、実施しない場合にはその理由を記載)。
2. 既存のバリデーション研究の結果を利用する場合には、そうすることが適切だとした理由を記載し、参考とした資料を研究実施計画書に添付。

② アウトカム判定(基準)について

1. アウトカムの判定基準は、関連する臨床ガイドライン等に定められた標準的な診断基準があればそれを参考に作成し、関連する臨床ガイドライン等がない場合であっても、可能な限り一般的に確立された診断基準を参考にした判定基準を設定することが望ましい。
2. 臨床的知識が必要とされるため臨床医が判定者となることが多いが、結果の客観性を担保するために複数の判定者によって実施し、判定結果のバラツキを抑えるために少数の判定者によって実施することが望ましい。

③ 妥当性を示す指標について

1. 感度、特異度はバリデーション研究の対象とした集団の特性の構成に影響されない指標値である。しかし、実施可能性の観点から多くの場合、陽性的中度しか求められないことも多い。
2. 陽性的中度は、研究のために用いる特定のデータベースと定義に特異的な指標

値である。しかし、研究に使用するデータの対象者が連結不可能匿名化されている場合等は、診療録等の情報を入手し妥当性の判定を行うことができないため、他のデータベースで同様の定義に関して実施したバリデーション研究の結果を参考にすることもある。

(イ)　製造販売後データベース調査実施計画書の記載要領(平成 30 年 1 月 23 日:PMDA)

本記載要領は、「医薬品の製造販売後の調査及び試験の実施の基準に関する省令等の一部を改正する省令」(平成 29 年厚生労働省令第 116 号)に製造販売後データベース調査が規定されたことを踏まえ、医薬品の製造販売業者等が製造販売後データベース調査実施計画書を作成する際の参考として提示されている。そのため、本要領におけるバリデーション研究に関する記述も、計画書への記載方法に特化している。以下にバリデーション研究に関連する箇所を引用する。

13.8. バリデーション

<u>バリデーションの記載要領</u>

- 調査で設定した全ての定義について、バリデーションスタディに関する既存の研究報告の有無及びその概要、及びバリデーションスタディ実施予定の有無を記載すること。
- バリデーションスタディが求められる対象として、対象集団の定義、曝露の定義、アウトカムの定義、共変量等が挙げられるが、特にアウトカムの定義については、既存の研究報告の有無及びバリデーションスタディの実施の有無について明記すること。
- バリデーションスタディを実施する場合は、計画の概要を記載するか、実施計画書等を付録とすること。付録とする場合、本章にはその旨を記載すること。
- バリデーションスタディを実施しない場合は、実施しない場合であっても適切な調査が実施可能と判断した理由を記載すること。
- 既存のバリデーションスタディの結果を利用してアウトカム等を定義する場合には、そのバリデーションスタディの結果を利用することが適切な理由及びバリデーションスタディの結果を記載するとともに、参考とした資料を付録とすること。

18. 付録

- 付録は以下のようなものが想定される。
  - バリデーションスタディ実施計画書又は結果報告書、もしくはアウトカム定義等の妥当性に関する資料、関連文献等

(ウ)　Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data(May 2013:FDA)

本ガイダンスは、診療報酬請求データ・電子カルテ等の医療情報データを用いた薬剤疫学的な安全性研究の実施・報告に関するベストプラクティスを記述している。本ガイダンス中、バリデーション研究の実施方法については、紹介の要素が強い以下の一文を除き具体的な記載はなかった:
「ICD のようなコードに基づくアウトカム定義での陽性的中度の評価は、データソースからそのコードを持つ症例の全て、もしくはサンプルを抽出し、その症例が本当にそのコードされたイベントを経

*験したか確認するために、その症例の一次医療データ(通常はカルテ)をレビューする」。*一方、アウトカムの妥当性に言及した「C. Study Design: Outcome Definition and Ascertainment」の章中の項(2. Validation of Outcomes)を中心に、いくつかバリデーション研究に関連する記載が下記のとおり存在した。

① バリデーション研究の必要性について

データベース研究を実施する際のデータソースの選択においては、アウトカムの妥当性に関する以下の点を検討の上、プロトコールへの記載も考慮すべきである。

1. アウトカムや他の研究因子(例えば、曝露、重要な共変数、選択／除外基準)の妥当性検証の実施可能性。
2. 選択したデータベースで実施されたアウトカムに関するバリデーション研究を含むあらゆる関連論文とその詳細(バリデーション研究が実施された集団、データベースや時期、及びその性能特性)。
3. 事前のアウトカムのバリデーションが無い研究では、研究者が使用したアウトカム定義を適切と考えた根拠。
4. アウトカムの重症度が把握できるか(患者がアウトカムについて異なった表現で訴えたり、重症度が様々なものについては特に重要である。例えば、アレルギー反応は、皮膚の発疹からアナフィラキシーショックのような生命を脅かすものまで様々である)。

② 妥当性を示す指標について

1. 研究のデザインにおいて、研究者は選択した臨床アウトカムの妥当性が検証されているかを確認すべきである。具体的には、臨床的に適切なアウトカム定義を確立し、その定義の陽性的中度を評価する。
2. 感度も重要であり、たとえば、あるアウトカム定義が高い陽性的中度を示しても、もしそのアウトカムが臨床的な関心をあまり引かず、それゆえ見逃されるのであれば、感度は低くなりうる。
3. したがって、研究者は、(1)そのアウトカム定義の陽性的中度と研究の内部妥当性に与える潜在的な影響、(2)アウトカム定義の感度、すなわち選択されたデータソースと研究対象集団においてアウトカムが特定される程度、について議論するべきである。
4. さらに、実社会での患者集団への外的妥当性(研究結果の一般化可能性)に与える潜在的な影響も議論すべきである。

③ 診断名をアウトカム定義に用いる際の考慮事項(米国でのプラクティス)

1. 退院時診断の主病名、副病名の区分はしばしば恣意的になされる。よって、退院時診断名の順番が、それらの医学的な重要性と一致しているとは限らない。
2. 入院時の保険請求データにおける ICD コードは、一般的により信頼性が高く、より重篤な疾病を反映しやすいことから、請求データを用いてアウトカムを定義する際には、可能な限り入院データのコードを使用しバリデーションするのが良い。
   - ✓ 外来の請求データにおける ICD コードは、保険償還を最大化するため、しば

しば「upcoding」や「downcoding」がされてきた（Strom 2005）。

3. 患者が療養中に死亡した時は多くの場合でシステムに記録されるが、療養中でない時に起こった死亡は電子的なヘルスケアデータシステムに記録されないことがある。よって死亡の信頼性のある確認は、他のシステムとのリンケージを通じて実施されることが好ましい。

（エ） Guideline on good pharmacovigilance practices (GVP): Module VIII – Post-authorisation safety studies (Rev 3)

European Medicines Agency (EMA)は、Post-authorisation safety studies（PASS）の実施についてガイダンスを発出している。このガイダンスには、バリデーション研究そのものの説明や実施方法に関する記述はないが、バリデーション研究に関連する記述はある。

バリデーション研究の必要性については以下のように記されている：*「利用するデータベースが研究で求められる程度に詳細で正確な情報を有していない場合、カルテに戻り診断名の確認を考慮する。関心のあるアウトカムによっては、症例ごとに行う判断もしくはランダムサンプリングした症例レビューによりバリデーションをしてもよい。」*

また、研究計画書には、データソース（1, 2）と品質管理（3）に関連して以下を記載すべきとある。

1. 利用するデータソースに含まれる記録やコーディングの妥当性に関するあらゆる情報
2. 診断名のバリデーションのための専門家委員会の関与の有無
3. 品質管理に関して、エンドポイントのバリデーションを含めたデータの質を確認する方法や手順

最終研究報告書にも、コーディングの妥当性に関して記述し、それらが研究に及ぼす限界について考察するべきである。

（オ） Scientific guidance on post-authorisation efficacy studies

EMA は Post-authorisation efficacy studies のガイダンスも公開しているが、その中でバリデーション研究に直接的に言及する記述は存在しない。医薬品のベネフィットを評価する観察研究においては、高い精度の曝露とアウトカムに関する情報が必要であるとの記述にとどまっている。

（カ） ENCePP: Guide on Methodological Standards in Pharmacoepidemiology (Revision 6)

本ガイダンスでは、以下のようなバリデーション研究に関連する記述が見られた。

1. 医療データベースを用いたアウトカムの適切な評価は、研究自体の正当性を示す上で極めて重要である。
2. アウトカム定義で用いられる因子の完全性と妥当性について十分な検討が必要であり、症例定義やアルゴリズムに用いられた仮説（前提）は検証されるべきである。バリデーションの実施方法については、専門家によるカルテレビューだけで

なく、がん登録や死亡レジストリー、もしくはそれらから算出された発生率や有病割合との一致性を確認することによる方法もある。

3. 研究によく用いられるデータベースでは、過去に重要な因子（key variables）のバリデーション研究が実施されその結果が文書化されていることがあるが、そのような過去のバリデーション研究の結果を外挿する場合には、（バリデーション研究における）因子や分析内容との差異、その後の医療、診療行為、コーディングの変化の影響を考慮する必要がある。つまり、医療制度/環境とデータが生み出される手順の両面を熟知することが好ましい。

なお、ENCePP による研究計画書のチェックリスト（Revision 3）では、Section 6: Outcome definition and measurement の中に、「6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, prospective or retrospective ascertainment, use of validation sub-study)」というチェック項目が存在する。

(キ) Guidelines for Good Database Selection and use in Pharmacoepidemiology Research

このガイドラインでは、薬剤疫学研究におけるデータベースの選択と使用に関するチェックリストを示し、それぞれのチェック項目について説明している。バリデーション研究に関連するものとして、以下の記述が確認できる。

1. レセプトデータ（claims data）の診断名は、最終的な診断名ではなく、“working”な診断名であることがあるため、偽陽性や偽陰性を考慮すべきである。
2. データのバリデーションの方法として、カルテ（clinical note）や死亡診断書のような外部資料との突合せがある。電子カルテ（EMR）がデータソースの場合では、死亡統計や大規模臨床試験のメタ解析などの外部の数値データとの比較といった方法によるバリデーションを実施するべきである。
3. 妥当性を示す指標については、主要アウトカムを特定するために使われたアルゴリズム/方法の陽性的中度を算出することがある。

(ク) GPS – Good Practice in Secondary Data Analysis: Revision after Fundamental Reworking

二次データを用いた分析を実施するための基本的基準を確立することを目的に作られた文書である。「Recommendation 6.4 – Data Quality」の項に、利用可能な情報を用いデータの信頼性と妥当性を評価すべき、といったバリデーションに関する概念的な記述はあるものの、それを実行するための具体的な手段に関する記述は存在しない。

(ケ) Guidelines for Good Pharmacoepidemiology Practices (GPP)

薬剤疫学研究の質と完全性の確保に向けた最低限のプラクティスと手順を記すことを目的としている文書である。バリデーション研究に特化した記述は無いが、使用したデータベース、アウトカムの定義、妥当性の検証方法、診断のバリデーションのための専門家委員会の有無、評価手順などについて研究計画書に記述すること、としている。

(コ)　A Checklist for Retrospective Database Studies – Report of the ISPOR Task Force on Retrospective Databases

レトロスペクティブなデータベース研究を実施する上でのチェックリストとして利用されることを想定した文書である。随所でデータの質の確保についての言及がなされているが、アウトカムのバリデーション研究についての詳細な説明はない。しかし、データの質の確保の一環として、アウトカムを特定するためのコードやアルゴリズムが妥当であるという根拠は提示されるべきとしている。また、その根拠としてしばしば過去の研究が引用されるが、理想的には一次データソースに対してのバリデーションがなされるべき、と記されている。

(サ)　Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide

The Agency for Healthcare Research and Quality による比較効果研究の研究計画立案に関するガイダンスであり、データベース研究を実施する上での注意点などについて記載されている。コーディングの妥当性検証の重要性については訴えているものの、バリデーション研究自体の方法については記述されず、バリデーション研究を実施する上で参考になる論文を紹介するにとどまっている。

(シ)　The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement

STrengthening the Reporting of OBservational studies in Epidemiology（STROBE）を基にした"routinely collected health data"を用いた研究報告に関する基準である。バリデーション研究自体についての方法論などの解説はないが、関係する記載として、集団の選択に用いられたコードやアルゴリズムに関するあらゆるバリデーション研究を引用すべきであり、また、本研究のためのバリデーションを実施したがその結果を論文化していない場合は、そのバリデーションの詳細な方法や結果を（本研究の論文に）記述すべきとしている。

以上、国内外のガイドラインなどでは「アウトカムのバリデーションは重要である」という考えは共通していたが、そのための研究方法や結果の解釈の仕方についての包括的な記述は確認できなかった。その理由の一つとして、研究の目的、用いるデータソース、注目するアウトカムの医学的性質（重症度や珍しさなど）、医療制度・医事会計制度を含めたデータが収集される環境などにより最適なバリデーションの方法が異なり、画一的なプラクティスを提供することが難しいからではないかと推察される。

「日本における傷病名を中心とするレセプト情報から得られる指標のバリデーションに関するタスクフォース」報告書

付録 3:過去に実施されたバリデーション研究のまとめ

本タスクフォースでは、バリデーション研究の要点について先例から学び、今後の日本の研究に活かすことを 1 つの目標として過去のバリデーション研究をレビューした。ただし、網羅的なシステマティックレビューではなく、どのような研究があるかを俯瞰することを目的とし、対象文献数は 105 に留めた。具体的には、2016 年 12 月に PubMed で(網羅性はないが重要な論文ができるだけ含まれることが期待できる)検索式
(((sensitivity[Title/Abstract]) OR predictive value[Title/Abstract]) AND database[Title/Abstract]) AND valid*[Title] でヒットした 412 論文から、タイトル・アブストラクトで論文を絞りこんだ。その際、(特定のアウトカムやデータベースについての)システマティックレビュー論文は除外したが、これから行う研究のためのプロトコール論文は残した。また、バリデーション研究の目的(例:バリデーションされた傷病名などを今後のデータベース研究におけるアウトカム、曝露、研究対象集団の特定に使うことが想定されるか)に制限は加えなかった。その後、検索式ではヒットしなかったがタスクフォースのメンバーが把握している要重要な論文をいくつか追加した。各研究の詳細は別途エクセルファイル(付録 3 文献リスト.xlsx)に示しているが、ここでは下記の 8 個の項目に焦点を当て、各項目において重要な研究を選択して紹介することとした。なお、引用文献番号は別途エクセルファイル(付録 3 文献リスト.xlsx)と対応している。

(1) 利用する医療情報データベースの種類

アウトカム定義のために使われたデータベースは下記のように大別された。①②③は administrative data とも呼ばれる。Administrative data は患者診療に用いるのではなく、主に病院の管理目的に集めたデータである。

① 診療報酬データ(claims data):

医療機関が診療報酬を請求書するためのデータであり、外来と入院の両方を対象にしているものが多い。代表的なものとしては米国の Medicare[1–3]、Medicaid[4] 、民間の診療請求データ[5]や、台湾の National Health Research Database[6,7]、カナダの Ontario Health Insurance Plan[8,9]、Alberta 州の Data Integration, Measurement, and Reporting (DIMR)[10]などがある。

② 退院時記録(discharge data):

患者退院時に主に担当医が作成する退院時記録のデータであり、カナダの複数の州の discharge abstract database (DAD)[11–13]、デンマークの Danish National Hospital Register[14,15]、フランスの French Hospital Discharge Database[16,17]などがある。

③ 外来の administrative data:

カナダの National Ambulatory Care Reporting System (NACRS)[18]は、診療報酬請求目的ではない一般外来と救急外来の administrative data である。

④ 電子カルテ(Electronic Medical Records):

日常診療の電子カルテをデータベース化したものであり、アメリカの退役軍人データベース[19–21]、

オランダの Integrated Primary Care Information（IPCI）[22]、イタリアの Health Search/CSD Patient Database(HSD)[22,23]、カナダの Canadian Primary Care Sentinel Surveillance Network(CPCSSN)の電子医療記録[25]、ノルウェーの Patient Administration System(PAS)[26]、スペインの Minimum Basic Data Set (MBDS)[27]のほか、イギリスのプライマリーケアデータベースの The Health Improvement Network(THIN)[28-33]などが代表的である。イギリスの General Practice Research Database（GPRD）[34]もプライマリーケアデータベースの 1 つであるが、GPRD は現在他のデータベースとのリンクが行える Clinical Practice Research Datalink に進化しており、CPRD については下記の⑥に含めた。

⑤ 疾患レジストリー:

特定の疾患を持つ患者を登録しているデータベースであり、日常診療データとは特徴が異なる。ノルウェーの Medical Birth Registry of Norway[26,35,36]やカナダの糖尿病レジストリー[37]や Trauma Registry[38]、米国の Colorado Violent Death Reporting System[39]、オランダの口唇口蓋裂のレジストリー[40]、英国の National Hip Fracture Database[41]、など様々な疾患を対象としたレジストリーがある。疾患レジストリーはゴールドスタンダードの定義に使われることが多いが、疾患レジストリーそのものをバリデーションする研究も散見された。

⑥ 上記の組み合わせ:

以上の①～⑤を 2 つ以上組み合わせたデータも存在した。特にカナダでは、オンタリオ州の外来の診療報酬データと退院時記録の組み合わせ[42-44]、さらに日帰り手術データ Same Day Surgery Database をも組み合わせたデータ[45]が評価されている。同様にアルバータ州[18]、ブリティッシュコロンビア州[24]、カルガリー州[46]、マニトバ州[47]、ノバスコティア州[48]でも Discharge Abstract Database (DAD)と Physician Claims や National Ambulatory Care Reporting System (NACRS)の組み合わせ、DAD、Medical Services Insurance (MSI)、Mental Health Outpatient Information System (MHOIS)の組み合わせ[48]など、リンクされたデータを評価した研究が多い。イギリスのプライマリーケアデータベースの 1 つである Clinical Practice Research Datalink(CPRD)[49,50]も入院・外来の Hospital Episode Statistics(HES)や Cancer Registry などとリンクさせて研究に使用することができる。また、死亡診断書、警察からの報告、監察医、検視官などかの情報を集めて作成されているアメリカの National Violent Death Reporting System[39]、薬、退院時データと検査に関するデータなどからなるイタリアの Drug Administrative Database (DAD)[51]、フランスの退院時記録（hospital discharge data）と診療報酬データ（health insurance data）の組み合わせ[16]などもバリデーション研究によって評価されている。

⑦ 薬剤疫学研究のために作られたデータベース:

少数ではあったが、米国の(Mini-)Sentinel Distributed Database[52-54]やヨーロッパ数ヵ国による European Commission-funded exploring and understanding adverse drug reactions (EU-ADR) project のデータベース[22]、アメリカの Brigham and Women's Hospital's における Medicare とリンクした EMR データベース[55]など、薬剤疫学研究のために作られたデータベースも存在した。

(2) バリデーション研究のセッティング

① 国:

今回レビューした研究の中では、カナダと米国が最も多く、英国、デンマークなども多かった。

② 研究対象期間:

データの研究対象開始年は 2000 年以降に集中し、2000 年以降バリデーション研究は年々増加する傾向が見られた。データの研究対象期間は 1 年未満のものが最も多く、5 年未満が過半数を占めた。

③ 研究の"population":

研究が対象としている"population"については、国や州といった地域の住民、Health maintenance organization (HMO)などの被保険者、グループ病院などの複数施設[23]、大学病院などの単施設で診療を受けた患者[56,57]、特定の疾患を有する患者[18,42,48,58-60]など様々であった。一方で、研究において想定される population は国や地域の医療体制に依存することも多く、論文内の記載だけでは詳細不明であるものもあった。一般的には、居住地域や加入する保険などが共通する健康人を含む集団を研究対象としており、さらにそこから検討症例または研究施設の全てまたはランダムサンプリングを用いるバリデーション研究は"population-based"とみなせる。一方で、複数／単施設を対象としている研究は"population-based"ではないと判断されることが多い。しかし、ある地域において特定の疾患を持つ患者が全例その複数/単施設に集まってくる状況では、研究が"population-based"とみなせる場合があり[61]、データベースの特徴を論文もしくは他の資料を参照して判断する必要があった。

(3) アウトカムの定義

アウトカムの定義として、疾患コードである International Classification of Diseases 9th Revision (ICD-9)もしくは ICD-10 が主に使用されていたが、英国プライマリーケアデータの Read code[29-32,62,63]のようなデータベース特有の疾患コードも見られた。また、薬剤処方コード[64,65]や処置・手術コード[66-68]、検査値[15,32]を用いてアウトカムを定義している研究も存在した。疾患コードに薬剤や手術のコードを組み合わせる[55,58,69]、あるいは疾患コードと入退院記録、受診回数情報を組み合わせる[53,58,70-72]ことで、陽性的中度や感度、特異度の値を高める工夫をしている研究もあった。

(4) リンケージの有無・方法

今回レビューした研究の中では，複数（2 つ以上）のデータベースがリンクされていることが多かった。すでにリンク済みのデータベースを使用していた研究と、その研究のために後述する方法によってリンケージさせた研究があった。

リンケージの目的は主に 2 つあり，1 つはアウトカム定義に用いるデータベースとゴールドスタンダードの定義に用いるデータベースなどとのリンケージであり，もう 1 つは(1)の⑥で言及したカナダの複数の州で実施されている入院と外来の記録を組み合わせるなどのより質の高いアウトカム定義を作成するための複数のデータベース同士のリンケージである。日本における個別の病院におけるバリデーション研究では、バリデーション研究を行う外部研究者に提供するデータと病院内部のデータを研究に固有の番号でリンケージしたものも見られた[67]。

リンケージの方法は以下の 2 つであった。

① Deterministic linkage:

2つ(以上)のファイルに共通で存在する1つ以上の個人固有の識別子(社会保障番号, 住民登録番号, 姓名等)を用いてマッチングさせる方法である。単独あるいは複数のマッチングに利用した識別子の全てが一致した場合にだけ同一個人とみなす。そのため, 例えば番号や姓名のスペルに 1 つでも違いがあれば, 同一人物であってもマッチされることはない。よって, この方法では, 誤って他人とリンケージされる事は少なく同一人物との一致割合は高いが, 不正確なデータが多ければ一致割合が低くなる。また, 特異性の高い識別子が使用できない(データベース上に存在しない)場合にも, 同一人物の一致割合が低くなることが起こりうる予想される。

使用された識別子としては, 国や地域で付与された健康情報を管理するための番号(カナダ[11,12,37,46,61,73-80]やイギリス[49]), 住民登録番号(デンマーク[14,15,81]やノルウェー[35]), 社会保障番号(米国[4,39,82])などが使われていた。他に一意の識別子)("unique identification code", "unique patient number"など)(イタリア[51], 日本[43,67], オランダ[64], 台湾[7], カナダ[8,10,18,24,43,59,83], 米国[84])を使用したと記載している研究も見られた。これらの研究では, これら一意の識別子の補足として他の識別子が使われている例もあった。複数の識別子を用いた deterministic な方法と考えられる。

② Probabilistic linkage:

複数の識別子を用いてマッチングさせる方法であり, 全ての識別子が一致する必要はない。この方法では、マッチングに用いる識別子毎に(同一人物と一致する確率に応じて)マッチングへ関与する重みを変え, その重みを足し合わせて一定の閾値を超した場合に, 同一個人とみなす。使用する識別子及び閾値を決定するのはリンケージを実施する研究者である。Deterministic な方法に比べて複雑なことから, probabilistic linkage を使用している研究は少なかった(台湾[6], 米国[17])。

リンケージの方法については、詳細が記載されていない研究も多かった。特に単施設で実施された研究は明確な記載がない場合が多かったが[19,21,26,38,56,57,63,84-89]、単施設ではカルテ番号や姓名等の匿名化されていない個人情報を用いることを前提にしている(ために敢えて記載しない)可能性が考えられる。また、米国の同一 HMO 内で実施された研究[19,43,52,90]、電子カルテ(Electronic Medical Records)における診断コードなどを、同一のデータ源の検査値などによって評価した研究[32,65]でも、リンケージの方法に特に言及されていないものが見られたが同様の理由によると考えられる。

(5) ゴールドスタンダードの設定

アウトカム定義の妥当性を確認するために参照となるゴールドスタンダードの種類は下記のように大別された。

① 疾患レジストリー:

国全体または地域単位で、がん[49,67,82,91]、心筋梗塞[11,18]、脳卒中[7,17]などのよく見られる疾患から、筋萎縮性側索硬化症[16]、先天性奇形 [46]、自閉症[48]、小児糖尿病[42,47,76]、など比較的稀な疾患まで、レジストリーをゴールドスタンダートとした研究が見られた。

② 検査結果:

血清クレアチニン値に基づいた慢性腎不全[32]及び急性腎障害[59]、血清ナトリウム値に基づいた

低ナトリウム血症[15]、経口ブドウ糖負荷(OGTT)試験に基づいた妊娠糖尿病[73]をゴールドスタンダードとしている研究が見られた。また、子宮頸がんのバリデーション研究において、病理診断をゴールドスタンダードとしているものがあった[92]。

③ カルテレビュー:

最も頻度が多く、様々な疾患のバリデーション研究で行われていた。医師、特に対象疾患の専門家が直接カルテレビューをして判断しているものと、chart abstractor と呼ばれるカルテレビューのトレーニングを受けた人達が専門家の作ったチェック項目に沿って判断しているものに大別された。前者は診断が比較的難しい疾患や専門的知識を要する疾患(例:特発性肺線維症[90]、多発性硬化症[75]、化膿性汗腺炎[85]、うつの緩解[65]、等)に多く、後者は比較的よく見られる疾患や診断基準が確立している疾患(例:脳卒中[45]、上部消化管出血[19,22],てんかん[71,93]、等)に多く見られた。2人の担当者が独立してカルテレビューをおこなっている研究が多かったが、1 人でおこなった研究もあり、また記載がない論文も見られた。また、chart abstractor がスクリーニングを行い、可能性の高いと判断された症例を専門家がさらにレビューし最終判断した研究もあった[71]。

その他の詳細事項についての記載も論文によって様々で、ゴールドスタンダードの詳細(実際に用いた調査票)を論文内または Appendix に示している模範的な論文[4,53]もあれば、「カルテレビューにより目的の疾患の有無を判断しゴールドスタンダードとした」と一文のみしか記載していない論文も散見された。

④ 医師への質問票:

イギリスのプライマリーケアデータでは、傷病名を入力した医師に質問票を送り、その解答をゴールドスタンダートとする研究が多く見られた[31,33,50]。これは、一応カルテレビューになるが、傷病名を入力したプライマリケア医自身がそれを行っているため、判断にバイアスが生じている可能性が高い。

⑤ その他:

妊娠登録における妊娠中の糖尿病や喘息を薬の使用をゴールドスタンダードとして評価した研究のほか[35]、性質の違う 2 つ以上の情報を組み合わせてゴールドスタンダードを設定した研究(例:糖尿病薬の処方歴、採血結果、または糖尿病病名のついた退院歴を糖尿病のゴールドスタンダードとした研究[58])や、インフルエンザ[8,94]やうつ[62]について住民に直接質問をし得られた解答をゴールドスタンダードとした研究、データベース上の診断コードをゴールドスタンダードとし薬剤の使用によるアウトカム定義を評価した研究[64]、異なるデータベースの登録情報をゴールドスダンダードとした研究(例:保険者が有するデータに資格喪失理由として入力された死亡の情報をゴールドスタンダードとした日本のレセプトのバリデーション研究[72]や、Nursing home Minimum Data Set への登録をケアホーム居住のゴールドスタンダートとした研究[1])もあった。また、ほとんどのバリデーション研究ではゴールドスタンダードの定義は単一であったが、敢えて複数のゴールドスタンダードの定義を示し、それぞれのゴールドスタンダードに対してアウトカム定義の妥当性を検討している研究が見られた[34,44,94](例:専門医または心臓カテーテルによる虚血性心疾患の診断を"hard"なゴールドスタンダード、一般医が診断した虚血性心疾患を"soft"なゴールドスタンダードとした[44])、疾患登録または(疾患登録にない時は)カルテレビューを行って判断するとした複雑なゴールドスタンダードを用いた研究も見られた[7]。

(6) サンプリング方法とサンプルサイズ

サンプリング方法とサンプルサイズは、何をゴールドスタンダードとするかで大きく異なるため、ここでは下記の 6 つに分類した。

① カルテレビューをゴールドスタンダードとする場合:

陽性的中度のみを求める研究ではサンプルサイズは比較的小さいものが多い。米国の Mini-sentinel における陽性的中度のみを求めるバリデーション研究のデザインでは 95%信頼区間を±10%の精度で求めるためには 100 例程度のカルテレビューで十分であることが示されている[54]。予測される陽性的中度を前提に求めたい 95%信頼区間から必要なサンプルサイズを計算した論文でも、必要症例数は 77 から 250 例で十分であることが示されている[50,95]。

これに対し、陽性的中度のみならず、感度、特異度、陰性的中度をも求めようとする研究では、サンプルサイズは格段と大きくなる。カナダで行われた、ランダムサンプルで妥当性に関する 4 つの指標を求めるバリデーション研究には合計 9,500 例のカルテレビューを実施した関節リウマチのバリデーション研究[96]、7,500 例のカルテレビューをした epilepsy のバリデーション研究[71]、5,000 例のカルテレビューをした脳卒中/TIA のバリデーション研究[45]が含まれる。また単一の病院においてではあるが、25,761 例のカルテレビューにより、会陰裂傷のバリデーション研究がノルウェーで実施されている[26]。

カルテレビューを前提にサンプルサイズを小さくする試みも行われており、ノルウェーの出生レジストリーの、カルテレビュー(および退院時の診断コード)をゴールドスタンダードとする妊娠高血圧腎症(preeclampsia)のバリデーション研究では、アウトカム定義を満たす 3,500 例全例とアウトカム定義を満たさない 75,311 例からのランダムサンプル 1,840 が検討され、サンプルの割合の異なりを適切に補正して妥当性に関する 4 つの指標を求めている[36]。アウトカム定義を満たす集団と満たさない集団から異なる割合でサンプルを求める研究は、他にもいくつか見られる。例えば、カルガリー大学耳鼻科の慢性鼻副鼻腔炎のバリデーション研究ではアウトカム定義を満たす症例と満たさない症例各 100 例[10]、スペインの脳卒中のバリデーション研究ではアウトカム定義を満たす 300 例と満たさない 100 例[27]、アメリカの全米の退役軍人 500 万人のデータベースを用いて行われた非アルコール性脂肪肝疾患のバリデーション研究ではアウトカム定義を満たす 450 例と満たさない 150 例[20]をカルテレビューで検討している。これらの研究では、感度と特異度を計算する際の補正が適切に行われておらず、示された妥当性に関する指標は陽性的中度と陰性的中度以外は適切とは思われないが、カルテレビューをする症例数を減らすために、アウトカム定義を満たす症例と満たさない症例を異なる割合でサンプルした例と考えられる。

また、カナダにおける有病割合の低い多発性硬化症のバリデーション研究では、ランダムサンプルでは妥当性に関する指標を検討する上で十分なケースが得られないとの理由から、①EMR の cumulative patient profile (CPP)、②free text entries for MS-related phrases、③physician billing codes)から多発性硬化症の全症例を含むと考える"all possible cases"943 例を特定し、カルテレビューにより 247 例の多発性硬化症の症例を特定している[75]。同様のアプローチは、ICD コードとテキスト検索により腸重積症例 565 例を特定したカナダにおける研究[61]、アメリカの Reasons for Geographic and Racial Differences in Stroke (REGARDS)における脳卒中のバリデーション研究で

Medicare の診断コードと脳卒中を疑わせる症状の自己報告のいずれかに該当する例のカルテレビューで真のケースを特定した研究[2] 、3つのデータベースのいずれかで特定されたNeural Tubal Defect による死亡、死産をさらにカルテレビューで真のケースを特定した研究[97]、などでもとられている。特にREGARDSコホートを利用した研究では、primary analysisでは特定されたケースを全ケースとしているが、特定されていない真のケースがどの程度存在すると、妥当性の指標にどのような影響を与えるかの感度解析が行われている[2]。”all possible cases”のアプローチに区分可能な研究としては、複数のサブグループから異なる抽出割合で得たサンプルのカルテレビューによって急性心不全を特定した研究もあるが[98]、抽出割合が異なるために、指標の求め方は複雑である。

② 調査票またはインタビューの結果をゴールドスタンダードとする場合:

この方法を用いた場合、調査されたサンプルサイズは 500 例以下がほとんどであった。郵送による調査票の回答は英国で多く用いられている。英国ではプライマリ医によって登録される医療情報データベース(The Health Improvement Network、Clinical Practice Research Datalinkなど)が多くの医学研究に利用されており、これらのデータベース上の傷病名の妥当性研究では、該当のプライマリ医に送付された調査票の結果が用いられている。大腸がん[29]、膀胱がん[30]、乾癬性関節炎[31]、C型肝炎[33]などの疾患が対象で、いずれも100から200例程度のサンプルサイズとなっている。

またインタビュー調査の結果は、インフルエンザワクチン接種の妥当性研究で使われている。オーストラリアの妊婦を対象にした妥当性研究[94]では、協力の得られた出産した女性に対して実施された電話インタビューによるセルフレポートが使われ、サンプルサイズは563人であった。一方、カナダのオンタリオ州で実施された研究[8]では、既に実施されていた地域住民を対象としたインタビュー調査の結果が使われたため、サンプルサイズは47,301人と大きかった。

③ 電子カルテなどの診療記録をゴールドスタンダードとする場合:

電子的に記録された情報からは、コードなどで該当の情報を特定する事が容易であることから,手間と時間を省くことが可能であり、カルテレビューに比べてサンプルサイズを大きく設定することが可能である。事実、数万～数十万の規模のサンプルサイズを扱った研究もあった[9,24]。

④ 疾患登録をゴールドスタンダードとする場合:

疾患が電子的に登録されている場合には情報の特定が容易であり、大きなサンプルサイズでの研究が実施可能である。カナダで実施された妊娠糖尿病の妥当性研究[37]では州の周産期レジストリーが用いられ、そのサンプルサイズは411,390例であった。該当症例全例が使用できたため,感度・特異度・陰性的中度・陽性的中度の全てが算出可能であった。他にもがん登録が用いられたイギリスのGeneral Practitioner Research Database (GPRD)のがんの妥当性研究[49]のサンプルサイズは 42,556 例と大きい。一般的には疾患登録の精度は高いと考えられるが、地域全体のがん登録情報を用いた研究で、感度・特異度・陽性的中度・陰性的中度を算出できても、陽性的中度が低い場合がある。米国のペンシルバニア州で州のがん登録が使用された研究[82]では陽性的中度が低く、その理由として疾患登録の不完全性の可能性を指摘し、がん登録で見逃された患者の割合とそれが陽性的中度に与える影響を検討する感度分析を行っていた。また、単施設等の小規模での妥当性研究で、地域の疾患登録ではなく、当該の医療機関の疾患登録を使用している

場合はその分サンプルサイズが小さい。例えば、単施設で実施された小児糖尿病の妥当性研究[42]では 1,323 名、乳がんの妥当性研究[67]は 633 例であった。その他使われた疾患登録には，小児糖尿病疾患登録[38]や先天性奇形疾患登録[37]なども使われていた。

⑤ 検査値や画像検査の結果をゴールドスタンダードとする場合：

画像診断の結果や、血液検査等の測定値が使用されており、単独あるいは少数の医療機関で実施された研究ではこれらの情報が入手しやすかったと考えられる。それらの研究のサンプルサイズは数百例と小さかった[59,70,81,99,100]。約 25,000 症例の大きなサンプルサイズで実施された研究は、妥当性の確認の対象となるレセプトデータに元々連結している電子カルテがあり、参照が容易であったと考えられる[92]。

⑥ その他：

レセプトや電子カルテ、検査値などの複合や、アルゴリズムの作成、保険資格喪失理由の記録や、複数のデータベースの記録の有無、疾患スクリーニングの結果などが使用されており、サンプルサイズも様々であった[41,48,72,101]。

(7) 妥当性を測定する指標

妥当性（validity）の指標として、陽性的中度、陰性的中度、感度、特異度が用いられていた。その内、求めた指標（の組み合わせ）は以下の 2 パターンに大別された。

① 陽性的中度

Mini-Sentinel[53,54]、THIN[30,33]や GPRD[34]等のデータベースにおけるバリデーション研究では陽性的中度のみを求めていた。ゴールドスタンダードはカルテレビューが多く、検査結果[92]や General practitioner （GP)[33]からの質問票の回答とした研究も見られた。いずれの研究もアウトカム定義を満たす患者群からのみランダムサンプリングを行っており、感度や特異度は求められていない。

② 陽性的中度、陰性的中度、感度、特異度

アウトカム定義を満たす患者群および満たさない患者群からサンプリングを行った研究では、4 つ全ての指標を求めていた。しかし、指標全てを求められるにも関わらず、陰性的中度を求めていない研究、陰性的中度および特異度を求めていない研究があった[29]。また、一病院で特定された真のケースとその病院の患者の administrative database での該当コードを照らし合わせて感度と陽性的中度のみを求めている研究もあった[74]。

偽陽性、偽陰性の症例については、偽陽性・偽陰性となった理由などの詳細を表などに記載している研究もみられた[4, 7,8,38,46,49,51,73,74,102,103]。

4 つの指標以外に、複数人でカルテレビューを行っている研究では κ 係数を算出した研究も見られた[5,12,44,65,80,104,105]。異なる評価者間の一致度（inter-rater agreement）のみならず、同一の評価者の異なる時点での一致度（intra-rater agreement）を評価した研究もみられた[8]。

(8) 指標の閾値・利用法

感度や陽性的中度の値としてどのレベルが望ましいか、または指標の利用方法については、殆どの論文では言及はなかったが、ここでは記載のあった少数の研究を紹介する。目標とする陽性的中度を 80%以上[89]や 85%以上[80]と予め定めている研究や、複数のアルゴリズムを比較する際に

陽性的中度 70%を「低い」[67]、67%未満を「低い」[81]と判断している研究があった。予防接種有無の自己報告を実際の接種記録と比較した研究[8]では、低い感度（49.8%）かつ高い陽性的中度（88.4%）という集団であっても、自己申告で既接種と回答した部分集団は予防接種の安全性を判断するための自己対照研究(Self-Controlled Study)には用いることができる、と筆者らは考察している。また、パーキンソン病／パーキンソン症候群の疫学調査のためのバリデーション研究で「陽性的中度が高く、かつ感度特異度も高い指標を選択した」旨が記載されている研究もあった[9]。

本報告書の「2. バリデーション研究の概要」の「バリデーション研究に関する教科書のまとめ（Pharmacoepidemiology, 5th Edition（Brian L. Strom ら）」の紹介でも言及されている通り、同一の感度や特異度であっても、薬剤疫学研究の結果への影響は研究集団における曝露の割合等で変わりうる。指標の望ましい値に言及されている研究が少ないのは、「望ましい値」は研究対象集団、研究の目的などで異なりうることを反映していると考えられる。

参考文献

別途エクセルファイル（付録 3 文献リスト.xlsx）参照

文献番号	出版年	文献	対象疾患	(1) 利用する医療情報データベースの種類	(2) バリデーション研究のセッティング (i) 国(地域)	(2) バリデーション研究のセッティング (ii) 調査した年	(2) バリデーション研究のセッティング (iii) 調査した病院数	(2) バリデーション研究のセッティング (iv) "population-base"とみなせるか	(3) アウトカムの定義	(4) リンケージの有無・方法	(5) ゴールドスタンダードの設定	(6) サンプリング方法とサンプルサイズ	(7) 妥当性を測定する指標	(8) 指標の閾値・利用法	筆者らが挙げている問題点	備考
1	2008	Koroukian SM, Xu F, Murray P. Ability of Medicare claims data to identify nursing home patients: a validation study. Med Care 2008 Nov;46(11):1184-1187.	Nursing Homer Care(擁護施設におけるケア)	Medicareのcarrier file (MedicareのPart B claim file)	アメリカ オハイオ州	2001/11～2002		yes	Physician service CPT codesの"Evaluation and Management (E&M) procedure codes"	social security numberを含む複数のPatient identifiersを使用した。	(US連邦政府によるMedicare、Medicaid対象のNursing Home Careを受けている人に関する)Nursing home Minimum Data Set (MDS)を使用した。	1997-2001年にがん診断のあった人で、2002年にnursing homeに入居した65歳以上の人。N=5327	PPV/Sensitivity			
2	2014	Kumamaru H, Judd SE, Curtis JR, Ramachandran R, Hardy NC, Rhodes JD, et al. Validity of claims-based stroke algorithms in contemporary Medicare data: reasons for geographic and racial differences in stroke (REGARDS) study linked with medicare claims. Circ Cardiovasc Qual Outcomes 2014 Jul;7(4):611-619.	急性虚血性脳卒中(AIS), 頭蓋内出血(ICH)	REGARDコホートのうち、Medicareとリンク可能であり、かつMedicare PartAとPartBの期間が1か月以上ある者(Any-stroke cohort)。 このうち、cohort entryの前にself-reported physician-diagnosed strokeがある患者を除いたものがfirst-stroke cohort	米国	2003 - 2009	(地域住民ベース)	Yes Population base(REGARDSコホート)	primary discharge diagnosis としてのICD9-CMの①虚血性脳卒中、②出血性脳卒中それぞれ単独および③虚血性/出血性脳卒中 (1) acute ischemic stroke (AIS) algorithm (ICD-9-CM code 433.x1, 434. x1, or 436) (2) intracranial hemorrhage (ICH) algorithm (code 430 or 431) (3) AIS/ICH algorithm (codes 430, 431, 433.x1, 434.x1, or 436)	REGARDSデータベースとMedicareのリンケージをExact matchingで。 Exact matchingはSSN、性別、生年月日の3つの要素の2つ以上(ただし年について許容したのは最大1年の差のもののみ)	チャートレビュー self-reported eventまたはself-reported eventはないがMedicareにstrokeの診断(AIS/ICH algorhythm)がある者に関して実施したmedical recordのレビュー Stroke nurseによる明らかなnon-caseの除外ののちstroke expertsのcommitteeがレビュー。WHO-definedおよびclinical strokeを含めた。	15089名	PPV/NPV/Sensitivity/Specificity	3つのアウトカムに関して評価されたアルゴリズムはいずれも、感度は低いがPPVは高いので、比較の指標を用いる研究では有用だが、発生率の絶対値を求める研究には不向き	1) 診療録のないものは、全て、イベントを特定する過程に生じた(真の脳卒中の割合は診療録のないものとあるものとで同一)と仮定したこと。 2) 65歳以上のREGARDS登録者全てが、Medicare にリンクできた訳ではないこと。 3) first-stroke cohortにおいて、既往のある人をREGARDS studyのデータのみを使って除外していること。	primary analysisでは感度などは"all possible cases"すなわち、Claimsまたはself-reportでStroke/Stroke疑いの患者のうちカルテレビューにより真のStrokeであることが判明した患者をPopulationに含まれる全ての真のStrokeのケースとして感度などを計算している。 Secondary Sensitivity analysisでは、"no event stratum"の把握できていないケースの存在が推定された指標に与える影響を検討している。 日本で行うバリデーション研究でも対象者全員、またはランダムサンプル全員のレビューが現実的でない時に実施可能な方法として有用と考えられる。
3	2007	Zuckerman IH, Sato M, Hsu VD, Hernandez JJ. Validation of a method for identifying nursing home admissions using administrative claims. BMC Health Serv Res 2007 Dec 10;7:202-6963-7-202.	養護施設へ入所していること	MarketScan Medicare Supplemental and Coordination of Benefit (COB) database 支払の情報なども含むがenrollment informationなどを含むadministrative database	アメリカ	2000-2002		yes	Table 1に記載されたProvider type, place of service, Procedure codeを使用して養護施設への入所の定義を行った。	明確には記載されていないが、patient identifierを使ってadministrative dataであるMarketScan Medicare Supplemental and Coordination of Benefit (COB) databaseとゴールドスタンダードのservice claimsがリンクされたと思われる。	inpatient and outpatient administrative service claimsを使って1名の専門家(clinical pharmacist)がNursing Home serviceを受けているかを判断。	Nursing Homeへの入所が"PROBABLE" 50例、"POSSIBLE" 50例、"UNLIKELY" 50例の150例	Percetage of agreement/Sensitivity/Specificity		評価されたDBsにはMarketScan とMedicareの"coordination"のサービスについてのデータしか含まれておらず、Medicareで100%支払われるNursing Home Serviceについては含まれていない。	Nursing Homeへの入所が"PROBABLE" 50例、"POSSIBLE" 50例、"UNLIKELY" 50例の150例という研究者が恣意的に設定した"population"の中で感度、特異度などが求められている。
4	2014	Palmsten K, Huybrechts KF, Kowal MK, Mogun H, Hernandez-Diaz S. Validity of maternal and infant outcomes within nationwide Medicaid data. Pharmacoepidemiol Drug Saf 2014 Jun;23(6):646-655.	子癇前症 新生児持続性肺高血圧 心臓奇形	US Medicaid Analytic eXtract	米国, 32 states	2000-2007	168	yes	入院または外来のICD-9コード 子癇は妊娠140日～出産後30日 新生児持続高血圧症は出産後30日以内 心奇形は出産後90日以内 子供の病気が母親のデータの中にある可能性を考慮して、母親のデータの中に入力されている新生児持続性肺高血圧と心臓奇形のコードもチェックしている	母と子供のSocial Security Number (SSN) 及び生年月日	カルテレビュー	もともとの母数は書いてない(数百～千万レベル?)が、available recordsがない症例を除いた後(available recordsがある人とない人の特徴が違わないことを確認)最終的な各病気が100－200になることを目標に、各病院のレスポンスレートを予想して大目にサンプリングして各病院に申請、結果的に 子癇前症183人 先天性心奇形158人 新生児持続性肺高血圧82人	PPVのみ	入院症例、出産後ほかの施設にうつらなかった症例を対象にするとPPVがあがる		今後の研究でうつ病・抗うつ剤との関連を検討することを意識してか、validationもうつ病・抗うつ剤の有無で層別化している medical recordのデータが得られた(response rate)のはわずか44%。SSNだけでは不十分で名前が必要、本人のICが必要などが得られなかった理由に含まれる。
5	2010	Wahl PM, Terrell DR, George JN, Rodgers JK, Uhl L, Cataland S, et al. Validation of claims-based diagnostic codes for idiopathic thrombotic thrombocytopenic purpura in a commercially-insured population. Thromb Haemost 2010 Jun;103(6):1203-1209.	特発性血小板減少性紫斑病(TTP)	HealthCore Integrated Research Database (HIRD)	米国の商業用の保険データベース (Southeastern, Mid-Atlantic, Northeastern, Central, Western Regions)	2001年1月から2008年5月			特発性血小板減少性紫斑病(TTP)のICD-9 446.6X が一回以上入院データベースに入っていること ICD-9とPlasma Exchangeなどのprocedure codeとの組合わせなども検討	"contact information"を得たと記載がある。同一HMO内なので容易だったと思われる	専門家2名がmedical records abstraction formを作成。A chart abstraction vendorがabstraction。医師2名(血液内科医)が評価(TTP)	2001-8年にデータベースで特定された714人のTPP病名のついた患者より、206人のチャートレビューできた患者	PPV	アルゴリズムを厳しくする(refinement)すると、PPVがあがる。PPVはICD単独で45.5%(86/189)。血漿交換と組み合わせると65%(67/103)に、さらにTTPと間違えそうな病名の除外により72.3%(34/47)になる。	2人の医者のカルテReviewのagreementがよくない 確立した診断基準がなく、類似疾患が多いことが要因と分析されている。	Abstractorがデータを抽出し、二人の専門家が独立に評価し、κ係数を求めている。
6	2015	Cheng CL, Chien HC, Lee CH, Lin SJ, Yang YH. Validity of in-hospital mortality data among patients with acute myocardial infarction or stroke in National Health Insurance Research Database in Taiwan. Int J Cardiol 2015 Dec 15;201:96-101.	AMI、Strokeの入院患者の死亡 (死亡例538例における合併症についても感度、PPVを求めている。)	National Health Insurance Research Database (NHIRD;全民健康保険研究資料庫)	台湾	2005 - 2010	1	No 台湾南部の三次医療施設1機関での研究。台湾南部の心筋梗塞、脳卒中の全ケースが入院するとは考えられず、Population-basedといえるかについては疑問	ICD9診断名 AMI (ICD-9-CM code: 410xx), hemorrhagic stroke (ICD-9-CM code: 430xx, 431xx, 432xx), or ischemic stroke (ICD-9-CM code: 433xx, 434xx) 死亡および退院時の合併症 ①死亡退院のコード 死亡については退院時のStatus"4"(死亡)、"A"(terminal stageのため退院)、"5"(medical adviseに反する自主的退院)を評価 ②イベント発生3日以内のNHIからの登録削除 ③CIRD(catastrophic illness registry data)での死亡情報	NHIRDのデータに含まれていた情報により特定の病院を特定。 さらにProbabilistic record-linkage (誕生日、性別、入院日、退院日)	EMRにおける「死亡」	6418名(1579 急性心筋梗塞, 1444 出血性脳卒中, 3395 虚血性脳卒中) がNHIRDから抽出された	死亡のPPV (ほかに死亡538例について、合併症のPPV/Sensitivity)	NHIRD に死亡診断はよく反映されている。 死亡の精度は、NHIからの登録削除を加えると良くなったことから、薬剤疫学研究で入院患者の死亡を評価するのに有用だろう。	単一施設のデータであること、死亡の定義が必ずしも死亡を意味しないこともありうること。	
7	2015	Hsieh CY, Chen CH, Li CY, Lai ML. Validating the diagnosis of acute ischemic stroke in a National Health Insurance claims database. J Formos Med Assoc 2015 Mar;114(3):254-259.	急性虚血性脳卒中	National Health Insurance Research Database	台湾	2006年8月-2008年12月	1	no	急性脳梗塞を示唆するICD-9-CM code(433xx or 434xx)が退院時病名の5つの中のどこかにあること	Bureau of NHIに報告された(院内の)claims dataとTaiwan Stroke Registry (TSR) のそれぞれで encrypted identifierを用いて匿名化し、リンケージに使った。 [久保田注:最終的には削除] TSR+でclaim(-)が42例いたが、リンクできなかったということではない?	TSR とリンクできた場合は、その中の診断がゴールドスタンダード TSRとリンクできなかった場合は、神経学者による電子的な退院情報と脳画像(CTかMRI)のレビュー	1病院の中でのNational Health Insurance Research DatabaseでICD9コードがついている人全例1736人と+Taiwan Stroke Registry (TSR) に登録されている人全例1341人(かなり重複があり、最終的に1778人、そのうち真のケースは1576人)	SensitivityとPPV(とfalse positive rate=1-PPV)			TSRにあるか、Claims+でnurologistのレビューにより特定されたケースが真のAISの全ケースとするall possible casesのアプローチ。ほかにも真のケースがありうることが言及されていない。 二つのゴールドスタンダード(TSRまたはneurologistのレビュー)が使われている点も問題。Limitationsの二点目で言及はされているがダブルスタンダードであることは認識されていない。 Claims dataはNHIRDからの抽出ではなく、病院内に保管されているBureau of NHIに報告されたclaims dataを用いており、日本で可能なバリデーション研究と同様。 Table 1(とTable 2)にFalse positive cases202例について、FPの理由が記載されている。Precerebral of cerebral artery occlusion without cerebral infarction21.8%, subacute ischemic stroke17.8%、old ischemic stroke 15.3%
8	2016	Schwartz KL, Jembere N, Campitelli MA, Buchan SA, Chung H, Kwong JC. Using physician billing claims from the Ontario Health Insurance Plan to determine individual influenza vaccination status: an updated validation study. CMAJ Open 2016 Aug 22;4(3):E463-E470.	インフルエンザワクチン接種	Ontario Health Insurance Plan (OHIP) databaseにおいて、インフルエンザ予防接種の医事会計請求	カナダ Ontario州	2007年1月 - 2009年9月	地域住民	yes	インフルエンザ予防接種のbilling codes (G590, G591, 追跡コードQ130(別の所で予防接種を受けた者に付与するコード)	Institute for Clinical Evaluative Sciences (ICES)のOHIPと"unique encoded identifiers" を用いてリンケージ。	The Canadian Community Health Survey (インタビューによる住民本人からの直接取得したインフルエンザワクチン接種に関する情報)	住民からサーベイで得た情報 provincial health administrative dataとリンクすることに同意した47301人の住民	PPV/NPV/Sensitivity/Specificity	感度は低い(49.8%)がPPVは高い(88.4%)。 ワクチンの安全性を評価するようなSelf-controlled study designには使える		感度、特異度から真の接種割合を推定するFoxらの方法でワクチンの有効性を求めている。
9	2014	Butt DA, Tu K, Young J, Green D, Wang M, Ivers N, et al. A validation study of administrative data algorithms to identify patients with Parkinsonism with prevalence and incidence trends. Neuroepidemiology 2014;43(1):28-37.	パーキンソン病(PD)/パーキンソン症候群(PKM)	診療報酬、調剤薬局、入院などの5つのデータベース(CIHI DAD[入院]、OHIP physician claims[外来]、NACRS[病院ベースの外来]、SDS[外来手術]、ODBP[調剤])	カナダ オンタリオ	2011年6-11月	84人のfamily physiciansのEMRのデータのEMRALDによるadministrative dataによるアウトカム定義の評価。	yes	22のアルゴリズムでPD/PKMを定義 ICD-9codes (332.0, 332.1)、ICD-10 codes (G20, G21.0-0.4, G21.8-9, G22, F02.3)、OHIP code (332) physician billing codesについては特定の期間におけるコード数 (例:physician billing codeの±6 monthsの期間)、退院時の記録(におけるコード)の有無、専門家の関与、治療薬の有無(≥ 65歳の集団で)。20歳以上の集団と65歳以上の集団(治療薬の情報使用可能)で評価。	Institute for Clinical Evaluative Sciences (ICES)において"an encrypted, anonymous manner"でリンクした。	オンタリオ州の家庭医のデータ(EMRALD)	73,003のレコードから"automated electronic text-matching search strategy"で2319人の候補を抽出。その上で、6人のレビューワーがchart review。"definite"(231:191PDと40PKM)、"possible"(294)、"not"(118)、"no mention"(1676)に区分。5%サンプルでintra- inter-rater reliabilityを評価。 最終的には全ての記録を専門家がrevalidateし、判断の食い違いを解決。	感度、特異度、陽性的中度 intra- inter-raterの一致についてはκ	PPVが高く、かつ感度、特異度も高い指標を選択した。 「1年以内に2回以上の医師への支払い記録にコードがある」が有用。	調剤データベースには65歳以上しか含まれていない	
10	2016	Rudmik L, Xu Y, Kukec E, Liu M, Dean S, Quan H. A validated case definition for chronic rhinosinusitis in administrative data: a Canadian perspective. Int Forum Allergy Rhinol 2016 Nov;6(11):1167-1172.	慢性鼻副鼻腔炎 (CRS)	アルバータ州の皆保険のデータ(Data Integration, Measurement, and Reporting (DIMR))	カナダ カルガリー大学病院 鼻科	2011年7月から2015年6月(カルガリー大学) 2010年3月から2015年3月(administrative database)			ICD-9 diagnosis codes for CRS (471.x and 473.x)	unique lifetime identifier (ULI)	カルガリー大学耳鼻科に該当期間に受診した2167人についてチャートレビュー(ガイドラインに準じた診断)	2011.7から2015.6までに、カルガリー大学鼻科に受診した(sinonasal complainsで紹介された)2167名の成人患者。 CRS群、非CRS群に分け、それぞれ100名をランダムサンプリング。	Sensitivity/Specificity/PPV/NPV	2年間に外来ClaimでCRSで受診2回以上(a CRS case definition that involves 2 physician claims of CRS (ICD-9 code 471.x or 473.x) within 2 years) がいいといえる		"Population"は一大学の耳鼻科を(sinonasal complainsで紹介され)受診した2167名の成人患者指標を求めた"population"はおそらくCase100名とControl 100名の200名で抽出割合の違いは考慮されていない?
11	2016	Jiang J, Southern D, Beck CA, James M, Lu M, Quan H. Validity of Canadian discharge abstract data for hypertension and diabetes from 2002 to 2013. CMAJ Open 2016 Oct 28;4(4):E646-E653.	高血圧、糖尿病	退院情報データベース(Discharge Abstract Database)	カナダ Alberta州	2002 to 2013		yes	高血圧 (ICD-10 codes I10.x, I11.x-I13.x and I15.x) 糖尿病 (ICD-10 codes E10.x, E13.10, E13.12, E14.10, E14.12, E11.x, E13.0 and E14.0)	Discharge Abstract Database とAPPROACHデータベースを、patient provincial health numberでリンク	Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) database,	63,483名の患者が、2つのデータベースでリンクされた	PPV/NPV/sensitivity /specificity	2002年から2013年にわたり、Discharge Abstract Databaseの質は高く、安定していた。	この期間中、糖尿病や高血圧の定義が若干変化したが、その変化が疾患の有病率を変えたか否かは分からない。 また、この期間中の外来患者を検証していない。冠動脈カテーテルを行った患者でしか検証していない。	
12	2015	Jolley RJ, Quan H, Jette N, Sawka KJ, Diep L, Goliath J, et al. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. BMJ Open 2015 Dec 23;5(12):e009487-2015-009487.	敗血症	退院サマリデータベース(DAD)	カナダ Alberta州	2009-2012	3	分からないが、おそらくNo 二つのコホート。①Calgaryの3つの病院のICUに入院した18歳以上の成人、②Foothills Medical Centre (FMC)のnon-ICUの入院患者。Generalizabilityに問題がある点については著者ら自身が"Limitations"で言及。	25のcoding fieldsにあるいずれかの敗血症のコード(ICD10CA)。さらにorgan dysfunction code.があれば"severe"の敗血症。重症度について特定されない敗血症とsevere敗血症両者が評価されている。 Table 2 の右側に示されている著者らが追加したコードを含む(optimised)"New ICD-10-CA"についても評価している。	アルベルタのAlberta personal health number	ICU敗血症臨床データベース, メディカルチャート	3病院からのランダムサンプリングで 1001 ICU patients, 202 non-ICU patients	PPV/NPV/Sensitivity/Specificity 評価者間のκについては、方法において、ゴールドスタンダードのChart Reviewを改善していく過程に関する記述において言及されている。	敗血症はundercodedになりがちだが, 特異度と陰性的中度が高いので, 敗血症のコホートを特定するには, 感度がより高い"optimised"のICD10CAコードによる敗血症の定義は有用。 ICD case definitionのOptimisedの感度71.9%について"Discussion"で"only moderately sensitive"と評価。	チャートレビューによる診断の誤分類の可能性。3病院からのサンプルであるため一般化可能性の問題がある	
13	2015	Pang JX, Ross E, Borman MA, Zimmer S, Kaplan GG, Heitman SJ, et al. Validation of coding algorithms for the identification of patients hospitalized for alcoholic hepatitis using administrative data. BMC Gastroenterol 2015 Sep 11;15:116-015-0348-5.	アルコール性肝炎	退院アブストラクトデータベース(DAD)	カナダ Alberta州 カルガリー地域	2008/1-2012/8	記載なし	yes	肝炎のICD-10コード(K70.1)	記載なし	医療記録のレビュー(アルコール消費量、ASTまたはALT、ビリルビン、INR、急性肝障害による定義)	2008年1月から2012年8月に肝炎のICD-10コード(K70.1)を付与されて入院した18歳以上228例	陽性的中度 評価者間のκ係数も出している	肝炎のICD-10コードを使う場合、医療記録のレビューも行った方がいい	生検実施例が少なく、アルコール性肝炎(AH)の診断が不十分であり、カルテレビューにおける真のケースの判断も「生検の結果を含まない5つの基準の全てを満たす」が用いられた。Primary diagnosisがAHである例に限ると、また定義に腹水、消化管出血を含めると、PPVは上がる。	
14	2010	Djurhuus BD, Skytthe A, Faber CE. Validation of the cholesteatoma diagnosis in the Danish National Hospital Register. Dan Med Bull 2010 Oct;57(10):A4159.	真珠腫(cholesteatoma)	Danish National Hospital Register (DNHR)	デンマークのDanish National Hospital Registerの「真珠腫」の正確性検討を1病院のOdense University Hospital (OUH)の症例				ICD8かICD10の中耳か乳頭状のコレステリン腫かつそれに関連する手術コードの記録がある	individual's civil registration number	Odense 大学病院の耳鼻科の Database (EDOUH)を用いた(電子的またはpaper formの)チャートレビュー。外科医の記録を重視した。	ランダムサンプリングされた286人。	PPV			
15	2014	Holland-Bill L, Christiansen CF, Ulrichsen SP, Ring T, Jorgensen JO, Sorensen HT. Validity of the International Classification of Diseases, 10th revision discharge diagnosis codes for hyponatraemia in the Danish National Registry of Patients. BMJ Open 2014 Apr 23;4(4):e004956-2014-004956.	低ナトリウム血症	Danish National Registry of Patients (DNRP)	デンマーク	2006-2011	記載なし	yes	退院時傷病名の主傷病名か副病名にICD10の低ナトリウム血症のコード(E87.1,E87.1A,P74.2B)	unique 10-digit civil registration number (assigned to all Danish residents since 1968)を使用,	laboratory information systems (LABKA)の血清ナトリウム値が <135 mmol/L (生後30日以内の幼児は<133 mmol/L)	2 186 642入院全例(819 701患者)	PPV/NPV/Sensitivity/Specificity	カットオフを下げると、Sensitivityがあがることを報告	低ナトリウム血症の長さ(duration).は考慮していない 感度が重症のナトリウム血症を含めて低い理由は低ナトリウム血症の多くが基礎疾患の"bystander"であるためと推測している。	(Na値<135に対して)感度1.8% (95% CI 1.7% to 1.8%)は低すぎる
16	2016	Vasta R, Boumediene F, Couratier P, Nicol M, Nicoletti A, Preux PM, et al. Validity of medico-administrative data related to amyotrophic lateral sclerosis in France: A population-based study. Amyotroph Lateral Scler Frontotemporal Degener 2017 Feb;18(1-2):24-31.	筋萎縮性側索硬化症 (ALS)	hospital discharge data (HDD) とhealth insurance data (HID)	フランスLimousin地域のregional hospitals			Yes. フランスのLimousin地方	HDDに'G12.2'(ICD-10)もしくはHIDに'G12' HID code があること。	特に記述なし	疾患レジストリ:French register of ALS in Limousin (FRALim)	2000.1から2013.7までFRALimレジストリに登録された322名の患者。 HDDでは、G12.2コードのある451名の患者(うち290名が真のALS)。 HIDでは、G12コードのある184名の患者(うち142名が真のALS)。	SensitivityとPPV	ALSでは、medico-administrative data 単独では不十分である		

17	2013	Aboa-Eboule C, Mengue D, Benzenine E, Hommel M, Giroud M, Bejot Y, et al. How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke registry as control. J Neurol 2013 Feb;260(2):605-613.	Stroke	French Hospital Discharge Database (FHDDB),	フランス	2004-2008	1	わからない（その病院がDijon地域を代表するかによる）	ICD-10 codes: I61, I63, I64, and G46のいずれかが第一診断として入っている	probabilisticrecord linkage (last names, first names, and dates of birth)	Dijon Stroke Registry	DijonHospitalのDischarge DBにありかつDijon市在住の903例とRegistryの811例。（ダブりもあるため合計1089例）	sensitivity/PPV	PPVは時間とともに良くなったが、sensitivityは変わらなかった。		
18	2016	Youngson E, Welsh RC, Kaul P, McAlister F, Quan H, Bakal J. Defining and validating comorbidities and procedures in ICD-10 health data in ST-elevation myocardial infarction patients. Medicine (Baltimore) 2016 Aug;95(32):e4554.	ST上昇型心筋梗塞（STEMI)の合併症・procedure	Discharge Abstract Database (DAD)と The National Ambulatory Care Reporting System (NACRS)の組み合わせ	カナダ Alberta州	2011	5病院 (Edmonton, Alberta, Canada)	yes	STEMIによる入院：I21.0*, I21.1*, I21.2*, or I21.3* が"most responsible diagnosis field"にあるほか、3つのprocedures(PCI, CABG, angiography)のCanadian Classification of Health Interventions (CCI)による定義。 7つの合併症（previous agnina, previous MI, hypercholesterolemia, HTN, DM, previous HF, AF)のICD10による定義。 7つの合併症については①index hospitalization DAD record② Index hospitalization or any DAD/NACRS records in the previous 1 year③the index hospitalization or any DAD/NACRS records in the previous 2 years ④at least 2 ccurrences in any DAD/NACRS records in the previous 2 yearsにある、という4つの定義。	unique patient identifiersによるVHRとDAD（入院）およびNACRS（外来）とのリンケージ	a comprehensive registry of STEMI (The Vital Heart Response (VHR) registry)	3236 patients	STEMIについてはPPV STEMI患者のProcedures、合併症については感度、特異度、PPV、NPV（κも求めている）	ICD-10 and CCI codes により、入院STEMI 患者を高いsensitivity, accuracyで検出できる。 合併症は、過去2年における2入院／救急外来で定義すると高いPPVが得られる		
19	2006	Abraham NS, Cohen DC, Rivers B, Richardson P. Validation of administrative data used for the diagnosis of upper gastrointestinal events following nonsteroidal anti-inflammatory drug prescription. Aliment Pharmacol Ther 2006 Jul 15;24(2):299-306.	NDSAIDs性上部消化管イベント Gastric/Duodenal/Pepetic/Gastrojejunal ulcer, GI haemorrhage	National administrative data VA Pharmacy Benefit Management (VA-PBM) Patient Treatment File (PTF) Out-patient Clinic File (OPC)および Beneficiary Identification and Records Locator Subsys- tem Death File databases	アメリカ テキサス州	1999～2003	1病院	yes 1病院であるが、28州をカバーするため。	NSAIDの処方の記録とTable 1に記載されたICD-9 CM code（アウトカムは「NSAID処方後の消化管イベント」）	Standard algorithmsを使用したと記載されているが、詳細は不明であった。	Trained research assistantsによるstandardized abstraction formを用いたA standardized search strategyによるレビュー	906例（possible case 606例およびnon-case 300例）を使用した。	PPV/NPV/Sensitivity/Specificity	上部消化器異常は、ICD-9に加え内視鏡など処置コードがあるとよい。		Possible caseおよびnon-caseのレビュー後、Sensitivity、Specificityなどをどのように求めたか記載はない。
20	2014	Husain N, Blais P, Kramer J, Kowalkowski M, Richardson P, El-Serag HB, et al. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. Aliment Pharmacol Ther 2014 Oct;40(8):949-954.	非アルコール性脂肪肝疾患	退役軍人データベース	アメリカ	2001-2011	1	Yes 全米の退役軍人500万人のDB	アルゴリズムで判定（ゴールドスタンダードのアルゴリズムとすこし異なる）	Corporate Data Warehouse (CDW)のデータプラットフォームでリンケージされている	従来のチャートレビュー（ALT値と既往、メタボかどうか）	Original のアウトカム定義を満たす150例。その後、新たに600例（アウトカム定義を満たす450例と満たさない150例）。	感度、特異度、陽性的中度、陰性的中度	当初の定義の感度(78.4%)、特異度(74.5%)、PPV(64.1%)、NPV(85.6%)は「2年以内に6カ月以上の間隔で最低2回のALT上昇」の条件を付加することにより、特異度は92.4%PPVは80.8%に上昇したが、感度は55.0%にNPVは78.0%に低下した。	ALT上昇しないNAFLDもある	直接のNAFLD病名ではなく、肝機能と除外病名からNAFLDをアウトカムとして、それをchart reviewと比較したややめずらしい研究 さらに研究をしながら、アルゴリズムをrefineしているという意味でもめずらしい。 "weighted specificity, sensitivity, PPV and NPV"の語は見られるが、指標の計算方法に関する記述を見る限り、定義を満たす450例と満たさない150例からなる600例を"population"として感度、特異度を求めている。
21	2013	Singh JA. Veterans Affairs databases are accurate for gout-related health care utilization: a validation study. Arthritis Res Ther 2013;15(6):R224.	痛風に関連したhealth careの利用	Veterans Affairs (VA) databases	アメリカ	2005年10月-2006年9月	1(Birmingham VA hospital)	no	痛風のICD-9-CM code for gout (274.x or 274.xx)が VA inpatient or outpatient databases にあること	1病院の研究なのでリンケージは必要なかったと思われる	Abstractorが盲検下に VA electronic health recordsのチャートレビューを行った（gout-related terms (gouty arthritis, tophaceous gout, tophus/tophi, acute gout, chronic gout, podagra, urate stones, urate or uric acid crystalsなど）があるかどうか）	母数は記載なく、そこからランダムサンプリングされた108人と書いてあるが、詳細の記載がなく、またアウトカムの割合が2割含まれているため単純なランダムサンプルではないと思われる	sensitivity/specificity/PPV/NPV			Cohort selectionのセクションの「A data programmer created a dataset of two samples that were mixed randomly」がおそらくサンプル作成を記述しているが、内容不明。最大（all visits）108名に真のgoutが23名（TP19+FN4)約2割含まれており、単なるランダムサンプルではないと思われる。
22	2014	Valkhoff VE, Coloma PM, Masclee GM, Gini R, Innocenti F, Lapi F, et al. Validation study in four health-care databases: upper gastrointestinal bleeding misclassification affects precision but not magnitude of drug-related upper gastrointestinal bleeding risk. J Clin Epidemiol 2014 Aug;67(8):921-931.	上部消化管出血	EHR Database（①②）とAdministrative Database（③④）: Europian Commission-funded exploring and understanding adverse drug reactions(EU-ADR) ネットワークの4つのElectronic health-care records データベース。①オランダのIntegrated Primary Care Information (IPCI), ②イタリアのHealth Search/CSD Patient Database (HSD), ③イタリアの ARS-Tuscany ④デンマークのAarhus University Hospital Database	研究で用いられた4つのデータベースは1996-2010年の800万人の欧州の住民をカバーする。				ICD9（イタリアのHSDとARS）、ICD10（デンマークAarhus）、International Classification of Primary Care.(ICPC)（オランダIPCI）の上部消化管出血のコード。オランダIPCIについては，追加的にフリーテキストによる定義。さらに薬による出血の可能性のない食道静脈瘤破裂などを除いたnarrow algorhithmとこれらを含めたbroad algorhithmを設定。	特に記述なし	訓練された"assessors"が入力した標準化された電子的質問票に入力された内容を判断。①単一の評価者による判断か，②判断自動的に生成されるアルゴリズムに基く判断。	各データベースから指定したコードをもつ者からの200人をサンプリング。ICPIについてはフリーテキストで定義されたアウトカムをもつ者200名を追加でサンプル。	PPV			
23	2016	Gini R, Schuemie MJ, Mazzaglia G, Lapi F, Francesconi P, Pasqua A, et al. Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from Italian General Practitioners' electronic medical records: a validation study. BMJ Open 2016 Dec 9;6(12):e012413-2016-012413.	高血圧、2型糖尿病、虚血性心疾患、心不全	Health Search/CSD Longitudinal Patient Database (HSD).	イタリア	2013	イタリア全土の900のGPsのうち、12名のボランティア	12名の属性等は不明だが、Population basedとみなしてよいと考える。	4つの疾患について、重症度別に3～4個の定義を作成 （組み入れ基準、除外基準を、それぞれICD9CMの組み合わせで定義し、必要に応じて文字情報や検査結果も併せて定義した）	Linkageはしていない。HSDの症例をリンクさせたのではなく、HSDで使うアルゴリズムを用いて、バリデーション研究で新たに12GPのMedical Recordから"plug in"により特定した患者を評価。	チャートレビュー 参加した12名のGPのシステムにおいて"PlugIn"により25例の患者がランダムされ、同時に疾患とその重症度の定義とも表示される。当該GPが（自分の患者の疾患名を改めて）真か偽か判断、また重症度を判定（アルゴリズムによる重症度は示されない"blinded to the level of severity"） 判断にあたって、GPはmedical recordにアクセスすることができたので、medical chart reviewに区分できる。	参加したGPのMedical Recordからランダム抽出したDM、IHD、HTN各300例(25X12)、HF243例	PPV	2型糖尿病、高血圧、IHDについては、本方法でとらえることができた。心不全については、アルゴリズムの改良が必要である。		
24	2013	Nosyk B, Colley G, Yip B, Chan K, Heath K, Lima VD, et al. Application and validation of case-finding algorithms for identifying individuals with human immunodeficiency virus from administrative data in British Columbia, Canada. PLoS One 2013;8(1):e54416.	HIV陽性者	BC Ministry of Health (BCMoH)のMedical Service Plan(MSP)とDischarge Abstract Database(DAD)	カナダ（ブリティッシュコロンビア）	1995-2010	記載なし	yes	MoHを構成するMSP（外来、Claims）とDAD（入院）の情報を組み合わせる4つのHIVに関連する外来受診のphysician claims and/or入院の回数の組み合わせ。①3 physician claims ②3 physician claims or 1 admission ③2 physician claims or 1 admission ④ 1 physician claims or 1 admission。このうち最終的に②が採用された。	provincial health number (PHN)	検査または治療薬の使用：BC Centre for Disease Control (CDC)のHIV-positive testの記録またはBC Centre for Excellence in HIV/AIDS (CfE)のplasma Viral Loard(pVL)または HIV-related medicationの記録。 この基準を満たすConfirmed cases11500人のうち、CDC、CfE、MoHのいずれにも記録がある者を5,039人を"Gold standard cases"と定義している（その5039人の中でsensitivityを計算している）。	18 month以上で3つのタイプのDB（CfE、CDC、MoH）のいずれかにHIV関連の記録がある25,673人	Sensitivity			Sensitivityが評価されている。Gold standard casesではない（が、gold standardの基準を満たす）"confirmed cases"6461例はそれ以上は調べていない(MoHのデータにない[リンクできない]non-nominal reportを選んだ人たち？）。研究は、最新のHAART治療を受けていないHIVの患者が多いことの克服のために実施されている。Unconfirmed cases9454人の中にも真のケースが存在する、と考えており、HIVに対する積極的治療が提供されているCfEやCDCに見出されないUnconfirmedの9454人から真のケースを効率良く見出すためにはどうすべきか？が課題。Unconfirmedの9454人にアルゴリズムを当てはめたときに選ばれた集団がHIVの患者の特徴（男性に多い、より多く死亡、health serviceをより利用）を満たすかをテストしている。
25	2014	Williamson T, Green ME, Birtwhistle R, Khan S, Garies S, Wong ST, et al. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. Ann Fam Med 2014 Jul;12(4):367-372.	慢性疾患	Canadian Primary Care Sentinel Surveillance Network（CPCSSN）を構成する10のネットワークのうち、British Columbia、Alberta（2つ）、Ontario、Nova Scotia、Newfoundlandの電子医療記録6つ	カナダ	2012年6月30日に10のprimary care research networks（PCRNs）の6つからデータ抽出。	33	yes	CPCSSNの8つの疾患定義（I糖尿病、高血圧、COPD,うつ、骨関節炎、てんかん、パーキンソン病、認知症）ICD9のほか、EHRの様々な箇所からのテキスト、薬、検査値によって定義	なし	リサーチアシスタントによるレビュー。不明点については研究に参加した疫学者または医師がレビューした。	各データベースからランダムに抽出された1920例 60歳以上から90%をランダムサンプル。Epilepsyとparkinsonismについてはcase positive各25例/PCRNをサンプル。	感度、特異度、陽性的中度、陰性的中度	CPCSSNデータは公衆衛生上のサーベイランス、primary care、ヘルスサービス、行政活用で有用。	特になし	Epilepsy、Parkinsonismについてはランダムサンプル+ case positiveの例を抽出しているが、どのように感度、特異度を求めたかは十分明らかではない。
26	2007	Baghestan E, Bordahl PE, Rasmussen SA, Sande AK, Lyslo I, Solvang I. A validation of the diagnosis of obstetric sphincter tears in two Norwegian databases, the Medical Birth Registry and the Patient Administration System. Acta Obstet Gynecol Scand 2007;86(2):205-209.	会陰裂傷	Medical Birth Registry of Norway (MBRN)と Patient Administration System (PAS)	ノルウェーの1病院（Haukeland University Hospital ）	1990-1992年と2000-2002年			MBRNはmidwifeまたは医者が入力したフリーテキストまたはPerineal tear有り無しの記録 PASでは、ICD-9/10とProcedure codeのコンビネーション	特に記述なし（1病院）	チャートレビュー 25761全例について、midwifeのbirth logからケース（候補）を特定し、surgical repairのprocedure recordを含む medical recordで特定した。	この病院の全出産例、MBRNは12380例（2000-2002年）、PASは13381例（1990-1992年）と12380例（2000-2002年）	Sensitivity/Specificity/PPV/NPV	2つのデータベースを比べてMBRNのほうがいいと言っている。	1990-1992年のデータについてはprocedure recordがない例が多く、ゴールドスタンダードの正確性に疑問があると書いてある	25000例以上のchart reviewから774例（1990-1992）+813例（2000-2002）のケースを特定。
27	2013	Ramalle-Gomara E, Ruiz E, Serrano M, Bartulos M, Gonzalez MA, Matute B. Validity of discharge diagnoses in the surveillance of stroke. Neuroepidemiology 2013;41(3-4):185-188.	脳卒中とTIA	The Minimum Basic Data Set (MBDS)	全てのスペイン住民の退院データベース。The Minimum Basic Data Set (MBDS)				ICD9-CMの430-438の全てまたは一部を組み合わせる4つのアルゴリズム。	特に記述なし	2人の神経科医によるカルテレビュー	MBDSからICD-9コードでランダムに"chronic cerebrovascular disease diagnoses"の300人と"mimic diseases"の100人、合計400人。研究者が恣意的に集めた400人。	Sensitivity/Specificity/PPV/NPV 二人のneurologistsの一致については"excellent"との表現があるのみ。	ICD9-CM code 430-438。先行研究と比べてもこのコードが最も感度も特異度もPPVも高く，お勧めだと述べている。		二種類の特定のICD-codeをもつ者が3：1で構成される400人（"chronic cerebrovascular disease diagnoses"の300人と"mimic diseases"の100人、合計400人）の中で感度、特異度が計算されており、これらの指標があてはまるのは同様にサンプリングされた「集団」においてのみ。
28	2009	Hall GC. Validation of death and suicide recording on the THIN UK primary care database. Pharmacoepidemiol Drug Saf 2009 Feb;18(2):120-131.	Death（死亡の事実および死亡日） Suicide Cause of death	THIN UK primary database	イギリス	2002年1月から2004年12月		Yes 英国GPデータ	Death：Index dateから365日以内のAdditional Health Data(AHD), medical file, administration fileにおける死亡の記載。死亡日についてはAHDまたはmedical fileにおける死亡日。 Suicide：Read codeまたはComments sectionにおける記載。 Cause of death (COD)：electronic recordにおける記載	必要があれば"third party"を介して追加データを得た	Death ：deathの記録があり、かつcause of deathが別に記録されているか、死亡診断書などのexternal documentがある 死亡日：external documentに記載された日 Suicideについては明確に記載されていないが、external documentによると思われる。 Cause of Death：死亡診断書	BMIの研究で対象となった142,082人。1年以内の死亡の可能性のある1621人（365日以内の死亡1399、366-485日の死亡<実は365日以内の死亡が含まれうる>222人の合計）。 この中からさらにSuicide7人が見いだされ、死亡例から40人が死亡診断書のCause of deathとの比較のためにランダムサンプルされた。	死亡：感度、PPV 死亡日、suicide、Cause of death：PPV	特に記述なし	BMI情報が取得でき、かつ、妊娠、emotional eventsやがんのない集団に限定されていることが限界の一つ。	対象集団のうち、アウトカム定義に含まれる「Index dateから365日内の死亡」がDBから見出されるものと（365日以内の死亡の可能性がある）「Index dateから366-485日内の死亡」をcause of deathが別途記載されているか？External documentで見いだされるか？で評価したもの。二つのアプローチ（365日以内、366-485日）で見いだされた死亡を「全ての真の死亡」とみなしている。
29	2016	Cea Soriano L, Soriano-Gabarro M, Garcia Rodriguez LA. Validity and completeness of colorectal cancer diagnoses in a primary care database in the United Kingdom. Pharmacoepidemiol Drug Saf 2016 Apr;25(4):385-391.	大腸がん	イギリスのプライマリケア医によって登録される匿名のEMRのデータベース（THIN:The Health Improvement Network）	イギリス	2000-2011	記載なし	yes	Readコード（イギリスのプライマリケアで用いられる診断コード）+manual review(フリーテキストで大腸がんが否定されない）	HESとのリンケージ方法の記載なし	①PCP質問票による医師調査（検査有無、癌の部位、診断票や紹介状の確認） ②HESデータベース(ICD-10コード）	2000年1月から2011年12月に初めて大腸がんのReadコードが付与された40から89歳の3805例のうち100例をランダムサンプリング	陽性的中度、感度（直接には偽陰性の割合が示されている）。	癌のステージや部位に関する情報は、臨床アウトカムのリスク評価や、交絡調整に使えるのでは。	特になし	複雑な研究：①ReadコードにCRCがある3805例のうち、manual review（gold standard）でCRCとされたのが3033例。PPVは79.7%（3033/3805）。 ②質問票送付対象100例のうちmanual reviewでCRCの80例のうち回答のあった87例（回答でケース71例、非ケース16例）に含まれる71例は全て真のケース。（Manual Review後の）THINのCRCのPPV=100%（71/71）。 ③Manual ReviewでCRCの568例のうち、HESでもCRCは509例。Manual ReviewでCRCの47例は入院しておらず、これを真とするとPPV=(509+47)/568＝97.9%。Manual ReviewではCRCではなく、HESにのみCRCのレコードのあったのが36例。これに加えHESでCRCの509例と入院しなかった47例の合計592名が「真のCRC」。このうち（Manual review後の）THINのCRCとして把握できなかったのが36例。FN=36/592=6.1%。すなわち、（Manual Review後の）THINのCRCの感度=100-6.1=93.1%.
30	2015	Mamtani R, Haynes K, Boursi B, Scott FI, Goldberg DS, Keefe SM, et al. Validation of a coding algorithm to identify bladder cancer and distinguish stage in an electronic medical records database. Cancer Epidemiol Biomarkers Prev 2015 Jan;24(1):303-307.	膀胱がん	イギリスのプライマリケア医によって登録される匿名のEMRのデータベース（THIN）	イギリス	2001以降	記載なし	yes	すべての膀胱がんの疾患コード、あるいは、筋浸潤膀胱がんの疾患コード+膀胱摘除のコード（疾患コードはReadコード）		医師記入の質問票、病理レポート、紹介状	2001年以降の最初の膀胱がんの診断前に6ヵ月以上のフォローアップ期間がある21歳以上194例 質問紙を送ったのは210例	陽性的中度	ICD-9コードのアメリカのデータセットにも一般化できるかどうかは要検討	特になし	GPに対する質問票を210例分送付。194(92%)例分の回答。
31	2014	Ogdie A, Alehashemi S, Love TJ, Jiang Y, Haynes K, Hennessy S, et al. Validity of psoriatic arthritis and capture of disease modifying antirheumatic drugs in the health improvement network. Pharmacoepidemiol Drug Saf 2014 Sep;23(9):918-922.	乾癬性関節炎	THIN	イギリス	2012年	記載なし	yes	乾癬性関節炎の疾患コード（Read coad)を有する。6個のアルゴリズムを比較	なし	医師の判断（調査票を医師へ送付）	疾患コードをもった18-89歳からランダムに抽出された100例	陽性的中度 もっとも広いアルゴリズム（該当診断コードを1回有する）以外については、「感度」「特異度」の用語で数値が示されている（「⑨問題点」参照）。	DMARDsだけでは十分に乾癬性関節炎の患者を拾うことができない	調査に協力した医療機関のデータのみなので選択バイアスの可能性がある	「感度（sensitivity）」「特異度（specificity）」については、調査対象の87例（最低1回の診断コード」のうち、真の74例をpopulationにおける真の全ケース（のランダムサンプル）、否定された13例をpopulationにおける全ての非ケース（のランダムサンプル）として計算したもの。著者らも"overall sensitivity and specificity"ではない、と断ってはいるが、適切な用語法とはいえない。

【付録3文献リスト】

32	2011	Denburg MR, Haynes K, Shults J, Lewis JD, Leonard MB. Validation of The Health Improvement Network (THIN) database for epidemiologic studies of chronic kidney disease. Pharmacoepidemiol Drug Saf 2011 Nov;20(11):1138-1149.	慢性腎不全(Chronic Kidney Disease)	The Health Improvement Network (THIN)	イギリス	2008年10月のデータセット(2002年以降2008年までのデータをクロスセクショナルに検討)		yes	事前に定義した中高度のCKDに対する80個のRead code (Table 2)	THIN (GPRD)は、診療コードと検査値はもともと両方入っているのでこのためにリンクさせたわけではない	血清クレアチニンから計算された腎機能(eGFR): 90日以上間隔をおいた2回以上でeGFR <60 ml/min/1.73m2	2002年より前に死亡・THINのプラクティスを離れた者を除く1,632,819人	PPV/NPV/Sensitivity/Specificity	PPVの95%信頼区間の下限80%以上のRead Codeが45あったとしている。ただし、筆者らはlaboratory-based definitionを基本とし、これを補完する上でRead codeを用いることを提唱しており、そのような使用を前提に「PPVの下限80%」を基準として示している。	2002年以降のクレアチニンデータしか使用していないため、それ以前にCKDと診断されていた人のコードが入る誤分類の危険性がある。	病名は感度が低いので、最初からクレアチニンを使って慢性腎不全や急性腎不全を定義するべき
33	2009	Lo Re V,3rd, Haynes K, Forde KA, Localio AR, Schinnar R, Lewis JD. Validity of The Health Improvement Network (THIN) for epidemiologic studies of hepatitis C virus infection. Pharmacoepidemiol Drug Saf 2009 Sep;18(9):807-814.	C型肝炎	The Health Improvement Network (THIN)	イギリス	2000～2007	380のGPからなるデータベース	yes	①Table 1にある10のHCV感染症のREADコード②診断コードはTable 2にある25の"viral hepatitis NOS"だが、フリーテキストに"hepatitis C"または"C"の記載がある。	リンケージ機能を有したTHIN's Additional Information Servicesにより実施された。	GPへの質問票の郵送により得た①HCV抗体, ②positive recombinant immunoblot assay (RIBA), and/or ③定量的なHCV RNA.	HCVのコードをもつものから75例、nonspecificのコードをもつものから75例、合計150例をランダムに選択	PPV		HCVの診断コードのみの使用であったため、THINに登録されている真のC型肝炎患者割合を決定できず、Sencitivity、Specificityあるいはκについても求められなかった。	GPへの質問紙を用いてゴールドスタンダード情報を取得しているが、取得したのは検査結果であり、GPの判断の余地は少ない。
34	2004	Fombonne E, Heavey L, Smeeth L, Rodrigues LC, Cook C, Smith PG, et al. Validation of the diagnosis of autism in general practitioner records. BMC Public Health 2004 Mar 3;4:5-2458-4-5.	自閉症(広汎性発達障害、pervasive developmental disorders)	General Practitioner Research Database (GPRD)	イギリス	1973～1997		yes	診断名	記載なし	GPから提供を受けたpatient recordの記録、hospital clinic letters, specialist reportなどの関連する臨床的報告のレビュー。GPの記録、consultant, speech therapy 、educational psychologists の報告。computer-based algorithmとraterのjudgementの2通りで判定。前者のPPVは80.6%、後者は92.5%		PPV 評価者間の一致についてはκ			
35	2009	Engeland A, Bjorge T, Daltveit AK, Vollset SE, Furu K. Validation of disease registration in pregnant women in the Medical Birth Registry of Norway. Acta Obstet Gynecol Scand 2009;88(10):1083-1089.	妊娠に合併した糖尿病、てんかん、喘息	Medical Birth Registry of Norway	全Norwayをカバーする妊娠登録	2004年3月から2006年12月		Yes 全ノルウェー	妊娠中の糖尿病・喘息・てんかん(Medical Birth Registry of Norway (MBRN)のICD-10コード	ノルウェーのNational identification number(11桁)	糖尿病・喘息・てんかんのための処方(Norwegian Prescription Database, NorPD)	Norwayの妊娠レジストリ2004/3から2006/12(before January 1, 2017)における108,489例(の糖尿病926例、喘息2098例、epilepsy314例)	Sensitivity/Specificity/PPV/NPV	1型糖尿病(感度90%)以外のany typeの糖尿病(感度72%)、epilepsy(感度74%)、喘息(感度51%)については"should be dealt with carefully"		ゴールドスタンダードが処方にもとづいていること
36	2014	Klungsoyr K, Harmon QE, Skard LB, Simonsen I, Austvoll ET, Alsaker ER, et al. Validity of pre-eclampsia registration in the medical birth registry of norway for women participating in the norwegian mother and child cohort study, 1999-2010. Paediatr Perinat Epidemiol 2014 Sep;28(5):362-371.	妊娠高血圧腎症(preeclampsia)	ノルウェー出生レジストリ(MBRN)	ノルウェー	1999-2010	47	Yes	妊娠高血圧腎症の記録 いくつかのTickBoxがある人ICD10で記録	調査への参加を同意した病院から出生前チャートおよび退院フォームを入手	出生前チャートに高血圧と尿たんぱくの記録、あるいは退院フォームにICD-10コードの付与	妊娠高血圧腎症の記録がある女性4081例と、記録をもたないランダムサンプル2000例	positive predictive value (PPV), sensitivity and specificity	1999-2010年のMBRNの"preeclampsia"のPPV と特異度は十分高いが、感度は低い	特になし	アウトカム定義を満たす全例と満たさない例の2.4%のランダムサンプルをChart Reviewで評価し、感度、特異度については、適切な重みをつけて補正、95%信頼区間についてはbootstrap法を適切に用いて計算している。
37	2015 (online), 2017 (journal published)	Bowker SL, Savu A, Lam NK, Johnson JA, Kaul P. Validation of administrative data case definitions for gestational diabetes mellitus. Diabet Med 2017 Jan;34(1):51-55.	妊娠糖尿病	Discharge Abstract Database、Physician claims、Alberta Health Care Insurance Population registry の組み合わせ	カナダ Alberta州	1999年4月から2010年3月			①糖尿病のためのNDSS algorithm(糖尿病のための1回の入院または2年以内の2回のphysician claims) を満たし、糖尿病の発生の180日前から120後までの間に産科に関連するclaim(obstetric claim)が見いだされる(Fig 1の"NDSS")。 ②妊娠糖尿病に特化したICD codeが出産に関連する入院の診断フィールドに見出される(Fig 1の"ICD")。	カナダのmother's unique personal health number	周産期レジストリ(APHP)の妊娠糖尿病のコード。	アルバータ州で1999年4月から2010年3月までの411390件の出産(273152人)。該当症例全例。	PPV/sensitivity/specificity	NDSSアルゴリズムを用いた妊娠糖尿病(アウトカム定義①)は妊娠糖尿病を過小評価するので、ICDを用いた定義(アウトカム定義②)の方がいい。過去の妊娠糖尿病の有病率の算定データを見る際にはこの点を考慮すべき。		
38	2011	Furlan JC, Fehlings MG. The National Trauma Registry as a Canadian spine trauma database: a validation study using an institutional clinical database. Neuroepidemiology 2011;37(2):96-101.	Spinal Trauma/Spinal cord injury	Canadian National Trauma Registry (NTR)		2003年5月から2007年7月	カナダの1病院(Tront Western Hospital)	Canadian National Trauma Registry (NTR)。Registryはpopulation-basedだたバリデーション研究は1病院で実施されておりpopulation-basedではない。	Spinal Trauma/Spinal Cord Injury(SCI)に区分されうる10のICD-10codes(Table 5と6にその一覧)	記載はない。一病院における研究であり、Linkageは問題にならない。	Chart ReviewによりSpinal Trauma (with/without coud innjury)であるか、Spinal cord Injuryであるかを決定。	Spinal TraumaのためにTronto Western Hospitalに入院した92人(この92人において感度、PPVなどを計算)	Sensitivity/Specificity/PPV/NPV		Table 3にInjuryの程度により、Traumaに関してとcord injuryに関して、True/falseの結果が異なることを示す。Neurologicalには正常のAmerican Spinal Injury Eの時にcord injuryについてはfalse resultsが多い。 Table 5(Spinal Trauma(with or without cord injury))とTable 6(spinal cord injury)に関するFalse Positive、False Negativeの一覧。	Spinal TraumaのためにTronto Western Hospitalに入院した92人が当該病院における全true casesとりあげているICD-10コードがYesの全例を含んでいるのなら感度、PPVは適切。特異度、NPVは「当該病院でspinal traumaで入院した患者」というpopulationという特殊な集団においての値であり、そのような特殊な集団以外の研究では不適切。
39	2012	Bahraini NH, Gutierrez PM, Harwood JE, Huggins JA, Hedegaard H, Chase M, et al. The Colorado Violent Death Reporting System (COVDRS): validity and utility of the Veteran status variable. Public Health Rep 2012 May-Jun;127(3):304-309.	Veteran status (退役軍人かどうか)	National Violent Death Reporting System (NVDRS) の利用を念頭にコロラドのColorado Violent Death Reporting System (COVDRS): NVDRSは死亡診断書、警察からの報告、監察医、検視官、crime laboratoriesからの情報などをmergeしたもの。	アメリカ(コロラド州)	2004-2008	記載なし	yes	Veteran status information on the death certificate	SSNとfirst nameとlast name	Veterans Benefits Administrationへの登録(VA Beneficiary Identification Record Locator Subsystem (BIRLS))	データベースにいる全ての18歳以上の2004－2008年の自殺者3820人	sensitivity/specificity/PPV/NPV		特定の部署の退役軍人はVeterans Benefits Administrationへの登録がされやすいかもしれない	Gold standardに"No"の情報がないことがユニーク。False nagativeについては「Gold standardでは"Yes"だがdeath certificateではverans statusは"No"」,False positiveは「death certificateではverans statusは"Yes"だがGold standardに「ない」。)
40	2012	Rozendaal AM, Luijsterburg AJ, Mohangoo AD, Ongkosuwito EM, de Vries E, Vermeij-Keers C. Validation of the Dutch registry of common oral clefts: quality of recording specific oral cleft features. Cleft Palate Craniofac J 2012 Sep;49(5):609-617.	口唇口蓋裂	口蓋裂レジストリ	オランダ全体で始められた口蓋裂レジストリー(15の口蓋裂チームが入力)	1997-2003年		15の口蓋裂チームによって特定されるケース集団がpopulationのケースをどの程度把握しているか不明。	レジストリの口唇口蓋裂の診断	明確な記載なし。レジストリはanonymousであるが、写真、カルテ情報などをチームのメンバーから得ており、何らかの方法でリンケージが行われたと考えられる。	Chart review(写真、カルテ情報などによる判断)。著者の一人が実施。結果をreregisterした。	2553例の症例からランダムサンプルした250例のうち、写真、カルテ情報などが得られた237例	PPV/NPV/Sensitivity/Specificityの値が示されている。			PPVはこの研究で正しく求められる。感度、特異度、NPVは、レジストリからランダムサンプルで得られた237人を"population"として求めたもの。より広い集団(例:オランダ全体で実施される研究)ではPPV以外は適切な指標ではない。
41	2016	Cundall-Curry DJ, Lawrence JE, Fountain DM, Gooding CR. Data errors in the National Hip Fracture Database: a local validation study. Bone Joint J 2016 Oct;98-B(10):1406-1409.	Hip fractureに対する手術の術式と死亡	イギリス英国内科医師会の運営する National Hip Fracture Database (NHFD)	英国	2009年7月から2014年6月			Hip fractureに対する"no operation"を含む13の術式と死亡	記載なし	病院レセプト情報、EMR, Picture Archiving and Communication System (PACS), the NHS Spine and patient notes.	2009.7から2014.6までにhip fractureとなった 2036例	sensitivity, specificity, overall accuracy	なし		感度、特異度などは、NHFDの2036例についてゴールドスタンダードで正しい情報を得て、2036例を"population"として求めたもの。
42	2012	Vanderloo SE, Johnson JA, Reimer K, McCrea P, Nuernberger K, Krueger H, et al. Validation of classification algorithms for childhood diabetes identified from administrative data. Pediatr Diabetes 2012 May;13(3):229-234.	小児糖尿病(1型、2型)	British Colombiaのadministrative data (外来のfee-for-service claimsと入院のhospitalization discharge codesからなる)	British Colombiaのadministrative data(実質、1つの病院でバリデーションをしている)	1996年4月から2007年3月			2001-2年の間に、2つのbilling claim(ICD-9 250)または1回の入院(ICD-10 E10-14)(National Diabetes Surveillance System (NDSS)で使われている糖尿病を特定するアルゴリズム)、に加えて、年齢、Indian status、PharmaNet provincial prescription drug databaseの処方パターンによるType 1とType 2糖尿病	BC Ministry of Health(MOH) のスタッフが実施したとの記述があり、Personal health numbers (gold standard(BC Children's Hospitalの記録)との比較のためのLinkageではなく、PharmaNetとの記録とのLinkage?	BC Children's Hospital Endocrinology and Diabetes Unitで診断された糖尿病(登録データ)	administrative dataから特定された糖尿病の薬を受けている子供4060人 4060人のうちBC Children's Hospital (BCCH) Endocrinology and Diabetes Unit (EDU)で診療を受けた1323名を評価。	Sensitivity/Specificity/PPV			National Diabetes Surveillance System (NDSS)における糖尿病の定義のルール(2年以内に1回以上の糖尿病による入院、2回以上の外来における糖尿病コード)をみたし、(さらに糖尿病疑い例をのぞいた)"population"におけるtype 1とtype 2を区別するアルゴリズムのSpecificity、Sensitivity,PPV、NPVを評価する研究。
43		Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care 2002 Mar;25(3):512-516.	糖尿病	カナダOntario州のphysician service claims (PSCs)とhospital discharge abstracts (HDAs)をリンクしたものにおける入院と外来の診断コードの組み合わせ。	カナダ オンタリオ州: Population basedの糖尿病の診断codeのバリデーション研究。			おおむねYes。研究に協力した57人の医師により3337chartsを抽出した。	①2年以内の1回以上の糖尿病のICD8/9をもつ外来または入院。 ②2年以内の糖尿病のICD8/9をもつ1回以上の入院または2回以上の外来診療	"Unique identifier"。	①一般住民対象のサーベイ(National Population Health Survey (NPHS)) ②オンタリオ糖尿病 の研究で協力した51人のphysicianのphysician office charts。Standard data collection instrumentを用いたとあるがそれ以上の詳細はない。	520人の医師のうち、57人の医師が3337人のデータをチャートレビューした。	感度、特異度, PPV	カナダオンタリオ糖尿病 "We sought an algorithm that maximized sensitivity while providing at least 80% of PPV."との記述がある	カナダオンタリオにおける糖尿病の診断の評価。Gold standardとして患者の自己報告のDBのthe National Population Health Survey (NPHS)を使ったが、正確に評価できなかったために、医師のカルテを抽出した、などのプロセスが記載されており、探索的な部分もある。	
44	2010	Tu K, Mitiku T, Lee DS, Guo H, Tu JV. Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the Electronic Medical Record Administrative data Linked Database (EMRALD). Can J Cardiol 2010 Aug-Sep;26(7):e225-8.	虚血性心疾患	Ontario Health Insurance Plan Physician Claims DatabaseおよびCanadian Institute for Health Information Discharge Abstract Database	カナダ オンタリオ州	2007/06～2007/12	Convenience sampleによる17人の医師	no	Ontario Health Insurance Plan Physician Claims DatabaseおよびCanadian Institute for Health Information Discharge Abstract Databaseのコードを組み合わせ、9パターンのアルゴリズムを作成した。 ICD9, 10コード、PCIを記載する。	Patient health card numbersを一意の識別子に変換し、匿名でadministrative dataにリンケージされた。	EMRALDを使ったChart review "hard"と "soft"の二種類のGS	17人の医師の20歳以上の2007年12月における19376人の患者の5%のランダムサンプル969例を使用した。	PPV/NPV/Sensitivity/Specificityおよびκ	記載なし	以前のバリデーション研究と本研究を比較することは困難である。	Gold standardは2つ(Hard IHD, Soft IHD)あるが、2つに対してTable 1で、Hard IHDに対してTable 2で検討している。 κをintraobserver & interobserver reliabilityの指標として示している。
45	2013	Tu K, Wang M, Young J, Green D, Ivers NM, Butt D, et al. Validity of administrative data for identifying patients who have had a stroke or transient ischemic attack using EMRALD as a reference standard. Can J Cardiol 2013 Nov;29(11):1388-1394.	一過性脳虚血発作か脳卒中	CIHI Discharge Abstract Database (DAD), the National Ambulatory Care Reporting System Database(NACRS), and the Same Day Surgery Database (SDS)	カナダ, オンタリオ州の83人の家庭医の患者。2011年6月～2011年11月。				ICD9かICD10のTIAかStrokeのコード, OHIPのTIAかStrokの診断コードがそれぞれのデータベースに存在しているか, OHIPに出現したコードの回数と対象となる期間の長さによる組合せ。	特に記述なし	ランダムサンプルされた5000例を5人のChart Abstractorsが①definite ②probable ③rule out される ④ stroke/TIAの言及がない、に分類。①②③は専門家が2人が最終判断。	73014人からランダムサンプリングされた5000人の成人。	Sensitivity/Specificity/PPV/NPV 5人のchart abstractorsの一致については、10%サンプルを2人がレビューし,κ係数を求めた。	特定するのに一番よいアルゴリズムは入院しているか, 2回以上の診断記録が1年以内にあること, といっている。		
46	2014	Metcalfe A, Sibbald B, Lowry RB, Tough S, Bernier FP. Validation of congenital anomaly coding in Canada's administrative databases compared with a congenital anomaly registry. Birth Defects Res A Clin Mol Teratol 2014 Feb;100(2):59-66.	先天性奇形(CA)	1) Discharge Abstract Database (DAD) 2) Ambulatory Care Classification System (ACCS) 3) Physician Claims (PC) database それぞれ単独に使用した時、および組み合わせて使用した時における指標を求めている	カナダ・カルガリー州の妊娠情報				ICD-10、外観からの判断、他	personal health number (PHN), name, date of birth and gender	レジストリ(Alberta Congenital Anomalies Surveillance System)	カナダ・アルバータ州のカルガリー(市)での 2007.11から2008.2までの妊娠とそのフォローアップで確認された5862 名の新生児・死産児	Sensitivity PPV Kappa	administrative dataに関しては、入院データは、救急外来・外来受診データよりもまだ精度よく検出できた。		
47	2011	Dart AB, Martens PJ, Sellers EA, Brownell MD, Rigatto C, Dean HJ. Validation of a pediatric diabetes case definition using administrative health data in manitoba, Canada. Diabetes Care 2011 Apr;34(4):898-903.	小児糖尿病	The Manitoba Health Services Insurance Plan (MHSIP) 入院(hospital discharge abstract),外来のclaims(physician claims), 薬の記録を含む	カナダ マニトバ州	2004 ～ 2006	記載なし	yes 小児1型糖尿病患者90%以上、多数の小児2型糖尿病患者の記録があるという記載からpopulation-basedと考えられる。	18のアルゴリズム(最終的に選ばれたのは、one or more hospitalizations or two or more outpatient claims over 2 years)	PHIN codeおよびcross-walk fileを使用した。	疾患レジストリ The Manitoba Diabetes Education Resource for Children and Adolescents (DER-CA) Registry	1000例前後の小児糖尿病患者の記録を使用した。	PPV/NPV/Sensitivity/Specificity	PPVおよびSencitivityの両方が高い指標を採用した。	Prevalenceをadministrative databaseから求めるという目的の下に実施された研究	
48	2009	Dodds L, Spencer A, Shea S, Fell D, Armson BA, Allen AC, et al. Validity of autism diagnoses using administrative health data. Chronic Dis Can 2009;29(3):102-107.	自閉症スペクトラム症	The Hospital Discharge Abstract Database, the Medical Services Insurance (MSI) Physician Billings Databaseおよび Mental Health Outpatient Information System (MHOIS) Database	カナダ ノバスコシア州	1989～2002に生まれた小児を2005まで追跡した。	記載なし	yes, Nova Scotia全てのhospital birthsから特定されており、population-based	診断コード記載欄の一番目もしくは二番目に記載されたICD-9 299もしくはICD-10 F84を用い、Hospital Data, Physician billing data, MHOISのいずれに何回見られたかの組み合わせにより7通りの定義を行った。	Encrypted health card numbersを使用した。	IWK Health Centreの自閉症チームの診断を使用した。	Study cohortの270例のうち診断が確定していない6例を除く264例。	PPV/NPV/Sensitivity/Specificity/C-statistic	Administratibe health databaseの使用は、費用効率の高い情報源である。		IWK Health Centreの自閉症チームが評価した264例を"population"として感度、特異度などが評価されている。
49	2012	Dregan A, Moller H, Murray-Thomas T, Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. Cancer Epidemiol 2012 Oct;36(5):425-429.	癌	GPRD	England	2002年1月から2008年5月	158practices "the subset of English practices"	英国のGPのデータであるという意味では"Yes"だが、感度などを求めているのは「がんの診断または警告症状をもつ"population"」において。	GPデータベースの 癌診断コード(Read code):大腸直腸、肺、胃食道、泌尿器、および死亡	NHS番号、郵便番号、性別、生年月日によりTrusted Third Party(TTP)がGPデータと癌レジストリーをリンク deterministic linkage	癌レジストリーの登録病名・死亡情報 2人の専門家が該当症例(4つがん種と死亡)をレビューした。 登録と個別症例のレビューの組み合わせ	GPデータベース(イングランド)の158GPに2001年から2007年まで継続して登録していた患者で、かつ警告症状(血尿、血痰、嚥下困難、血便)または癌の診断がついていた42,556人 "convenience sample"と記載されているが、複数の定義(病名と警告症状)を満たす人を選択しており、"all possible cases"のサンプルとも考えられる。	PPV/sensitivity/specificity ただし、Specificityは42556records with either alarm symptoms or eligible cancer codesにおける値		False PositiveとFalse Negativeの例についての詳細(Table 2)が示されている。	がん登録をgold standardにしたGPRDのがんのコードの正確性を検討した研究。単純にレジストリをゴールドスタンダードとするのではなく、さらに4万例以上という多数について個別症例をレビューしている。レビューの詳細は不明だが、レジストリにおける個別症例のレビューでは相当多数が対象になりうることを示す例。 "specificity"は42556records with either alarm symptoms or eligible cancer codesという特殊な"population"における値であり、より広い"population"における値とは異なる。
50	2014	Quint JK, Mullerova H, DiSantostefano RL, Forbes H, Eaton S, Hurst JR, et al. Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). BMJ Open 2014 Jul 23;4(7):e005540-2014-005540.	COPD	CPRD	イギリス	2004年1月1日から2012年12月31日	660 primary careからなるデータベース	yes	8つのアルゴリズムでCOPDと判定	なし	郵送法による質問票(COPDか?のGPの判断、Spirometryの結果、専門医からの紹介状送付を要請)から得た情報を2人の呼吸器専門の医師による独立した判定、意見が一致しない場合は3人目の医師も判定を行う	2004年1月から2013年12月に登録された、喫煙歴があり少なくともひとつのCOPDを示唆する疾患コードが付与された35歳以上の951例。解析に使えたのは696人分	陽性的中度	PPVはCOPD codeのみではなく、薬やspirometryを組み合わせた方が高くなるが、それほど違いはなく、簡便なCOPDのcodeだけを用いることを推奨している。		ゴールドスタンダード自体に問題(呼吸器内科医が、COPDを記録したGP自身の判断+追加資料を判断)

【付録3文献リスト】

51		Franchi C, Giussani G, Messina P, Montesano M, Romi S, Nobili A, et al. Validation of healthcare administrative data for the diagnosis of epilepsy. J Epidemiol Community Health 2013 Dec 1;67(12):1019-1024.	epilepsy	Drug Administrative Database (DAD) 薬、discharge diagnosis、EEGなどの検査、医療費免除(disease-specific exemption)のコードなどがリンク可能な状態で管理されている。	イタリアのLombardy Regionのレッコ県(district of Lecco)における「全epilepsy」	2000年から2008年		Yes、イタリアの一地域全体	①Epilepsy のICD9code,②Epilepsyに関する医療費免除コード(excemption code)③1回以上のEEG④antiepilepticの処方、それぞれおよび組み合わせ。	"linked by a unique identification code"	11人のGPのelectric records(Italy, epilepsy)の記録。 「in Italy, epilepsy is confirmed only after a neurology (or child neurology) consultation」とGSとしてよい理由を説明。	11GPsのadministrative data15,728例 から71例のepilepsyのコードをもつ患者を検討	PPV/NPV/Sensitivity/Specificity 地域の全ケースを11GPで特定可能			イタリアのLombardyにおけるepilepsyの研究。11人のGPの電子的なレコードにおけるepilepsyが地域の全epilepsyとされている。
52		Lo Re V,3rd, Haynes K, Goldberg D, Forde KA, Carbonari DM, Leidl KB, et al. Validity of diagnostic codes to identify cases of severe acute liver injury in the US Food and Drug Administration's Mini-Sentinel Distributed Database. Pharmacoepidemiol Drug Saf 2013	検索式によって見出したのではなく事後に追加した論文(US HMOの論文を含めたかった)	Mini-sentinel Distributed Database (MSDD)	Mini-sentinelの5つのデータパートナーから得たデータによるsevere acute liver injury				Severe Acute Liver Injury(SALI)に該当する複数のICD-9 codes、External-Cause-of-Injury codes (E-codes)、biopsy codeの組み合わせ	特記はされていないが、MINI-sentinelの各パートナー内ではリンケージは可能。	Chart review Hospital medical recordsからREDCap databaseに抽出されたデータ 20以上の検査値、腹水や肝性脳症、黄疸、薬、肝移植、pre-existing liver disease, などからadjudicatorが判断	肝臓の基礎疾患がないが入院の主病名がtoxic hepatitis (ICD9 2種類)の75例、入院の主病名がacute liver failure (ALF)(ICD9 5種類)の74例、とchronic liver disease (CLD)をもつもので入院の主病名がSALI(toxic hepatitisかALF)の75例 を5つのデータパートナーから同数をランダムに選択。数については" To ensure sufficient sample sizes to evaluate the validity"	PPV		100-200例程度のchart reviewでPPVのみを求めるMini-sentinelの研究。	severe acute liver injury (SALI)
53	2013	Walsh KE, Cutrona SL, Foy S, Baker MA, Forrow S, Shoaibi A, et al. Validation of anaphylaxis in the Food and Drug Administration's Mini-Sentinel. Pharmacoepidemiol Drug Saf 2013 Nov;22(11):1205-1213.	アナフィラキシー	Mini-Sentinel Distributed Database (1) HealthCore, Inc.; (2) Humana; (3) three member health plans within the Kaiser Permanente Center for Effectiveness and Safety Research; (4) two member health plans within the HMO Research Network; and (5) Vanderbilt University	アメリカ	2009年1月-2010年12月	不明(5 data partners の8 health plans)	Data Parnersからのランダムサンプルなので、population-basedと考えてよいか?	3つの基準を設けている: A: アナフィラキシーの入院または外来コード B: Aに加えて気管支スパズム、喘鳴、低血圧、エピネフリン使用、ジフェンヒドラミン使用、CPRのどれかの条件 C: Aに加えて、気管支スパズム、喘鳴、低血圧のどれか、かつエピネフリン使用、ジフェンヒドラミン使用、CPRのどれか 入院患者は基準に該当した人。外来は該当してもその30日前以内にアナフィラキシーコードがあったら除外。	記載ないが内部で必要な情報に遡れるようになっていると思われる	2人(救命救急医, アレルギー専門医)による診療録から抽出された情報のレビュー。一致しなかったら3人目のレビューも加える。	150人のランダムサンプル	PPV			わかりやすい研究 Supporting InformationにAbstraction Formを示している。
54	2012	Cutrona SL, Toh S, Iyer A, Foy S, Cavagnaro E, Forrow S, et al. Design for validation of acute myocardial infarction cases in Mini-Sentinel. Pharmacoepidemiol Drug Saf 2012 Jan;21 Suppl 1:274-281.	急性心筋梗塞	Mini-Sentinel Distributed Databaseを用いた。研究には4つのHMOが参画した。SASによるランダム抽出を実施。 ①HealthCore, Inc. ②Humana ③Kaiser Permanente Center for Effectiveness and Safety Research (CESR) ④HMO Research Network (HMORN)	アメリカ	2009	記載なし(複数)	yes 4つのHMO内で患者を特定し、SASを用いランダムサンプリングを行っているため、population-basedと考えられる。	急性心筋梗塞の退院時主病名(ICD-9 410.x0 and 410.x1)	4つのHMO(データパートナー)内で直接患者を特定したのでリンケージは実施していない。	研究チームが作成したabstraction formを用いてnurse abstractorsが抽出した情報と心電図やその他検査結果をもとに、複数のcardiologistsが判断	100例の予定(100例あればPPVを計算しvalidationを行うのに十分と考えている)	PPV		Lessonsが9項目にわたって記載されている。病院ごとに異なる情報の抽出の仕方などに関してabstractor trainingの重要性などが強調されている。	
55	2016	MacFarlane LA, Liu CC, Solomon DH, Kim SC. Validation of claims-based algorithms for gout flares. Pharmacoepidemiol Drug Saf 2016 Jul;25(7):820-826.	痛風発作	Medicare	米国	2006 - 2010	1	no	痛風の診断名と薬剤処方、あるいは、痛風の診断名+処置、の組合わせ	研究が実施された施設では、リンクされたMedicare linked with an electronic medical record (EMR) dataがあると考えられ、そこから該当患者を特定している。Linkage の詳細については記載されていない。	Medical record review 発作に関する医師の記録	3952人におけるgout flareになりうるencounters(Index date[ICD9code、薬、procedureのうち遅い方]の前後7日)。総encounters数の記載はない。 薬に関する503encountersとProcedureに関する290encouters及び(NPV計算のための)ICD9だけの200encouters。	PPV, NPV	痛風発作についてPPVは低く痛風発作は正確には特定できなかった。一方、NPVは高かったので、逆に、痛風発作のない高尿酸血症の症例を抽出することは可能。	ゴールドスタンダードの痛風発作の診断にはAmerican Rheumatism Associationの基準を使用していない。使ったらPPVはもっと下がるかも。高齢者のデータを使っているので, PPVを過大評価しているかも知れない。	NPV計算のための200encoutersは全encoutersからのランダムサンプルの方がよかったのではないか?NPV計算を正しくやるためにはソースの「指標の基準を満たさない者」からのランダムサンプルであるべきだが、「指標の基準を満たさない者」は指標によって異なる。 痛風そのもののPPVは高いので, 研究デザインや目的によっては, このアルゴリズムが有用かもしれない。ただし, 一つの病院でしかやっていないので, 他の医療機関における研究も必要だと考えられる。
56	2010	Semins MJ, Trock BJ, Matlaga BR. Validity of administrative coding in identifying patients with upper urinary tract calculi. J Urol 2010 Jul;184(1):190-192.	尿路結石	Administrative database (Johns Hopkins Hosital Medicine Datamartの入院・外来データ) claims dataを含む	USの単一病院(Johns Hopkins)	2007年11月から2008年10月			上部尿路結石の診断(ICD-9 コード592, 592.0 592.1, 592.9)	1病院の中の保管データ	カルテ(electronic patient record)の中の情報(病名と治療と画像診断)	2007年11月から1年間で上部尿路結石のICD-9を受けた8245人よりランダムサンプルされた100人	PPV			1病院 レントゲン写真が全員にない
57	2016	Moulis G, Germain J, Adoue D, Beyne-Rauzy O, Derumeaux H, Sailler L, et al. Validation of immune thrombocytopenia diagnosis code in the French hospital electronic database. Eur J Intern Med 2016 Jul;32:e21-2.	免疫性血小板減少症(ITP)	FranceのFrench health insurance systemdatabase (SNIIRAM)を構成する外来のデータと入院のデータ(PMSI)	単一病院 (フランスにおけるToulouse 大学病院:ITPのICD10を評価する研究)				入院のprimary, related or associated Dx がICD10D69.3 IPT (Immune thrombocytopenia)	記載はないが、一病院内なので可能と思われる。	チャートレビュー(血小板数[<100 × 10^9/mL] かつ、その他の血小板減少の原因がない)	2014年にToulouse大学病院に入院し、PMSIにICD-10のD69.3が診断名のある265名 (sensitivityのためには、CARMENレジストリ[Gold Standard]で、新規にITPと診断されToulouse大学に入院した成人患者16名)※	PPVとSensitivity フランスの一病院で実施し、入院のClaims dataの入院の診断がITPのN=265からPPVを、地域のITPレジストリに登録され、当該病院に入院した患者16例からSensitivityを推定。		特にassociated diagnosisにITPが含まれる場合、原発性か、骨髄異形成症候群に伴う二次性ITPかの区別がむつかしかった。	一病院にITPのICD10 をもつ入院n=265からPPVを推定。 また、当該病院をカバーする地域の疾患登録にあり当該病院に入院したn=16からSensitivityが推定されている。当該病院で真のケースは200例以上見いだされており、ICD以外の情報からも真のケースを見出す努力をするなどして当該病院の真のケース全例を見出すよう努める方法の方がベターであったと考えられる。
58	1999	Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. J Clin Epidemiol 1999 Mar;52(3):199-207.	糖尿病の15の合併症もしくは併存症	GHCデータベース(アメリカのHMO)	アメリカ 西ワシントン			Yes 米国の1HMOでの研究	糖尿病の15の合併症および併存症をICD-9-CM, CPTで定義。	HMO内でリンケージは不要と考えられる。	"Medical record abstractors"によるChart review	糖尿病患者からランダムサンプリングされた471人が対象となった。 糖尿病の定義は以下のPORT Studyの定義を利用し、の5つの項目のうち1つでも条件を満たす場合を糖尿病とした ①インスリンもしくは経口血糖降下作用薬を3年以上処方されている ②前年のヘモグロビン合計値が7.5 mg/dL以上 ③前年の空腹時血糖が140 mg/dL以上 ④前年の血漿グルコース値が200 mg/dL以上 ⑤GHCに糖尿病での退院歴がある	感度、特異度、PPV	特に記述なし		
59	2016	Molnar AO, van Walraven C, McArthur E, Fergusson D, Garg AX, Knoll G. Validation of administrative database codes for acute kidney injury in kidney transplant recipients. Can J Kidney Health Dis 2016 Apr 7;3:18-016-0108-7. eCollection 2016.	腎移植患者の急性腎不全	オンタリオ州南西部における、 Institute for Clinical Evaluative Sciences (ICES)のCanadian Institute for Health Information Discharge Abstract Database (CIHI-DAD),	カナダ Southwestern Ontario and Ottawa, Ontario, Canada,	2003 - 2012		yes OttawaまたはOntarioで腎移植を受け、6カ月以上経過後に再入院した患者で入院前のcreatinine値が測定されているもの。複数回入院の時にはそのうちの1つをランダムに選択。	初回の腎移植を受けた患者を対象に、 入院時診断名、主病名、としての急性腎不全のICD-10-CA code 、その他の診断名 ICD-10 N17x(Acute renal failure)[1]いずれかのDxとして[2]Admission diagnosisとして[3]main/most responsible diagnosisとして与えられている場合を区別して評価。	encrypted unique identifiers.基本はICESのadministrative databaseとOHDW(Ottawa)またはGamma-Dynacare(Ontario)からのLaboデータのリンクだが、移植のデータやvital statusのデータなど7つのデータベースをより正確な情報を得るためにリンケージ	オタワ病院データウェアハウスから得られた検査値。 Serum Creatininの変化 [入院前の(最後の)クレアチニン値と入院中の最大値の差を検討。] (Acute Kidney Injury Network (AKIN) staging systemのStage1(SCreの1.5-2倍/0.3mg/dL以上上昇)、Stage2(2-3倍上昇)、Stage3(3倍以上上昇または0.5mg/dL以上上昇後で>354umol/L[4.6mg/dL]以上)についてstage1/2/3以上という3つの基準	524例。該当症例全て使用	sensitivity /specificity /PPV/ NPV	腎移植患者の急性腎不全をレセプト上のコードによる特定は, 過小評価あるいは誤分類の可能性がある。コードによる当該事象の特定は有効ではない可能性がある。 「感度が最大42.1%」に関してIncidenceやAKIのconsequenceの研究ではunusableと評価。		
60	2010	Ingeman A, Andersen G, Hundborg HH, Johnsen SP. Medical complications in patients with stroke: data validity in a stroke registry and a hospital discharge registry. Clin Epidemiol 2010 Aug 9;2:5-13.	脳卒中の合併症	治療の質をモニターし改善するためのプロジェクトのDanish National Indicator Project (DNIP)および入院患者のレジストリ Danish National Registry of Patients (NRP)	デンマーク	2003/02~2006/12	急性脳卒中に特化したコペンハーゲンにある10病院	yes	DNIP: 以下の7つ合併症とその診断日:①pneumonia,②urinary infection, ③pressure ulcer, ④falls, ⑤deep venous thrombosis(DVT), ⑥pulmonary embolism (PE), ⑦constipation. NRP: 以下の ICD-10コード: ①pneumonia (DJ12-DJ18), ②urinary infection (DN30.0, DN30.8, DN30.9, and DN10), ③pressure ulcer (DL899), ④falls (DR297 and EUHE), ⑤DVT (DI82.9, DI82.9A-E), ⑥PE (DI26), ⑦ (DK590).	Civil registry numberを使用した。 デンマーク市民を識別する固有のものであり、性別、生年月日の情報が含まれる。レジストリとのリンケージに用いられる。	Chart review Table 2に記載されたアルゴリズムを用いシングルレビューが行われた。	2003-2006年に急性のStrokeの8024例の5%ランダムサンプルの417例に加えて、DNIPコードの7つの合併症に関し、25例ずつランダムサンプリングで使用した。レビューしたのは合計589例	PPV/NPV/Sensitivity/Specificity PPVとNPVについては全589例から、感度と特異度については、ランダムサンプルの417例から計算されている。	SensitivityはおおむねDNIPがDNPより高い。	シングルレビューによる評価およびレトロスペクティブなデータ収集が限界点として挙げられている。	脳卒中の患者という"population"で実施されている。 ①全体のランダムサンプル+②指標を満たす者のランダムサンプルを得て、①+②でPPVとNPVを求め、①のみから感度、特異度を求めている。
61		Ducharme R, Benchimol EI, Deeks SL, Hawken S, Fergusson DA, Wilson K. Validation of diagnostic codes for intussusception and quantification of childhood intussusception incidence in Ontario, Canada: a population-based study. J Pediatr 2013 Oct;163(4):1073-9.e3.	腸重積症	1) 入院に関するadministrative data:Canadian Institute for Health Information Discharge Abstract Database (CIHI-DAD), 2) 救急外来のdata: emergency department (ED) visits from the National Ambulatory Care Reporting System (NACRS), 3) 外来のClaims:physician claims data from OHIP	Children's Hospital of Eastern Ontario			Yes: ゴールドスタンダードはChildren's Hospital of Eastern Ontario (CHEO)でのchart reviewによるが、CHEOは地域の「sole tertiary pediatric health care」	診断名(ICD9か10) +CHEOのprodecure code(intestinal reduction/laparotomy/diagnostic enema)または医事請求コード	ユニークな識別子:All patient-level data were linked across the multiple databases using the encrypted ICES unique identifier	Chart review	カナダのオンタリオ州で 1995-2010年度で腸重積症.で入院した患者。Chart reviewの結果565casesが確認された。	Sensitivity/Specificity/PPV/NPV			オンタリオにおける全ての腸重積が単一の病院に集まるとの仮定のもとにintussusception-related ICD codesと"intussusception"というテキスト検索をして特定したケースが地域における全ケースとして、Sensitivity, Specificityなど指標を計算。All possible casesのアプローチ。
62	2016	John A, McGregor J, Fone D, Dunstan F, Cornish R, Lyons RA, et al. Case-finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. BMC Med Inform Decis Mak 2016 Mar 15;16:35-016-0274-7	不安、うつ	WalesのGeneral Practice Database (GPD)	英国(Wales)	2000 - 2009		yes Wales のCaerphilly地区の住人	12 algorithms for Read codes to define cases according to various criteria.	引用文献によればMACRAL (Matching Algorithm for Consistent Results in Anonymised Linkage)と呼ばれる方法で、氏、名、性、郵便番号、DOBの5変数によるDeterministic LinkageとProbabilistic Linkageの両方を用いている。(BMC Medical Informatics and Decision Making 2009, 9:3)	郵送法によって得た住民からの調査から求めた(SF-36の簡易版の5項目からなる) MHI-5.≤60(値が大きいと健康)をcut-point.	SALE databankとCaerphilly Health and Social Needs Survey (CHSNS)とでリンク可能だった対象集団は以下の通り: Wave1(2001年baseline survey):2584人 Wave2(2008年 follow-up survey):1195人 (2799 individuals)	PPV/NPV/Sensitivity/Specificity	アルゴリズム10が、今後用いるとしては一番いい指標だった。		
63	2001	Hassey A, Gerrett D, Wilson A. A survey of validity and utility of electronic patient records in a general practice. BMJ 2001 Jun 9;322(7299):1401-1405.		EMIS clinical systemの電子的データ	イギリス 北ヨークシャー州スキプトン	1999	1病院(Fisher Medical Center)	yes	15疾患(乳がん、前立腺がん、糖尿病、肥満、甲状腺機能低下、甲状腺機能亢進、痛風、鉄欠乏性貧血、緑内障、パーキンソン病、虚血性心疾患、高血圧、喘息、関節リウマチ)のRead Code	リンケージは実施されていない。	病名コードおよび検査や治療を組み合わせたアルゴリズム。	13606例を使用した。	PPV/Sencitivity			
64	2016	Mulder B, Groenhof F, Kocabas LI, Bos HJ, De Vries TW, Hak E, et al. Identification of Dutch children diagnosed with atopic diseases using prescription data: a validation study. Eur J Clin Pharmacol 2016 Jan;72(1):73-82.	小児アトピー(喘息、皮膚炎、鼻炎)	オランダ北東部の開業医の登録データベース(RNG:Registration Network Groningen database)	オランダ	2001/01/01-2010/12/31	3つのGP診療所(17人のGP)	yes	アトピー性疾患の治療薬とその処方回数	DB内で、患者はunique patient numberで特定されるとある。同一DB内の薬剤によるアウトカム定義のvalidityを診断コードをgold standardとして評価したもので、外部データとのリンケージはされていない。	アトピー性疾患に該当するInternational Classification for Primary Care(ICPCコード)、アトピー性皮膚炎:S97、喘息R96、アレルギー性鼻炎R97	2001年1月1日から2010年12月31日に1回でもデータベース上に記録がある0から10歳の7439例	感度と陽性的中度	皮膚炎やアレルギー性鼻炎の薬剤は疾患の特定に使うのは難しい	GPによるICPCやATCの誤分類の問題	
65	2010	Sicras-Mainar A, Blanca-Tamayo M, Gutierrez-Nicuesa L, Salvatella-Pasant J, Navarro-Artieda R. Clinical validity of a population database definition of remission in patients with major depression. BMC Public Health 2010 Feb 11;10:64-2458-10-64.	大うつ病の寛解	スペインのprimary care data	スペイン	2003/01～2007/03	Primary careの6施設	no	初発の治療終了後、少なくとも6か月以上新規のADが処方されていないことを定義とした。	リンケージは実施されていない。	精神科専門医による臨床経過の記載を判断基準をもとに寛解か否か判断された。	133例を使用した。	PPV/NPV/Sensitivity/Specificityおよびκ	記載なし	一般化可能性については留意する必要がある。	
66	2001	Segal JB, Ness PM, Powe NR. Validating billing data for RBC transfusions: a brief report. Transfusion 2001 Apr;41(4):530-533.		Medicare and Medicaid administrative data	アメリカ			No:1病院での研究	Johns Hopkins Medicine DataMartからのICD-9 Procedure Code 9904	medical record number	Blood bank computerにおける記録。Blood bankにない(false positiveの)17例についてはdischarge summaryも使って輸血を受けていないことを確認した。	FPとFNの割合から850人を目標にしたが、実際には716人が対象となり、358人が赤血球輸血を受けた。	感度、特異度	特に記述なし	特に記述なし	赤血球輸血

【付録3文献リスト】

67	2015	Sato I, Yagata H, Ohashi Y. The accuracy of Japanese claims data in identifying breast cancer cases. Biol Pharm Bull 2015;38(1):53-57.	乳がん	日本のレセプト情報	日本	2011	1	No 特定の病院で実施。 Population baseではない。	診断コードや処置コードの組み合わせで、14通りに乳癌を定義： 1 Diagnosis of breast cancer 2 Diagnosis of breast cancer+diagnosis code related to breast cancer 3 Diagnosis of breast cancer+diagnostic imaging code 4 Diagnosis of breast cancer+biopsy code 5 Diagnosis of breast cancer+marker test code 6 Diagnosis of breast cancer+surgery code 7 Diagnosis of breast cancer+chemotherapy code 8 Diagnosis of breast cancer+medication code 9 Diagnosis of breast cancer+radiation procedure code 10 Diagnosis of breast cancer+the other code related to breast cancer 11 Diagnosis of breast cancer+diagnosis code related to breast cancer or marker test code 12 Diagnosis of breast cancer+surgery, chemotherapy, medication, or radiation procedure code 13 Diagnosis of breast cancer+diagnosis code related to the breast cancer, marker test code, surgery, chemotherapy, medication or radiation	一病院内のデータでありリンクは不要	医療機関内のがん登録レジストリ	院内がんレジストリーに登録されている患者50056名のなかから、乳癌患者 633例を抽出	PPV/NPV/Sensitivity/Specificity (PPVに重点を置いている)	診断コードや処置コードの組み合わせで、レセプトデータから、高い精度で乳がん患者を特定できる。 考察でPPV70%は「低い」との記載あり		
68	2009	Harboe KM, Anthonsen K, Bardram L. Validation of data and indicators in the Danish Cholecystectomy Database. Int J Qual Health Care 2009 Jun;21(3):160-168.	胆嚢摘出術を受けた患者に関連する情報	Danish Cholecystectomy Database (DCD)に含まれるNational Patient Registryのデータ	デンマーク	2004～2005	2病院	no Gold standardがコペンハーゲンの2病院であるため。	胆嚢摘出術の種類、手術後入院期間、再入院の有無等Box 2に記載された定義を使用した。	Unique personal identification numberを使用した。	Medical recordsを2名がレビューし判断した。	National Patient Registryに記録されている腹腔鏡または回復でのcholecystectomyの2病院における症例1363例が対象となった。	PPV/NPV/Sensitivity/Specificity		胆嚢摘出術および総胆管の結石摘出術を区別することが困難であった。	National Patient Registryにcholecystectomyの記録のある1363名の中で感度、特異度、PPV、NPVが求められているが、PPVしか求められないはず。
69	2007	Zgibor JC, Orchard TJ, Saul M, Piatt G, Ruppert K, Stewart A, et al. Developing and validating a diabetes database in a large health system. Diabetes Res Clin Pract 2007 Mar;75(3):313-319.		Electronic Medical Record Archival etrieval System (MARS)というrepositoryにためられたadministrative data	アメリカ University of Pittsburgh Medical Center(UPMC)	2000年1月から2003年12月			糖尿病治療薬の使用 外来の診断 ERにおける診断 A1cの測定 Glucose>200mg/dl 入院における診断。 以上の組み合わせ	特に記述なし	Chart Review	①ICD-9 250で入院、救急もしくは外来の有無 ②A1cの結果 ③血糖値＞200mg/dl ④糖尿病治療薬の使用 のいずれかを満たす99,144人から ①622人がランダムサンプリングされた。(そのうち254人がレビューされた) ②10% random sample from general internal medicine clinic data(N=94、Table 2ではN=95)	感度、特異度、PPV	特に記述なし	糖尿病のうち1型と2型の診断を分けることが困難であった。	糖尿病
70	2016	Schulman KL, Lamerato LE, Dalal MR, Sung J, Jhaveri M, Koren A, et al. Development and Validation of Algorithms to Identify Statin Intolerance in a US Administrative Database. Value Health 2016 Sep - Oct;19(6):852-860.	スタチン非耐性	Henry Ford Health System administrative database. EMRとadministrative dataを含む。	米国	2005-2010	地域のHMO	Yes 米国中西部のHMO(規模の記載ない)	まずはinclusion criteria: スタチン関連イベント(※)が期間中にあり、かつ、スタチン関連イベント前に下記ICD9CMコードがある：(高脂血症(272.4)、高コレステロール血症(272.0)、混合高脂血症(272.2)) ※：スタチン処方の中止、変更(強スタチンから中・弱スタチンへ、あるいは、中・弱スタチン間の変更；論文内定義あり)、あるいは、減量。 スタチンの有害事象は、Table2に示すICD9CMコードで定義。 スタチンのIntensity：スタチン種類ごとにdaily doseで"moderate to"(中・弱)と"high"(強)に区分(Table 1)。 上記患者から、以下の患者を抽出し、CV-risk(合併症の種類で判断)で層別 アウトカム定義はadministrative dataのみを用いる。 【スタチン絶対的不耐性】は、以下3つの全て満たすものとして定義： 1)2種以上のスタチン薬の中止(横紋筋融解症の場合は1種) 2)そのスタチンがSIやAEに関連している 3)最終のスタチン使用時にCK値上昇またはAEの出現 【スタチン用量依存的不耐性】は、以下3つの全て満たすものとして定義： 1)強スタチンから弱・中スタチンへ処方変更 2)その後、同じ強度のスタチンの処方はない 3)その100日間以上の観察期間あり	リンクされている	EMRより、スタチン処方の日付、量、AE種類、スタチン中止／減量の理由を調べ、以下の2つに該当するか否かを調査。 【スタチン絶対的不耐性】は、以下3つの全て満たすものとして定義： 1)2種以上のスタチン薬の中止 2)その後、他のスタチン処方を受けない 3)スタチン薬中止の理由は有害事象である、薬と有害事象との関連に関する記述、患者がスタチンに不耐性/アレルギーであるとの記述。 【スタチン用量依存的不耐性】は、以下3つの全て満たすものとして定義： 1)強スタチンから弱・中スタチンへ処方変更 2)中止した強スタチンは、SIやAEに関連している 3)その後、同じ強度のスタチンの処方はない	スタチン処方のあった46,783名より、組入基準に該当する者が5473名いた。このなかから1000名をランダム抽出	sensitivity /specificity /PPV/ NPV/ Cohen κ	感度、特異度、PPV、NPVの標準(閾値)は存在しないが、 ・絶対的不耐性における感度は低い(50%)ので、疾患の発生率や有病率の検討はできない。 ・NPVは高い(99%)ので、絶対的不耐性ではないスタチン使用者を特定することはできる。 ・指標の閾値の標準値がない限り、研究者は「感度86%」が疾患の発生率や有病率を検出するのに十分かどうか、だとか、「PPV 77% 」が用量依存的不耐性の患者を特定するのに十分かは判断できない。 ・著者は研究者に適宜感度分析を実施するように奨励する。		
71	2014	Tu K, Wang M, Jaakkimainen RL, Butt D, Ivers NM, Young J, et al. Assessing the validity of using administrative data to identify patients with epilepsy. Epilepsia 2014 Feb;55(2):335-343.	てんかん	以下の4つのadministrative databases(を組み合わせてえられるアルゴリズム) ①Canadian Institute for Health Information hospital discharge abstracts (CIHI DAD) ②the National Ambulatory Care Reporting System (NACRS) ③the Same Day Surgery Database (SDS) ④Ontario Health Insurance Plan (OHIP) physician billing database	カナダ(オンタリオ)	2011(6～11月)	83プライマリーケアドクター	yes	受診回数と期間で22のアルゴリズムを作って下記のものが一番いいと結論している An algorithm of three physician billing codes (separated by at least 30 days) in 2 years or one hospitalization using ICD-10 codes G40.x after the year 2002 and 345.x (excluding 345.2 petit mal status and 345.3 grand mal status) ICD-9 codes prior to 2002	明確には記載されていないが、valid health card number と date of birthのある者からランダムサンプルをとっており、これらをリンケージに使ったと思われる。	6人のchart abstractorによる ElectronicMedical Record Administrative data Linked Database (EMRALD)のチャートレビュー 評価シートはAppendix のFig S1に示されている。 Epilepsy(definite, probably, possible)と区分されたabstractionについてはfamily physicianとepileptologistがレビューした。	対象となりうる 73,013 人のうち7500人をランダムサンプリング	sensitivity/specificity/PPV/NPV Youden Index(ROC曲線下面積のような意味合い)＝sensivity + specifiity - 1			EMRALDの70000人強からの10%ランダムサンプルでadministrative databases(を組み合わせた指標)を評価し、その指標で全Ontarioのepilepsyのprevalenceとincidenceを推定 Abstractorsのinter- intra-rater reliabilityを見るために、10%サンプルで一致度を評価。6人全員でκ＞0.8と記載されている。
72	2013	Ooba N, Setoguchi S, Ando T, Sato T, Yamaguchi T, Mochizuki M, et al. Claims-based definition of death in Japanese claims database: validity and implications. PLoS One 2013 May 31;8(5):e66116.	死亡	a commercially available claims database covering working population	日本	2005-2009	不明	yes	クレームの中の死亡と入院記録などを組み合わせて9個のアルゴリズムを作成(死亡の後に請求がある状況をゾンビと呼びそれもアルゴリズムの一部に使用)	特に記述はないがデータの中にクレームデータと保険者の情報が両方含まれていたと考えられる	保険の資格喪失理由の部分に書かれている死亡	データベース内の195,193人の保険受給者全員	sensitivity/specificity/PPV		our study population did not include subjects aged 19 years or younger and 75 years or older	
73	2016	Bowker SL, Savu A, Donovan LE, Johnson JA, Kaul P. Validation of administrative and clinical case definitions for gestational diabetes mellitus against laboratory results. Diabet Med 2017 Jun;34(6):781-785.	妊娠糖尿病	カナダ・アルバータ州のadministrative claim database。5つのDBがリンクされているが、①the Ambulatory Care Classification System (the 'outpatient' database) ②the hospital Discharge Abstract Database (the 'inpatient' database) ③the Alberta Perinatal Health Program (APHP) databaseに関して[1]①only、[2] ②only、[3] ①+②、[4]③が評価されている。。	カナダ・アルバータ州の情報(レセプトデータと州の妊婦レジストリ)	2008年10月から2010年3月			妊娠糖尿病の病名(ICD-10)	カナダのa unique, scrambled maternal personal health number 評価される'outpatient' databaseと'inpatient' databaseのリンクのほか、ゴールドスタンダードのpopulation-level laboratory data	laboratory data (妊娠中のOGTT試験で2つ以上の異常値を妊娠糖尿病と判断)	アルバータ州で2008年10月から2010年3月までに出産した妊婦のうち、多胎児出産、臨床検査値や属性データ欠損、すでに糖尿病患者等を除外した、58338人。	PPV/sensitivity/specificity	入院＆外来情報を統合すれば、ICD10は使える。 感度、特異度は高いがPPVハ57%。	False Positiveの78%超がimpaired glucose tolerance (IGT) in pregnancyの基準に合致。IGT in pregnancyがgestational diabetesとしてコードされることがPPVが低い理由。	
74	2015	Stringer E, Bernatsky S. Validity of juvenile idiopathic arthritis diagnoses using administrative health data. Rheumatol Int 2015 Mar;35(3):575-579.	若年性特発性関節炎(JIA)	Medical Services Insurance (MSI) database	カナダのノバスコシア州のMedical Services Insurance (MSI) database。4年間(Fiscal years 2005/2006 to 2008/2009)				JIA(若年性特発性関節炎)のICD9 Codeに関する4つのアルゴリズム。①any physicianが2年内2回以上、小児リウマチ医が(②1回または③2年内に2回以上)、④小児リウマチ医以外が2年内2回以上。	暗号化された健康カード番号。	Chart Review: IWK Health Centre (IWKHC) in Halifaxにある，MSIデータベースと同時期の小児リウマチクリニックの臨床データベースから，小児リウマチ医が2004 International League of Associations for Rheumatology (ILAR) criteriaに基き診断した症例をクリニカルチャートを用いてレビュー	母集団(IWKHC)は900,000人。Chart Reviewの対象はincidentの 若年性特発性関節炎(JIA)94例+いずれかのアウトカム定義Positiveの60例程度。	PPV, sensitivity	小児リウマチ医による診断名の感度が高かった。他の研究ではリウマチ医と，2つ以上のリウマチ関連検査結果の組合せが最もJIAを特定できたが，検査値データがなかったので実施できなかった。	感度が高いがPPVが低い。専門医による診断に限ってもPPVは低く、偽陽性が多い理由について、著者らは"rule out JIA"の可能性、暫定的な診断(provisional diagnosis)であった可能性を指摘している。	
75	2014	Widdifield J, Ivers NM, Young J, Green D, Jaakkimainen L, Butt DA, et al. Development and validation of an administrative data algorithm to estimate the disease burden and epidemiology of multiple sclerosis in Ontario, Canada. Mult Scler 2015 Jul;21(8):1045-1054.	Multiple Sclerosis	Ontario Health Insurance Plan (OHIP) Databaseに含まれる①入院に関するCanadian Institute for Health Information's Discharge Abstract Database (CIHI-DAD)、②外来に関するNational Ambulatory Care Reporting System (NACRS)	カナダのオンタリオ州の130万人の住民				Ontalio Health insurance planのMSのICDコードを有する。一定期間における診断の数と入院の有無の組合せによる複数の定義。	encrypted health card numbers	73003症例の中からsystematic search strategy(引用されているKryskoらの論文によれば、①EMRのcumulative patient profile (CPP) 、②free text entries for MS-related phrases、③physician billing codes)で全症例を含むall possible cases特定した。 Kryskoらの論文によれば、5人の訓練を受けたabstractorsが①definite MS ②possible MS ③MSがrule outされた ④MSへの言及が全くない、に分類。最終的に専門家がチェックして必要があれば修正。10%サンプルを使って、inter-intra rater reliabilityをκ係数で評価。	EMR Administrative data Linked Database (EMRALD)の73003人から得た"all possible cases"(引用されているKryskoらの論文によれば943例)のChart Reviewにより247人のMS患者を特定。	Sensitivity/Specificity/PPV/NPV 引用されているKryskoらの論文によれば、inter, intra-rater reliabilityをkappa係数で評価。	感度とPPVが良かったアルゴリズムは一回の入院か，2年以上にわたり5回以上の診断結果がある症例である。		有病割合の低いMultiple SclerosisではRandom Sampleでは十分な数の真のケースを見出せないとして、複数の方法(free textの関連する記載、physician claimsなど)からall possible casesを特定する方法を用いたもの
76	2011	Amed S, Vanderloo SE, Metzger D, Collet JP, Reimer K, McCrea P, et al. Validation of diabetes case definitions using administrative claims data. Diabet Med 2011 Apr;28(4):424-427.	糖尿病	administrative databases of the British Columbia Ministry of Healthy Living and Sport	カナダ British Columbiaの20歳未満の971 492 children and adolescentsにおける糖尿病のPrevalence特定の研究の一環として 単一病院(British Columbia Children's Hospital)で実施				Icd9かICD10の傷病名コード，入院や診断名の回数，インシュリンの処方，経口糖尿病薬の処方等の組合せ。	unique personal health number	British Columbia's Children's Hospital の Endocrinology andDiabetes Unit databaseに登録されている糖尿病患者(すでに専門家により診断されており、バリデーション研究のための新たなChart Reviewなどは行われていない)	2611人。	Sensitivity			
77	2015	Lam NN, McArthur E, Kim SJ, Knoll GA. Validation of kidney transplantation using administrative data. Can J Kidney Health Dis 2015 May 18;2:20-015-0054-9. eCollection 2015.	腎移植	①Claims：Ontario Health Insurance Plan [OHIP] ②移植に特化したレジストリ：Canadian Organ Replacement Register [CORR] ③入院患者に関するDB：Canadian Institute for Health Information Discharge Abstract Database [CIHI-DAD]	カナダ オンタリオ	2008/01/01-2011/12/31	3	オンタリオの3つの主要な移植センターからの情報で実施されており、population-basedといえるかどうかは不明	各データベースの疾患コード CORR:treatment code、CHI:CCIcode、OHIP:Fee code	each patient's encypted healthcard number	3つの主要な移植センターからの直接の情報	2008年1月から2011年12月までに肝移植を行った1112例	感度、陽性的中度		同じ地域の他の3つの病院のデータが含まれていない(が結果に影響はない、という見解)	
78	2013	Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. Chronic Dis Inj Can 2013 Jun;33(3):160-166.	心不全	Canadian Institute for Health Information (CIHI) のhospital discharge abstract database (DAD) と the same-day surgery database (SDS), , および OHIP database of physician fee-for-service billings or shadowbillings.	カナダ(オンタリオ州)	2007年6-12月(EMRALD) 2004年12月2005年8月(76のFamily phycician practices)	Canadian Cardiovascular Outcomes Research Team (CCORT)の17physicianと76 fee-for-service family physician practices	yes	心不全を示唆するICD-9とICD-10コード 9個のAlgorithm (入院か外来か、1回か2回か、1-3年、のパターン)を、それぞれについてNarrow(心不全のみ)とBroad(心筋症と肺水腫のコードも含む)で検証	暗号化された health card number	訓練されたチャートレビューワーによるチャートレビュー	17physicianからの20歳以上のランダムサンプル969人と76のFamily physicianからの38歳以上のランダムサンプル2472人のうち、EMRALDに含まれ、38歳以上で過去3年に2回以上の受診があり、OHIP card numberをもつ2338 patients	sensitivity/specificity/PPV		二つのデータソース。Chart reviewは3人(17 physicians)、2人(76 family practices)が実施し、inter-ovserverのκはいずれも>0.8だった。	
79	2012	Rezaie A, Quan H, Fedorak RN, Panaccione R, Hilsden RJ. Development and validation of an administrative case definition for inflammatory bowel diseases. Can J Gastroenterol 2012 Oct;26(10):711-717.	炎症性腸疾患	4つのadministrative database(から作成されたアルゴリズム) 1. Health Care Insurance Plan Registry (HCIPR) 2. Physician billing claims (PC) database 3. Hospital Discharge Abstract Database (DAD) 4. Ambulatory Care Classification System (ACCS) database	Western Canadian health region (アルバータ)	2000-2004	記載なし	yes	150のアルゴリズムたとえば2-4-2-2(ICD9またはICD10で定義した)IBDによる入院2回または外来4回またはACCS(ambulatory care)2回「以上が2年以内」	Development phase(定義の作成)では、内視鏡データベースとHCIPR(Albertaの全住民のregistry)のリンケージは名前とdemographics(性、年齢。人種、その他も？)。HCIPRとその他のDBはa unique lifetime personal health number (PHN)で。 Validation phase(定義を外部のDBでテスト)のEdmonton地区内視鏡DBにはPHNが含まれている(らしい)。(それ以外のDBとPHNでリンケージしたとある[PHNs were linked to --])。	Development phaseでは内視鏡の報告またはChart Reviewで"Lennard-Jones criteria"を満たす。 Validation phaseでは内視鏡の報告	Development phaseでは内視鏡DBの21193人。うち、他のDBとリンクできたのは16,838人(IBD1399人、非IBD15,439人)。 Validation Phaseでは31832人の内視鏡の記録のうち、5433人のIBDpatientsを特定。その他のDBとリンクできたのは5201人。Validation phaseでは感度のみを調査したようだ。	Development phaseではPPV/NPV/感度/特異度 Validatoin phaseでは感度。感度が、Development phaseと同じ値をもつ、年によって変化せず安定しているなどを基準に2-4-2-2のアルゴリズムを採用			この研究の"Population"は内視鏡検査を受けた人たち(Development phaseでは21193人。ただし、他のDBとリンクできたのは16838人。Table 1にもIBDのN=1399、Non IDBのN=15,439となっている。Validation phaseではendoscopy databaseで特定できた5433人のうちリンクができた5201人)、よって地域のPopulationではSpecificityやNPVは異なるはず。ある特定の検査を受けた人をPopulationとしているという意味で、特定の疾患の人たちにおける合併症などのValidation 研究と共通点をもつ。 150もの定義を評価している点、Development phaseで作成した定義をValidation Phaseで確認している点などがユニーク。

80	2012	Lacasse Y, Daigle JM, Martin S, Maltais F. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. Can Respir J 2012 Mar-Apr;19(2):e5-9.	COPD	Quebec州の医療保険RAMQがもつ入院患者に関するuniversal medical insurance registration (MedEcho database)	カナダ（ケベック州）	2003-2004	40	yes	COPDの退院主病名 (International Classification of Diseases, Ninth Revision codes 491, 492 or 496)	health insurance番号とand hospital file 番号のリストを各病院に送って、各患者のデータを抽出してもらい、それを匿名化して返してもらった	二人の呼吸器内科医が病歴とスパイロメトリーの結果から判断	まず全97病院を入院数に応じて4グループに分け、それぞれから10病院を選択、さらに、それぞれの病院からCOPDの主病名がついた入院患者を同確率でランダムで抽出、合計1221人	PPV 評価者間の一致についてκ	PPVが85%以上なら"valid"と事前に決めていた 高齢者と女性の診断は正しい可能性が低く、教育病院は正しい病名である可能性が2倍高かった	COPDの診断がどこか他でされているかもしれないこと（その場合はスパイロメトリーなど客観的な数値がpatient fileになくてもCOPDと判断した）	入院主病名としてのCOPDだけ。（外来病名、secondary diagnosisは検討されていない）
81	2016	Jorgensen LK, Dalgaard LS, Ostergaard LJ, Andersen NS, Norgaard M, Mogensen TH. Validity of the coding for herpes simplex encephalitis in the Danish National Patient Registry. Clin Epidemiol 2016 May 31;8:133-140.	単純ヘルペス脳炎	Danish National Patient Registry (DNPR).	デンマーク	2004 - 2014		yes	単純ヘルペス脳炎の病名（ICD10）。初回のみ。15歳未満は除く	ten-digit civil registry number (Civil Personal Registration [CPR] number) によりDNPRを ①the Danish Microbiology Database (MiBa) ②departments of clinical microbiology ③patient medical records とリンク	①the Danish Microbiology Database (MiBa) ②departments of clinical microbiology ③patient medical records から得られるPCRによるDNAまたはELISAが確認され、かつ臨床的な3つの条件（[1]症状[2]画像[3]髄液中の細胞）の2つ以上	442。該当症例全て使用。	PPVとSensitivity（Sensitivityについては本文のみ） Sensitivityについては"Viral encephalitis without specification"でHSV1の5例を見出しsensitivity=105/(105+5)として求めている。"all possible cases"のアプローチと考えられる。	PPVが低い理由の一つは臨床所見をゴールドスタンダードに含めたためだが、臨床所見を含めない場合でもPPV<67%。（DNPRではなく）microbiological dataのみの場合PPVは十分高い。"要求されるdata qualityのレベルはresearch questionとstudy designによる"。 PPVが11年間ほぼ一定だが、PCRがよりルーチンに使われるようになっており、検査値のみを使うと発生率が上昇しているように見えるなど注意が必要。	ミスコーディングがどのデータベースにも存在した。	
82		Setoguchi S, Solomon DH, Glynn RJ, Cook EF, Levin R, Schneeweiss S. Agreement of diagnosis and its date for hematologic malignancies and solid tumors between medicare claims and cancer registry data. Cancer Causes Control 2007 Jun;18(5):561-569.	がん (lymphoma, leukemia, breast, colorectal, stomach, and lung cancer)	Medicare 薬の情報はMedicareの情報をPharmaceutical Assistance Contract for Elderly (PACE)にリンクさせて取得	アメリカ ペンシルベニア州				定義1：がんのDx ICD9（1回）と治療（手術、抗ガン剤、放射線）の組み合わせ。定義2：2年以内にがんのDx2回以上。定義3：1または2。定義4：1回以上の診断。 定義3を採用。	social security number, gender, and date of birth	Pennsylvania State (PA) Cancer registry data	US ペンシルベニア州のmedicareの肺がん、大腸がん、胃がん、乳がん、リンパ腫、白血病のコードをもつ6996人	PPV/NPV/Sensitivity/Specificity		PPVが特に稀なリンパ腫（56.6%）や白血病(37.7%)で低い理由として、がん登録が不完全である可能性を指摘し、がん登録でとらえられていない患者の割合とそれがPPVに与える影響について検討する感度分析も行っている。	
83	2004	Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. J Clin Epidemiol 2004 Feb;57(2):131-141.		ケベック州のmedical service claims diagnostic database	カナダ ケベック州				①Drug-disease contraindicationに関連する14疾患に関するICD-9 Heart block 426.0-426.9 Hypertension 401.0-401.9 Hypotension, postural 458.0 CHF 428.0-428.9, 429.3 Raynaud's 443.0 Renal failure 585.0-586.9 Diabetes 250.0-250.9 Asthma 493.0-493.9 COPD 490.0-491.9, 494.0-494.9, 496.0-496.9 Dementia 290.0-290.9, 294.1, 294.9 Glaucoma 365.0-365.9 Gout 274.0-274.9 Prostatic hypertrophy 600.0-600.9 Peptic ulcer 530.1, 578.0-578.9, 531.0-535.9 ②Charlson Comorbidity Indexの構成要素の18疾患に関するICD-9CMコード ③32 Ambulatory Diagnostic Group Clusters (ADGs)のうち30	unique identification number	MOXXI studyのmedical chart review	66歳以上でMOXXI study施設に少なくとも2回来院した14980人が対象となった。	感度、特異度	特に記述なし	データベース中のmissing or unusable diagnostic codesの存在がFNを上げる。	Heart block Hypertension Hypotension, postural CHF Raynaud's Renal failure Diabetes Asthma COPD Dementia Glaucoma Gout Prostatic hypertrophy Peptic ulcer
84	2009	Gonzalez-Fernandez M, Gardyn M, Wyckoff S, Ky PK, Palmer JB. Validation of ICD-9 Code 787.2 for identification of individuals with dysphagia from administrative databases. Dysphagia 2009 Dec;24(4):398-402.	嚥下障害 (dysphagia)	administrative database of tertiary hospital	アメリカの a tertiary hospital	2007年1月から6月		No：一病院で実施された研究	ICD-9 787.2	一病院における研究でリンケージはされていないと考えられる。	VFSS（sequential videofluorographic swallowing study）のレポート	VFSS（sequential videofluorographic swallowing study）に参加した168人患者を対象とした。	（特定の検査を受けた168人における）感度、特異度、PPV、NPV	特に記述なし	病院の滞在日数により、コードの欠損有無が異なっていた。	
85	2014	Kim GE, Shlyankevich J, Kimball AB. The validity of the diagnostic code for hidradenitis suppurativa in an electronic database. Br J Dermatol 2014 Aug;171(2):338-342.	化膿性汗腺炎	Massachusetts General Hospital database	米国	1980- 2013	1	no	最低二回のICD-9コード(705.83)	1病院の研究なのでリンケージは必要なかったと思われる	1人の医学生によるcomplete medical records review、診断に疑問がある時は（when the diagnosis of HS was questionable）皮膚科医による確定	1168人の該当患者全員	PPVのみ	2回だけではなくコードが入っている回数が増えるほどPPVがあがった。		
86	2014	Rishi MA, Kashyap R, Wilson G, Hocker S. Retrospective derivation and validation of a search algorithm to identify extubation failure in the intensive care unit. BMC Anesthesiol 2014 May 23;14:41-2253-14-41. eCollection 2014.	人工呼吸からの抜管	Mayoクリニックの神経科学ICU（NICU）に入院した患者の医療記録	アメリカ ミネソタ	2002年1月1日から2011年12月31日	1	no	抜管72時間後の再挿管 DDQBというテキストサーチのソフトにより抜管失敗を定義	なし	2名の医師による独立したマニュアルレビュー	2002年1月から2011年12月の入院患者からランダムに抽出された100名	感度、特異度、陽性的中度、陰性的中度	抜管失敗のリスクファクター同定の第一歩	特になし	
87	2013	Smischney NJ, Velagapudi VM, Onigkeit JA, Pickering BW, Herasevich V, Kashyap R. Retrospective derivation and validation of a search algorithm to identify emergent endotracheal intubations in the intensive care unit. Appl Clin Inform 2013 Sep 4;4(3):419-427.	緊急気管内挿管	Date from the Mayo Clinic Life Sciences System	アメリカ	2010-2011	1	no	商用ソフト（Data Discovery and Query Builder [DDQB]）を使い用語を入力して特定: サーチアルゴリズムは (Emergency: Yes) event type: "intubation" event type: "intubation" or "intubation V2"（medical noteが"finalize"される時のVersion 2の挿管。） or "sedated" event type: "intubation" or "intubation V2" or "sedated"	1病院の研究なのでリンケージは必要なかったと思われる	別々の2人の医師によるカルテレビュー	2,689人のうち、2010年の患者の中から 450人を選んでアルゴリズムを作成し、2011年の患者の中から450人を選んでvalidationした。	sensitivity/ specificity/PPV/NPV			Discussionの最後の"this comprehensive EMR system is unique to Mayo Clinic"にあるように、基本Mayo Clinic以外では本研究の結果を使うことはできない（特にintubation V2など病院固有のサーチ言語をアルゴリズムに入れている点）。1病院で実施する研究でももう少し、generalizabilityの高い指標を見出すことはできたかもしれない。 日本の病院単位の研究への教訓にもなりうる。
88	2013	de Achaval S, Feudtner C, Palla S, Suarez-Almazor ME. Validation of ICD-9-CM codes for identification of acetaminophen-related emergency department visits in a large pediatric hospital. BMC Health Serv Res 2013 Feb 21;13:72-6963-13-72.	子供のアセトアミノフェン関連の救急受診	HoustonのTexas Children's Hospital に保存されているadministrative database	アメリカ	2005-2010	1	no	救急の受診記録に加えて1回以上の中毒を示唆する ICD-9コード965, 977またはICD-9のsupplemental sectionにある外傷、中毒のexternal causeに関するE codes	1病院の研究なのでリンケージは必要なかったと思われる	リサーチアシスタント（複数）によるチャートレビュー	1,215例全例	PPV/sensitivity			sensitivityの求め方には無理がある（全てのアセトアミノフェン関連の救急受診が今回のInclusion criteriaで捕まってくると言う仮定を置いて、all possible caseを分母として計算している） 患者特定に複数のE-codesを含むICD-9 codesを使っているが、その全てについて（単独での）感度、PPVをもとめている。
89	2012	Goldberg D, Lewis J, Halpern S, Weiner M, Lo Re V,3rd. Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database. Pharmacoepidemiol Drug Saf 2012 Jul;21(7):765-769.	末期肝不全	University of Pennsylvania Health System (UPHS)	アメリカ	1997-2011	2	no	chronic liver disease, cirrhosis, hepatic decompensation, or conditions associated with ESLD (e.g., jaundice, coagulopathy)などESLD（末期肝不全）を示唆するICD-9コードを組み合わせたAlgorithm1から3	2病院なので、各病院の中でデータがリンケージできていたと思われる	チャートの中に肝硬変または hepatic decompensation event (ascites, hepatorenal syndrome [HRS], spontaneous bacterial peritonitis [SBP], hepatic encephalopathy [HE], or gastric/esophageal variceal bleeding) の確定診断があるか	今回対象の病名（chronic liver disease, cirrhosis, hepatic decompensation, またはjaundice, coagulopathyなどESLD（末期肝不全）を示唆するICD-9コード）がついた人（元は何人か不明）から300人ランダムサンプリング、実際には244人	PPV	PPVが80%以上のアルゴリズムを探すことが目標とあらかじめ書いている 2回以上 chronic liver diseaseの診断が入っているものを入れるとPPVが上昇	特になし	
90	2015	Esposito DB, Lanes S, Donneyong M, Holick CN, Lasky JA, Lederer D, et al. Idiopathic Pulmonary Fibrosis in United States Automated Claims. Incidence, Prevalence, and Algorithm Validation. Am J Respir Crit Care Med 2015 Nov 15;192(10):1200-1207.	特発性肺線維症	HealthCore Integrated Research Database (HIRD)（入院、処置、薬歴、healthcare useを含むデータベース）	米国 HealthCore Integrated Research Database (HIRD)（全米の商用医療保険に加入している人のデータベース）				ICD-9-CMを用いた広義のIPF ICD-9-CMを用いた狭義のIPF（広義の定義でさらに、診断した医師、診断のタイミング、検査や移植の実施があった） IPF Score Algorithm（ロジスティック回帰により求めた）	特に触れられていないが、単一HMO内なので可能と考えられる	Unstructured review of medical records。主治医（Treating clinicians）のIPF Yesとの判断。さらにTwo primary adjudicator (expert)のIPF Yesとの判断	2006.1から2012.9までの連続6か月以上に商用医療保険に加入していた50-100歳の人のレセプト情報	PPV			HealthCore Integral Research Databaseで、IPFのBroad definition Yesの50例からのPPVを求め、Broad definitionを満たす4598例からIncidence/Prevalenceを求めPPVの値で補正。Broad definitionのSensitivity=100%が満たされていないとこの計算法は正しくないが、Sensitivityは求められていない。
91	2004	Verkooijen HM, Fioretta G, Chappuis PO, Vlastos G, Sappino AP, Benhamou S, et al. Set-up of a population-based familial breast cancer registry in Geneva, Switzerland: validation of first results. Ann Oncol 2004 Feb;15(2):350-353.	乳癌、卵巣癌患者のFDRの乳癌既往	The Geneva Familial Breast Cancer Registry	スイス ジュネーブ	1990～1999		yes	Family Breast Cancer RegistryのFirst-degree relative（FDR、母、娘、姉、妹）の乳がん・卵巣がんの既往に関する情報	州のpopulation officeを介して、Familial Breast Cancer Registryで特定されたFDRに関する情報とがん登録（Cancer Registry database）をリンケージした。 注：Familial Breast Cancer Registryは家族歴の情報を含む1990年以後のデータを蓄積した新しいDB。Cancer Registry databaseは1970年以後のデータを含む古いDB、	Cancer registryにおけるFDRの乳がん・卵巣がんに関する記録	家族歴を持つ119例および家族歴を持たない患者からのランダムサンプリングで得られたFDR310例を使用した。	PPV/NPV/Sensitivity/Specificity			SensitivityとSpecificityについては、119例の乳がんの家族歴ありの患者のFDRと100例の乳がんの家族歴のない患者のランダムサンプル100例（合計219例）のFDR310人が全populationであるかのように扱って計算しており、正しいとは思われない。
92	2013	Kim SC, Gillet VG, Feldman S, Lii H, Toh S, Brown JS, et al. Validation of claims-based algorithms for identification of high-grade cervical dysplasia and cervical cancer. Pharmacoepidemiol Drug Saf 2013 Nov;22(11):1239-1244.	高悪性度子宮頸部異形成と子宮頸がん	アルゴリズム作成にはbilling data from the Partners' Research Patient Data Registry (RPDR). 外部バリデーションには the Harvard Pilgrim Health Care's(HPHC) administrative claims database linked to the EMR of Harvard Vanguard Medical Associates (HVMA)	アメリカ	2007-2009	不明	わからない（Partner's の方はNoでよいがHPHCはpopulation-basedと言ってよいか。アルゴリズムの作成のDBとexternal validationのDBが異なるので判断困難）	女性20－60歳で外来のデータにabnormal Pap smear tests (ICD-9 code: 795.0x, 622.1x,616.9, 233.x, 180.x, and V73.81)のコードがある。この診断コードと処置の組合せによる4つのアルゴリズム。最終的にいいとされたのは2回の診断コードに加えて1回のprocedure code	記載なし	組織または病理検査結果	少なくとも1回Pap smearに異常があった24,426の女性全員	PPV	1個だけの診断名だと PPV of 57.1% (95%CI 54.7-59.5%). 2個の診断名だと PPV increased to 60.2% (95%CI 53.9-66.1%). 2個の診断名と a procedure code（初回の診断から一カ月以内）だと PPV of 80.7% (95%CI 73.6-86.2%).		アルゴリズムを作ったデータベースと、バリデーションしたデータベースが違う（その結果、PPVは同じくらいだったから、作ったアルゴリズムはgenerarizabilityがあると主張している） 10-100%の感度、99.47-99.93%の特異度を仮定して、hypothetical dataを使って、リスク比、リスク差に与える影響を検討している。
93	2016	Lee SY, Chung SE, Kim DW, Eun SH, Kang HC, Cho YW, et al. Estimating the Prevalence of Treated Epilepsy Using Administrative Health Data and Its Validity: ESSENCE Study. J Clin Neurol 2016 Oct;12(4):434-440.	てんかん	Health Insurance Review and Assessment service (HIRA)	韓国	2009	43	Yes AEDの処方が1度以上あった者でrepresentative health institutesのサンプル(に含まれていた6774人)がバリデーション研究の対象	AED処方が1度以上あった者におけるepilepsy/seizureのICD10コード	記載されていないがHIRAには個人を特定する情報が含まれており、それをそのまま利用したと考えられる。	Medical record survey:訓練を受けたhealth record administratorsによって抽出された情報をboard of epileptologistsが評価（中央に集めたと思われる）	AEDを2009年に処方されていた1,119,360人からのMultistage cluster sampling 43病院で6,774人。 2000人は診断名+抗てんかん薬、4000人は抗てんかん薬の処方のみの記録、になるように見積もっていた。パイロット研究から推定。 Health record administratorsが抽出。Epileptologistsのboardが判断。	抗てんかん薬+HIRAでepilepsyを示唆する診断名ありの者からのサンプル2374人から求めた"PPV"と抗てんかん薬+HIRAでepilepsyを示唆する診断名なしの4400人から求めた"PPV"。	ケースの定義や年齢性別によってレセプトデータの診断名の妥当性は変わる。またてんかん薬で治療されている患者が必ずしもてんかんや発作の病名がついているわけではない。	未治療のてんかん患者は含まれない。他の費用でカバーされている患者の情報は含まれない。	AEDを一度以上処方された患者群における研究であることを明確にし、感度、特異度、PPV、NPVを示すべきだったのでは？ 最終的には韓国におけるPrevalenceを（正しく）計算しているが、PPVだけだと、Prevalenceは求まらないと感じられてしまう。 また、薬剤処方がなかったてんかん患者も評価することがより望ましかったと考えられる。
94	2015	Regan AK, Mak DB, Moore HC, Tracey L, Saker R, Jones C, et al. Surveillance of antenatal influenza vaccination: validity of current systems and recommendations for improvement. BMC Public Health 2015 Nov 23;15:1155-015-2234-z.	妊婦のインフルエンザ予防接種	①Western Australian Influenza Vaccination Database (WAAIVD)（政府による州全域の出生前インフルエンザ接種のDB）＜対象者：期間中に出産した563人＞ (2)Perth（州都）にある公立の一病院（antenatal hospital ）のdatabase ＜①の対象者のうち該当病院で出産した201人＞ (3) 州都北部の3つの私立の産科病院を管理するprivate health service のdatabase＜①の対象者のうち該当サービスを利用して出産した211人＞.	オーストラリア西オーストラリア州	2013年4月-10月	1	Yes	2013年妊娠時のインフルエンザワクチン接種	Western Australia's statutory Midwives Notifications Systemからランダムに抽出した対象者とGSの情報をリンケージした： ①WAAIVDとのリンケージについては論文に明確な記載がないが、WAAIVDには氏名を含む個人情報が含まれていると考えられる。 ②と③については電話インタビューで、出産した病院の情報を得て、該当の病院のデータベースのうち女性の氏名と生年月日が一致する者の情報を得た	（WAAIVDを管理する州のDepartment Healthが実施した）電話インタビューによるセルフレポート	西オーストラリア州助産婦お知らせシステム（Western Australian's statutory Midwives Notifiacations System）からランダムに抽出された、妊娠中に2013年にインフルエンザワクチンを接種し、2013年4月から10月に出産した女性831例のうち、協力の得られた563人	ワクチン接種有無の感度、特異度、陽性的中度、陰性的中度	この研究から、現存データベースではワクチン接種状況を知るための感度が低いことがわかった。出生前のインフルエンザワクチン接種を記録するシステムがあるとよい。	職場や新興のサービスプロバイダーの元で接種している場合、データベース上で拾えない。 女性の氏名や生年月日でデータベースからデータ照合するのは不十分かもしれない。	
95	2015	Brixval CS, Thygesen LC, Johansen NR, Rorbye C, Weber T, Due P, et al. Validity of a hospital-based obstetric register using medical records as reference. Clin Epidemiol 2015 Nov 26;7:509-515.	出産および出産に関連する診断、薬、診療行為	産科データベース（Obstetric database）	デンマーク	2013年	1	population-basedではない。一病院(Hvidovre hospital)のdataを用いた評価。 評価されたObstetric registerは首都の最大の出産数の3つの病院からのデータを集めたもの。	出産およびlive birthに関連する会陰裂傷、Apgar Score、oxytocinの使用など吸引分娩など11項目の記述	出産数に関する、Obstetric databaseとNational Patient Register（gold standard）については"unique personal identification number"。11項目については言及ないが、一病院内での研究であり、リンケージは不要だったと考えられる。	National Parient Register（出産数）とmedical recordのチャートレビュー（11項目）	出産については2013年の全出産,11項目は産科データベースから（Hvidovre hospitalでの出産に関する例を）ランダムに抽出した250例	出産については数（6718vs 6717）自体を示し（指標自体は示していないが）「ほとんど完全に一致」と評価。11変数については感度、特異度、PPV、NPV。2人の評価者間のκ	産科の介入やアウトカムのモニタリングに有用なデータベースである。	特になし	

【付録3文献リスト】

96	2014	Widdifield J, Bombardier C, Bernatsky S, Paterson JM, Green D, Young J, et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. BMC Musculoskelet Disord 2014 Jun 23;15:216-2474-15-216	関節リウマチ	Administrative databases ①Ontario Health Insurance Plan (OHIP)の physician billing ②Canadian Institute for Health Information Discharge Abstract Database (CIHI-DAD) ③National Ambulatory Care Reporting System (NACRS)	カナダ オンタリオ	1991年4月1日から2011年3月31日	（都市部と田舎を含む）83のFamily physiciansが寄与するEMRデータベース（EMRALD）でadministrative dataによるRAの評価を実施した。	yes	28のアルゴリズム Algorithms were derived using combinations of physician billing diagnosis (ICD 714), primary and secondary hospital discharge diagnosis (ICD9 714; ICD10 M05-M06), prescription drug claims (for DMARDs, biologics, and glucorticosteroids), by varying windows between diagnosis codes or the period in which diagnosis codes appeared, and whether the services were rendered by a musculoskeletal specialist.	encrypted health insurance numbers	5人の訓練されたレビュワーによるChart review 専門家の診断あるいは、プライマリケア医のエビデンス（検査結果など）付き診断をRA症例とした。	20歳以上のランダムサンプル7500例と65歳以上の追加的ランダムアンプル2000例。前者から69例、後者から38例、合計107例のRA患者が特定された。	感度、特異度、陽性的中度、陰性的中度 ほかにintra- inter-raterの一致度をみるために10%サンプルについて複数のレビューを行いκを算出	最終的なアルゴリズムを決める基準（PPVが高いことを一義とし、感度と特異度も高い指標を選択する）をあらかじめもうけている	かかりつけ医に定期受診していない人には一般化できない	
97	2006	Tairou F, De Wals P, Bastide A. Validity of death and stillbirth certificates and hospital discharge summaries for the identification of neural tube defects in Quebec City. Chronic Dis Can 2006;27(3):120-124.	胎児の神経管欠損 無脳症、脳脊髄髄膜炎、脳脊髄炎または脳脊髄炎	stillbirth certificates (SBC)、infant death certificates (IDC), computerized hospital discharge summaries (Med-Echo)および computerized results of prenatal ultrasound examinations (Res-Echo)を使用した。	カナダ ケベック州	1993～2002	1病院	no	ICD-9を使用した。ただし、2000年以降、死産に対してはICD-10を使用した。	リンケージは実施されていない。	Medical notesをレビューした。	Infant death certificates (IDC), Med-Echo, Res-Echoのいずれかに NTDの記載があり真と判断された173ケース	PPV/Sensitivity			Sensitivityはinfant death certificates (IDC), Med-Echo, Res-Echoのいずれかに NTDの記載があり真と判断された173ケースが真のケースの全てであると仮定した値。この173ケースのうちそれぞれのデータソースがもつ感度を計算している。"all possible cases"のアプローチとしてよい。
98	2014	Rosenman M, He J, Martin J, Nutakki K, Eckert G, Lane K, et al. Database queries for hospitalizations for acute congestive heart failure: flexible methods and validation based on set theory. J Am Med Inform Assoc 2014 Mar-Apr;21(2):345-352.	急性心不全	Indiana Network for Patient Care (INPC)	アメリカ（インディアナ州）	2002-2011	2	no	少なくとも一個の急性心不全を示唆するICD-9コード（428.xx, 402.x1, 404.x1, 404.x3, or 398.91）またはBNP＞500pg/ml 具体的なコードとBNPとエコー所見でベン図を作ってそれぞれに当てはまる人達をバリデーションしている	記載なし（データをくっつけてくれるa master person indexという人をやとったと書いてある）	チャートレビュー（心不全専門の循環器科に相談してクライテリアを作っておいた）	CHFのICD-9をもつ66942人からサンプルした810人とBNP>500 pg/mLの12149人からサンプリングした98人。	PPV/sensitivity			ベン図を使っためずらしい研究 複雑だが、診断コード、BNF、echocardiogram（の組み合わせ）に該当するpossible casesから真のケースを見出す"all possible cases"の方法。組み合わせごとに異なるサンプリング法をとっている。 PPVとSensitivityの点推定に関する式が正しいのはOKだが、varianceについては？（複雑な式が与えられている。）
99	1996	Gabriel SE, Crowson CS, O'Fallon WM. A mathematical model that improves the validity of osteoarthritis diagnoses obtained from a computerized diagnostic database. J Clin Epidemiol 1996 Sep;49(9):1025-1029.	osteoarthritis	Olmsted郡のEMR	アメリカ ミネソタ州 Olmsted 郡（county）	1975-1987		Yes、Olmsted郡の全患者をカバーするEMRのシステムを使っている。	性、年齢、診断コード、初回診断からの経過年数を用いてrecursive partitioning modelによるosteoarthritisの定義	同一のEMRを用いた数学モデルによるアウトカム定義をChart Reviewで評価	medical record review clinical diagnosisに加え、X線に関するレポート、当該関節の痛みなどで判断。	Olmsted Country のEMRにOsteoarthritisの診断をもつ者から400例をランダムサンプル。	Osteoarthritisの診断をもち検討できた387人の"population"における感度、特異度、PPV、NPV	特に記述なし	特に記述なし	アウトカム定義にrecursive partitioning modelを用いた初期のバリデーション研究
100	2013	Goldberg DS, Lewis JD, Halpern SD, Weiner MG, Lo Re V,3rd. Validation of a coding algorithm to identify patients with hepatocellular carcinoma in an administrative database. Pharmacoepidemiol Drug Saf 2013 Jan;22(1):103-107.	肝細胞がん	Pennsylvania Integrated Clinical and Administrative Research Database (PICARD)	アメリカ	1997-2011	2	no	2回の慢性肝炎または肝硬変のICD-9コードに加え、2回の肝細胞がんのICD-9コード	2病院なので、各病院の中でデータがリンケージできていたと思われる	臨床上確定された肝細胞がん（造影CTまたはMRIによる画像診断または組織のクライテリア）	algorithmを満たす120例（実際に調査できたのは101例）、そのほか、肝移植のために検査された6000例からランダムに選択したTrue positive60例、True Negative 60例を用いて、sensitivityとspecofocityも求めようとしている	PPV、sensitivity、specificity	なし	なし	2病院の調査なのでしかたないが、Source Populationが明確でない。PPVはOK。Randomにサンプルされた60名のTrue positiveケースから感度を91.7%(55/60)として求めているが、True positiveの例を"random"に選択する過程が不明。また、Randomにサンプルされた60名からSpecificityを求めている。60名中1名がFalse Positiveで特異度は59/60=98.3%としているが、False positiveの例を"random"に選択する過程が不明。
101	2015	Hure AJ, Chojenta CL, Powers JR, Byles JE, Loxton D. Validity and reliability of stillbirth data using linked self-reported and administrative datasets. J Epidemiol 2015;25(1):30-37.	死産	1つの自己報告（Australian Longitudinal Study of Women's Health (ALSWH)の14,247人のコホートメンバーからの郵送）データと、3つのデータセット（Midwives Data Collection (MDC)、Admitted Patient Data Collection (APDC)、Perinatal Death Review Database (PDRD)）	オーストラリア	2011年	記載なし	yes	各データセットの死産の記録	ヘルスレコードリンケージセンター（Centre for Health Record Linkage (CHeReL)）にて実施 Provabilistic matching	4つのデータセットのうち、2つ以上のデータセットで記録が一致した場合（2/2、2/3、3/4以上の一致）	1996年から2009年に死産であった女性41例 53の死産のうち49例	感度，特異度，PPV、NPV	自己報告データ（Self-reported data）は感度は高い（100%）が特異度は低い（30%）。即ち、死産を経験した女性は全てその旨回答するが、死産は「over-reporting」される。ALSWHの誤った報告の約半分はlongitudinal data.のinconsistencyを見出すことにより除外可能。	自己報告データの正確性が低い	the Centre for Health Record Linkage (CHeReL)の機能が参考になる。日本の「代理機関」のようにデータを結合し匿名化データを作成するのではなく、それぞれのデータ源の一致を調べ、それを各データ源に戻す作業。
102	2009	Joseph KS, Fahey J, Canadian Perinatal Surveillance System. Validation of perinatal data in the Discharge Abstract Database of the Canadian Institute for Health Information. Chronic Dis Can 2009;29(3):96-100	周産期の診断や治療	Canadian Institute for Health Information (CIHI) hospitalization database のDischarge Abstract Database (DAD)	カナダ ノバスコシア州	2002	1地域	yes カナダ1地域のpopulation-basedの研究	疾患についてはICD-10-CAを用い、Intervention/procedureについてはCanadian Classification for Health Interventions(CCI)を用いた。	不明	Nova Scotia Atlee Perinatal Database (NSAPD)の情報を用いた。	6194例の母親、6315例の新生児を対象とした。	Sensitivity/Specificity	より重症だと（more severe forms of the disease、例：挿管を要する呼吸困難）では正確性（感度）がより高い。	ノバスコシア州以外には一般化できない。	同一の人がNANSPとDADに入力することが多く、一致の割合が過大評価されている可能性がある。
103	2015	Coleman N, Halas G, Peeler W, Casaclang N, Williamson T, Katz A. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. BMC Fam Pract 2015 Feb 5;16:11-015-0223-z	慢性疾患（COPD、Depression、Diabetes、Hypertension、Osteoarthiritis）	Manitoba Primary Care Research Network (MaPCReN)（カナダプライマリケアセンチネル調査ネットワーク（CPCSSN）の一つ）	カナダ	2011年12月31日以前。	3	Yes	慢性疾患定義アルゴリズム（糖尿病、高血圧、うつ、変形性関節症、COPD）に関するCPCSSN 2011 diagnostic algorithms（主にclaimのICD-9、EMRのproblem list上のテキストで定義される）	リンケージはしていない（EMRデータベースの診断コードなどを同一EMRのChart Reviewで評価）	チャートレビュー	1つ以上の慢性疾患を有する60歳以上からランダム抽出した403例	感度、特異度、陽性的中度	検査値などの客観的指標を用いるexternal validationではなく（カルテの記載とICD-9などとの一致性を調べる）Internal validationにフォーカスした。	サンプルサイズが小さい	False Positive（Tables 3）、False Negative(Table 4)の特徴の記述があり、Case detectionの改良やdata qualityの改善に役に立つとの記述が（Disucssionに）ある。
104	2016	Cave AJ, Davey C, Ahmadi E, Drummond N, Fuentes S, Kazemi-Bajestani SM, et al. Development of a validated algorithm for the diagnosis of paediatric asthma in electronic medical records. NPJ Prim Care Respir Med 2016 Nov 24;26:16085.	小児喘息	EMRのDBを評価 Canadian Primary Care Sentinel Surveillance Network (CPCSSN) database（primary care抽出され"clean"にされたEMRのDB）のうちSouthern Alberta Primary Care Research Network (SAPCReN)	カナダ Alberta州	2015年まで	地域全体をカバー	yes	喘息患者を定義するアルゴリズム： ICD-9コードと、EMRからの文字情報（請求書情報（billing）、診断名、診療録、処方薬） の組み合わせ	リンクはされていない。 同一のEMRから抽出したデータについてケースを定義するアルゴリズムを専門家の判断によって評価している。	2名の医師による判断	1000例のランダムサンプル（1-17歳）	PPV/NPV/Sensitivity/Specificity 2名の専門家の"any discussion"の前一致に関するκを示すとともに、不一致のものについては第3の評価者も含めdiscussionにより一致に達した	小児喘息の case definition and algorithmは、CPCSSNプライマリケアの研究に適している	本研究は、元のpatient chart から抽出されたデータの変数や情報の質に依存している。Patient chartに含まれる情報が不足している場合や誤っている場合、ケースを抽出するアルゴリズムや、2名の医師によるチェックの際に、小児喘息が、誤分類または過小評価されている可能性がある。	
105	2016	Abraha I, Serraino D, Giovannini G, Stracci F, Casucci P, Alessandrini G, et al. Validity of ICD-9-CM codes for breast, lung and colorectal cancers in three Italian administrative healthcare databases: a diagnostic accuracy study protocol. BMJ Open 2016 Mar 25;6(3):e010547-2015-010547.	乳がん、肺がん、大腸がん	イタリアの3つの地域におけるadministrative database。 Umbria Region (910 000 residents), LocalHealth Unit 3 of Napoli (1 170 000 residents) and Friuli-Venezia Giulia Region (1 227 000 residents)	イタリア	2012-2014		Yes, Caseとnon-caseのランダムサンプル。	ICD-9-CMによる定義。 (1) 233.0 and (2)174.x for breast cancer; (3) 162.x for lung cancer; (4) 153.x for colon cancer (5) 154.0-154.1 and 154.8for rectal cancer. 5年間該当コードがない"incident cases"のみを把握	記載はない。administrative databaseに患者情報が含まれている？	Medical chart review (1) 胸、肺、結腸、直腸における a primary nodular lesion（イメージングか内視鏡か） (2)、一次もしくは転移性のがんについて、細胞学的もしくは病理上の記載	各Unit(対象となる3つの地域のそれぞれ）ごとに1120chartsを評価する。 5つのICD9コードをもつ者各130、持たない者94 合計224（224X5=1120）	感度と特異度。κを使って評価者間の一致率を上げることを意図。			5種類のICD9をもつ130例ともたない94例から感度と特異度を求めるとしているが、その方法が不明。特異度の期待値90%としているが、94例の全てがnegativeの可能性が高いのではないか？

【付録3文献リスト】

付録 4: バリデーション研究におけるサンプルサイズ

[1]陽性的中度（PPV）のみを求めるバリデーション研究

一般にアウトカム指標の定義を満たすn例のサンプルのカルテレビューを行い a 例の真のケースが見出されれば、PPV は

$$\mathrm{PPV} = \frac{a}{n} \quad (1)$$

であり、その 95%信頼区間（$95\%\mathrm{CI}_{\mathrm{PPV}}$）

$$95\%\mathrm{CI}_{PPV} = \mathrm{PPV} \pm \delta_{PPV} \quad (2)$$

のδ_{PPV}（絶対精度、absolute precision)は二項分布の正規近似によって

$$\delta_{PPV} = 1.96\sqrt{\frac{PPV(1-PPV)}{n}} \quad (3)$$

と与えられる。δ_{PPV} の値が最大であるのは PPV=0.5 の場合であり、逆に、絶対精度をあるδ_{PPV} の値以下にするための n の大きさは

$$\mathrm{n} = \left(\frac{0.98}{\delta_{PPV}}\right)^2 \sim \frac{1}{{\delta_{PPV}}^2} \quad (3)$$

である[1]。PPV の 95%信頼区間を±0.1 以内としたければ 100 例（$1/0.1^2$）、±0.05 以内としたければ 400 例（$1/0.05^2$）が必要である。

[2]陽性的中度（PPV）、陰性的中度（NPV）、感度、特異度を求めるバリデーション研究

[2-1]単純なランダムサンプル

バリデーション研究を実施する集団全体から単純なランダムサンプルによって得た n 例のカルテレビューを行い、陽性的中度（PPV）、陰性的中度（NPV）、感度、特異度を本報告書の表 1 に示した式を用いて求める場合、95%信頼区間が最も広くなるのは通常は感度である。ある特定の絶対精度（δ_{sen}）で感度を求める場合には、上記式(3)と同様にサンプルに含まれる真のケース（表 1 の a+c）の大きさが$\frac{1}{{\delta_{sen}}^2}$となることが求められる。n 例の単純なランダムサンプルに含まれる真のケースの割合は、集団における真のケースの割合（有病割合、prevalence、以下 prev と略す）と同一とみなすことができ、n･prev の大きさが$\frac{1}{{\delta_{sen}}^2}$に等しいような n が必要である。すなわち必要なサンプルサイズは

$$n = \frac{1}{prev}\frac{1}{{\delta_{sen}}^2} \quad (4)$$

と与えられる。たとえば、真のケースを 1%含む集団からの単純なランダムサンプルで感度の 95%信頼区間を±0.1 以内としたければ 10,000 例（$(1/0.01)(1/0.1^2)$）のカルテレビューが必要である。実際、関節リウマチ（RA）に関して、「RA の診断による入院または 2 年以内に 3 回の外来診療における

RA の診断をもち、うち 1 例は専門家による」という定義の感度が±0.1（0.77、95%CI：0.67-0.87）であることを示したカナダで実施されたランダムサンプルによるバリデーション研究では、7,500 人のカルテレビューが実施されている[2]。この研究では当該定義による TP のケースは 53 例、FN は 16 例だったので、有病割合は 0.92%程度であったと考えられる。

[2-2]層別サンプル

バリデーション研究を実施する N 人の集団のうち、アウトカム指標"Yes"の者が N_1 人、"No"の者が N_0 人である場合に、これら N_1 人と N_0 人から異なる割合でサンプルを得る層別サンプル（stratified sampling）によるバリデーション研究も可能である。

ただし、この方法への過大な期待は禁物であり、層別サンプルを使って、感度をたとえば±0.1 で求めるのに必要なサンプルサイズは、上述の[2-1]の単純なランダムサンプルで感度を±0.1 で求める場合のサンプルサイズと通常大きくは変わらない。この方法が有用なのは、たとえば PPV については必要な精度の情報を得る一方、感度についても、絶対精度は若干低くても何らかの情報を得ておきたい、というような場合である。たとえばノルウェーにおける妊娠レジストリにおける子癇前症(preeclampsia)のバリデーション研究では、この疾患コードをもつ 3,500 例全例と疾患コードをもたない 75,311 例の 2.4%の 1,840 例のレビューが antenatal chart と退院時の診断コードをゴールドスタンダードとして実施されている。疾患コードをもつ 3,500 例のうち、真のケースは 2,936 例であり、PPV=0.84、その 95%信頼区間は（0.83, 0.85）であるが、感度については 0.43、その 95%信頼区間は（0.39, 0.48）であり、PPV については 95%信頼区間が±0.01 で求められたのに対し、感度の 95%信頼区間については±0.04 であった。

本 TF のメンバーなどによる最近の研究において、層別サンプルを用いたバリデーション研究では二段階で研究を実施することが推奨されている[4]。第一段階（Step I）においては[1]と同様にアウトカム定義を満たす者 N_1 人からランダム抽出した n_1 人のサンプルについてカルテレビューを実施し、真のケース a 人と非ケース b 人を得て($a+b=n_1$)、PPV（a/n_1）を推定する。次いで第二段階（Step II）において、以下の式で与えられるアウトカム定義を満たさない者 N_0 人からの n_0 人のサンプルについてカルテレビューを実施し、真のケース c 人と非ケース d 人を特定する($c+d=n_0$)。

$$n_0 = \frac{N_0}{N_1}\frac{1}{PPV}f* \qquad (5)$$

式(5)で f*の値は感度の絶対精度して期待される値"δ_{sen}"と"a"の 2 つによって決まり、本付録末尾に掲げた表にその値を示した。本付録末尾の表における"a"の値が、Step I において n_1 人に含まれる真のケースの数よりも多い場合には、アウトカム定義を満たす者について追加のカルテレビューを行う。たとえば、Step I で PPV の 95%CI を±0.1 で求めるために 100 人のカルテレビューを実施し、真のケースが 80 人（a=80, b=20）で PPV=0.8 と推定されたとする。この場合、得られた PPV の 95%信頼区間は二項分布の正規近似により 95%信頼区間（0.72-0.88）と計算され±0.1 の範囲内だが、本付録末尾の表を利用し、たとえば a については a=80 ではなく a=100 を用いるのであれば、a=100 が満たされるように、アウトカム定義を満たす者 25 名程度のカルテレビューを追加で実施する。本付録末尾の表の f*の算出法については元論文[4]を参照されたい。

n_1 例と n_0 例のカルテレビュー実施し、a、b、c、d の値を得た後の PPV と NPV および、その 95%

信頼区間の推定方法は、単純なランダムサンプルの場合と同様であり、標準的な教科書における割合（proportion）とその 95%信頼区間の推定方法を参照する。

また、感度（sen）と特異度(spe)は以下のように求めることができる。

$$\text{sen} = \text{PPV}\cdot N_1/[\ \text{PPV}\cdot N_1+(1-\text{NPV})N_0] \quad (6)$$

$$\text{spe} = \text{NPV}\cdot N_0/[(1-\text{PPV})N_1+\ \text{NPV}\cdot N_0] \quad (7)$$

その 95%信頼区間については、bootstrap 法を用いるか、以下の近似式[4]を用いる。

$$95\%\text{CI}_{sen} = \ \text{sen}\cdot\ \exp\left[\pm 1.96(1-\text{sen})\sqrt{\frac{1}{a}+\frac{1}{c}}\right] \quad (8)$$

$$95\%\text{CI}_{spe} = \ \text{spe}\cdot\exp\left[\pm 1.96(1-\text{spe})\sqrt{\frac{1}{b}+\frac{1}{d}}\right] \quad (9)$$

参考文献

1. McNeil D. Epidemiological research methods. John Wiley & Sons; 1996, 266 p.
2. Widdifield J, Bombardier C, Bernatsky S, et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. BMC Musculoskelet Disord. 2014;15:216.
3. Klungsøyr K, Harmon QE, Skard LB, et al. Validity of pre-eclampsia registration in the medical birth registry of norway for women participating in the norwegian mother and child cohort study, 1999-2010.Paediatr Perinat Epidemiol. 2014;28:362-71.
4. Kubota K, Iwagami M, Yamaguchi T. Designing the validation study with the stratified sampling. In submission.

表　f*　（式(5)）

δ_{sen}	a=50	a=60	a=70	a=80	a=90	a=100	a=120	a=150	a=200	a=300	a=500
0.10	80.46	69.26	63.58	60.14	57.84	56.18	53.96	52.01	50.26	48.70	47.56
0.11	58.25	52.69	49.59	47.60	46.22	45.21	43.81	42.55	41.40	40.35	39.57
0.12	45.32	42.11	40.22	38.96	38.07	37.40	36.46	35.60	34.80	34.06	33.51
0.13	36.80	34.77	33.52	32.67	32.05	31.59	30.93	30.32	29.75	29.21	28.80
0.14	30.75	29.38	28.50	27.90	27.46	27.13	26.65	26.20	25.77	25.37	25.06
0.15	26.25	25.26	24.63	24.19	23.87	23.61	23.26	22.91	22.59	22.28	22.04
0.16	22.76	22.04	21.56	21.23	20.98	20.79	20.51	20.25	19.99	19.75	19.56
0.17	19.99	19.44	19.08	18.82	18.63	18.48	18.26	18.05	17.85	17.65	17.50
0.18	17.75	17.32	17.03	16.83	16.67	16.56	16.38	16.21	16.05	15.89	15.77
0.19	15.90	15.56	15.33	15.16	15.04	14.94	14.80	14.66	14.53	14.40	14.30
0.20	14.35	14.07	13.89	13.75	13.65	13.57	13.45	13.34	13.23	13.12	13.04

アウトカム定義を満たす者からのランダムサンプルのカルテレビューにより真のケースが a 人特定されていることを前提に、アウトカム定義を満たさない者から、本表に示された値(f*)を用いて得られる（(N_0/N_1)/PPV f*）のサイズのサンプルをランダムサンプルし、カルテレビューを実施すれば、感度の 95%信頼区間は推定値±δ_{sen} 以内である。詳しくは元論文[4]参照。

バリデーション研究のプロトコールや論文中に記載すべき情報一覧（チェックリスト）				
チェック項目	内容	確認結果		
		はい	いいえ	不明
リサーチクエスチョンとプロトコール	リサーチクエスチョン（バリデーション研究後の使用目的も含む）が明記されているか。			
	ガイドライン(Guidelines for Good Pharmacoepidemiology Practices など)にそってプロトコールが作成されているか。			
倫理審査と研究計画の登録	研究実施に関する倫理審査委員会による承認について記載されているか。			
	研究計画の事前登録(UMIN 臨床試験登録システムなど)がなされているか。			
パイロット研究	少数例を対象としたパイロット研究が予定されているか。			
(1)利用する医療情報データベースの種類の理解	データベースの種類（入院・外来の診療報酬請求データ・退院時記録・電子カルテ・疾患レジストリー、またはそれらの組み合わせ）が明記されているか。			
(2)バリデーション研究のセッティングの理解	対象の地域が定義されているか。			
	データの期間が明記されているか。			
	調査された医療機関の数が明記されているか。			
	医療機関の特徴（大学病院、救急指定病院、がん拠点病院、受診患者数、病床数、等）が明記されているか。			
	研究対象集団の特徴（年齢・性別構成、重症度の分布、等）が明記されているか。			
	バリデーション研究の対象集団と、利用するデータベースの想定する"population"の差異についての検討・考察がされているか。			
(3)アウトカムの定義	どの情報を使うか（傷病名のみ、薬剤名のみ、処置名のみ、検査結果のみ、またはその組み合わせによるアルゴリズム、など）。			
	傷病名をアウトカム定義に使う場合、傷病名コードの種類（傷病名マスター、ICD-10 コード、等）、選択した傷病名コードのリスト、傷病名入力の期間・回数に関する条件（例：異なる月に連続して 2 回、半年以内に 2 回、等）の記述があるか。			

(4)リンケージの有無・方法の確認	リンケージ(照合)を行うか(注: 各医療機関の中で保存しているデータを比べる作業はリンケージとはみなさない)。			
	リンケージ(照合)を行った場合、リンケージのタイプ(deterministic または probabilistic)、使用した因子(Social security number、年齢・性別・生年月日・居住地域、等)、リンケージが出来なかった人の数や割合などが記載されているか。			
	Probabilistic リンケージの場合、リンケージの質について記載されているか。			
(5)ゴールドスタンダードの定義	ゴールドスタンダードの種類(疾患レジストリー、検査結果、カルテレビュー、等)が明記されているか。			
	疾患レジストリーの場合、それ自体の妥当性に関する研究が過去に実施されている場合はその引用、実施されていない場合は疾患レジストリーがゴールドスタンダードとして妥当であるとの判断が合理的である旨の記述があるか。			
	検査結果の場合、検査結果が正確に測定・標準化・記録されているかの記載があるか。			
	カルテレビューの場合、対象疾患の判断基準または調査票を明示し、評価者の専門性を記載し、2 人(あるいは少数)の評価者が別々に判断、評価者間の一致度を表す κ 係数を計算し、一致しなかった評価結果の取り扱いについて明記されているか。			
(6)サンプリング方法とサンプルサイズの設定(右のいずれかを選択する)	[A]全患者を対象にする場合:アウトカムとゴールドスタンダードの定義を満たす患者を、それぞれコンピューターを用いて簡単に同定できるかを確認する(できない場合はデータの標準化・電子化による作業の効率化が必要)。			
	[B]全患者からランダムサンプリングを行う場合:アウトカム定義を満たす集団(陽性のケース)、満たさない集団(陰性のケース)、ゴールドスタンダードの定義を満たす集団(真のケース)、満たさない集団(偽のケース)をそれぞれ最低 100 人(求める指標の 95%信頼区間が最大±10%に対応)ずつ確保できるようサンプル数を計算する(確保が難しい可能性が最も高い「真のケースが最低 100 人」になるように計算する)。			
	[C]アウトカム定義を満たす患者群からのみランダムサンプリングを行う場合:最低 100 例			

	が目安となる。			
	[D]アウトカム定義を満たす患者群と満たさない群からランダムサンプリングを行う場合：アウトカム定義を満たす患者群からは最低 100 例、アウトカム定義を満たさない患者群からは付録 4 の近似式などを参考にサンプルサイズを計算する。			
	[E]“all possible cases”を想定したサンプリングを行う場合：研究対象集団（全体またはランダムサンプリングした症例）の中で真のケースを多く含むことが期待されるサブグループを 2 種類以上特定し足し合わせることでサブグループの集合体を作る。複数のサブグループからランダムサンプリングをする際は、抽出率を同じ割合にして、最終的に同定される真のケースが最低 100 例以上になるよう必要サンプル数を見積もる。			
(7)妥当性を測定する指標の計算	妥当性を測定する指標として求めるものは何かが説明されているか。			
	[A]全患者を対象にした場合、または、[B]全患者からランダムサンプリングを行った場合：2×2 テーブルが作成され、感度・特異度・陽性的中度・陰性的中度が計算されているか。			
	[C]：アウトカム定義を満たす患者群からのみランダムサンプリングを行った場合：陽性的中度が計算されているか。			
	[D]アウトカム定義を満たす患者群と満たさない群からランダムサンプリングを行った場合：ウエイトを考慮した 2×2 テーブルが作成され、感度・特異度・陽性的中度・陰性的中度が計算されているか。			
	[E]“all possible cases”を想定したサンプリングを行った場合：得られた情報をもとに 2×2 テーブルが作成され、感度・特異度・陽性的中度・陰性的中度が計算されているか。			
(8)指標の閾値・利用法に関する考察	同じ研究セッティングで複数のアルゴリズムを比較する場合、著者らが考える最適なアルゴリズムについての考察がなされているか。			

1