

# 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

## 報告書説明会

### プログラム

14:00

開会の辞 (小出)

14:05

報告書の概要 (岩上)

14:20

教科書・ガイドライン (宮崎)

14:40

レビューした文献の紹介 (久保田)

15:00

バリデーション研究の方法 (岩上)

15:20-15:40

休憩

15:40

DPCデータとDPCレセプト (今井)

15:55

大規模コホート研究におけるバリデーション研究 (後藤)

16:10

北欧におけるレコードリンケージ (今井)

16:20

匿名加工(医療)情報とバリデーション研究 (小林)

16:35-17:25

質疑応答 (タスクフォースメンバー)

17:25

閉会の辞 (久保田)

# レビュー対象文献の選定

- 2016年12月にPubMedで以下の検索式で412文献ヒット
  - (((sensitivity[Title/Abstract]) OR predictive value[Title/Abstract]) AND database[Title/Abstract]) AND **valid**\*[Title]
- 更にタイトル・アブストラクトを見て、絞りこみ。
- 担当メンバー（久保田、岩上、佐藤、草間、中根）の手持ちの重要な論文を追加。
  - 例：The **accuracy** of Japanese claims data in identifying breast cancer cases.
  - claimsデータのバリデーション研究に絞るか？
    - **絞らない**：“どのような研究があるかザッと見ておきたい。”
  - プロトコール論文は？
    - **含める**
  - systemic reviewは？
    - **除外する**
- スカイプでdiscussion：8回実施(130→最終105文献)
  - 「抽出すべき項目」の大枠を早い段階で設定
    - 「抽出すべき項目」については、その後微調整

# 抽出すべき項目

- 利用するデータベース
- セッティング
  - 国・地域、期間、“population-based”か？
- アウトカム定義
- リンケージの種類、方法、目的
- ゴールドスタンダード
- サンプルング方法とサンプル数
- 妥当性（信頼性）の指標
- 指標の閾値、利用方法
- 著者らが挙げている問題点
- 備考

# 「利用するデータベース」

- 「バリデーション研究で評価した指標を用いた研究に利用するデータベース」（評価対象のデータベース）
  - ①診療報酬データ（claims data）
    - Medicaid/Medicare、カナダの医療費償還請求書、日本のレセプト
  - ②退院時記録（discharge data）
    - カナダ、デンマーク、フランス
  - ③外来の”administrative data”
    - カナダ National Ambulatory Care Reporting System (NACRS)
  - ④電子カルテ（Electronic Medical Record）
    - 米退役軍人DB、カナダ、オランダ、イタリア、ノルウェー、英国 THIN
  - ⑤疾患レジストリ
    - カナダ糖尿病レジストリ、英国National Hip Fracture DBなど
    - ゴールドスタンダードとして使われることもある
  - ⑥これらの組み合わせ
    - カナダ：外来の診療報酬DB+退院時記録、英国CPRD（旧GPRD）
  - ⑦薬剤疫学研究のためのデータベース
    - 欧州のEU-ADR

# セッティング 国・期間

- 国 :

- 主にカナダ、米国
- 英国・デンマーク・フランス・ノルウェー・オランダ・スペイン・イタリア
- 最近はアジアでも（台湾、韓国、日本）

- 期間 :

- 開始年は2000年以後。年々増加
- アジアの少数の研究は2010年前後から

# セッティング population-basedか？

- 「利用する（評価対象の）データベース」はおおむね”population-based”
  - ”population-based“：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（本TFにおける定義）
    - 例：地域住民、特定保険の被保険者、退役軍人
- 評価対象のDB自体がpopulation-basedでない場合もある（1病院のデータなど。後述。）

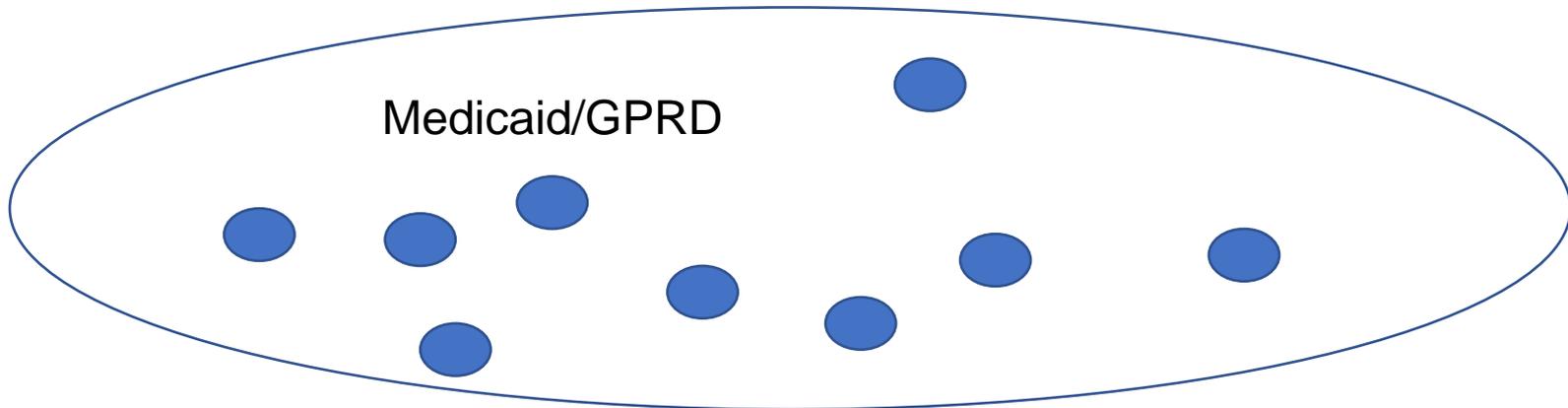
# セッティング：“Population-based”の研究

（付録3文献番号：例示。網羅はされていない。）

- ゴールドスタンダード（GS）がカルテレビューでない場合”population-based”であることが多い。
  - GS：疾患レジストリ
    - カナダの疾患登録（18, 47, 73）：administrative dataとregistry
    - 米国の州のがん登録（82）：Medicareとがん登録（SEER）
  - GS：処方DB
    - 妊娠糖尿病（35）：ノルウェー出生レジストリ（MBRN）と処方DB（NorPD）

# セッティング：“Population-based”の研究 GS：Chart Review 1（付録3文献番号 例示）

- 評価されるDBから特定の診断コードをもつ患者を直接（ランダム）サンプル。依頼状/調査票を送付。



- 例：
  - 米国32州の168病院に送付（3）
  - 韓国のような地域の43病院に調査依頼（93）
  - 英国GPに患者延べ951人分の質問票を発送（50）

# セッティング：“Population-based”の研究GS： Chart Review 2（付録3文献番号 例示。網羅はされていない。）

- 厳密なランダムサンプルではないが、評価対象のDBと性・年齢・疾患の分布などがほぼ同一と考えられるEMR DB/コホートを用いる。
- 例1：カナダ（例：オンタリオ州）のバリデーション研究：[地域の相当数のGPのEMRデータベースを用いたカルテレビュー]（8, 9, 44, 45, 75, 96）

外来のClaims data+入院患者のadministrative data

ElectronicMedical Record Administrative data  
Linked Database (EMRALD)

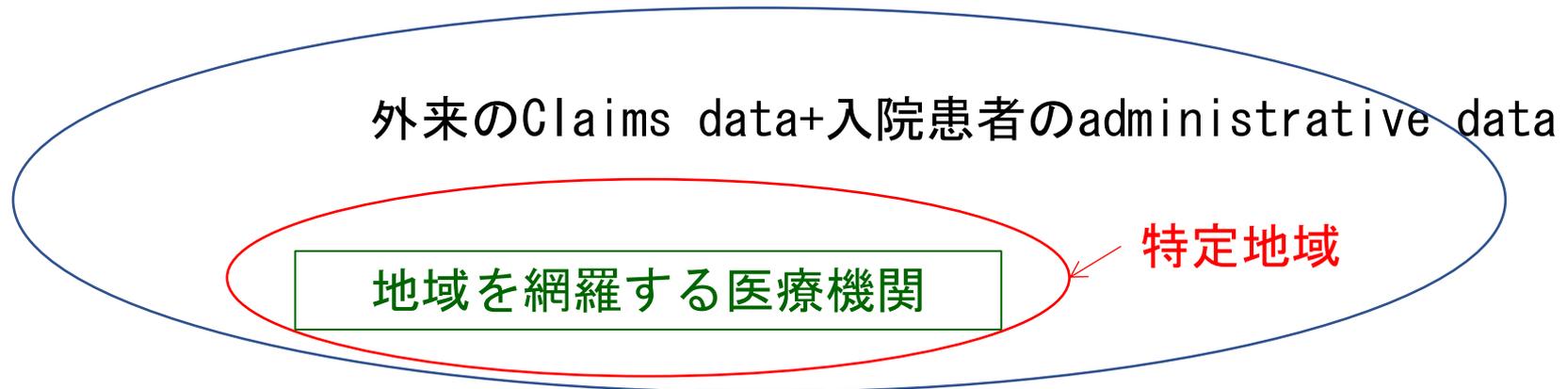
- 例2：米国MedicareのStrokeのバリデーション研究[当該疾患に関するコホートを利用]（2）

Medicareデータベース

reasons for geographic and racial  
differences in stroke (REGARDS) cohort

# セッティング：“Population-based”の研究 GS : Chart Review 3（付録3文献番号 例示。網羅はされていない。）

- 特定地域におけるバリデーション研究



- 当該地域をほぼカバーすると考えられる医療機関で実施
  - 1施設
    - Eastern Ontarioで小児腸重積を全てカバーする1病院(61)
    - 英国の1GPにおける研究（63）
  - 複数施設
    - コペンハーゲンのStrokeの10の専門病院（60）
- 当該地域をカバーするか否かを（文献からは）明確にしえないものも。
  - 当該病院がフランスDijon地域をかバーするか？が不明（17）

# セッティング：“Population-based”ではない研究（付録3文献番号 例示。網羅はされていない。）

- 評価対象DBの”population”を代表していない
  - カナダの退院サマリデータベースを3病院で評価(12)
  - カナダの疾患レジストリを1病院で評価(38)
  - フランスのdischarge databaseを1病院で評価(17)
  - 米国Medicareを1病院で評価(55)
  - スペインPrimary care databaseを6施設で評価(65)
  - デンマークの疾患レジストリを2病院で評価(68)
  - 台湾のNHIRDを1病院で評価(6, 7)
  - 日本のレセプトデータベースを1病院で評価(67)
- 評価対象のDBが”population-based”ではない
  - 米国の当該病院固有のDBを評価(56, 85-89)
- 特定の疾患/検査対象の患者集団(6, 18, 58, 79)

# アウトカムの定義

(付録3文献番号 例示。網羅はされていない。)

- ICD単独 (ICD-9、ICD-10、Read code) 多数
- 薬剤 (64, 65)、処置・手術 (66-68)、検査値 (15, 32)
- 疾患情報+受診回数
  - 例：2年以内に入院1回 or 外来3回 (71)
- 疾患情報+薬剤/手術(55, 58, 69)

# リンケージ（照合）

（付録3文献番号 例示。網羅はされていない。）

- タイミング
  - リンク済のDBを使用
    - Medicareの死亡情報
    - カナダの入院に関するadministrative dataと外来のclaims data
  - バリデーション研究の際にリンケージ
- リンケージの目的
  - ゴールドスタンダードによる評価を可能にする
  - よりよいアウトカム定義を得る
    - カナダの入院に関するadministrative dataと外来のclaims data
- リンケージの方法
  - deterministic: 1つまたは複数の識別子が完全に一致する者を1:1でリンケージ（多数）
  - probabilistic: 複数の識別子の一致度が一定の閾値以上であれば同一個人とみなす(17)
- リンケージに用いられた識別子
  - 国や地域で付与された健康情報を管理するための番号(11, 12, 37, 46, 49, 61, 73-80)
  - 社会保障番号(4, 39, 82)

# ゴールドスタンダード

(付録3文献番号 例示。網羅はされていない。)

- カルテレビュー (最多)
  - 専門家が直接レビュー：診断に専門的知識を要するなど困難
  - chart abstractorが (専門家の作成したチェック項目に従って) 実施
    - Screeningを実施し、最終的に専門家が判断する場合も(71)
- 疾患レジストリー
- 検査結果
  - 臨床検査値：急性(59)慢性(32)腎不全、糖尿病(73)、低Na血症(15)
  - 病理診断：子宮頸がん (92)
- 医師への質問票：英国primary care dataで多い
- その他：
  - 薬の使用 (糖尿病、喘息)
  - 住民の質問票への回答

# サンプリング方法とサンプル数（1）

（付録3文献番号 例示）

- [1]カルテレビューがゴールドスタンダード
  - [1-1]陽性的中度のみを求める
    - 「100例あれば±10%の精度」(54)
  - [1-2]感度・特異度・陽性/陰性的中度を求める
    - 一般にサンプル数は[1-1]より1~2桁大きくなる
    - ランダムサンプル
      - RA: 9,500例(96) Epilepsy: 7,500例(71) Stroke/TIA: 5,000例(45) 会陰裂傷: 25,761例(26)
    - アウトカム定義Y/Nの集団から異なる抽出率でサンプル
      - 妊娠高血圧(36)
        - 定義満たす：3,500/3,500 定義満たさない：1,840/75,311
        - 感度・特異度の信頼区間を”bootstrap”法で推定
  - “all possible cases “の方法
    - 多発性硬化症(75) 可能性の高い943例から247例を特定
    - 腸重積症例(61) 可能性の高い症例から565例を特定

# “All-possible cases”のアプローチ

(付録3文献番号)

## • “All-possible cases”

- カナダ多発性硬化症（MS）の研究（75）で使用された語。
- EMRALDからの9,500例のランダムサンプルによるRAのバリデーション研究（96）（2014公表）を実施したWiddifieldらが同時期（2015公表）にEMRALDを用いてMSのバリデーション研究（75）を実施。
  - 「MSは稀であり、ランダムサンプルでは、得られる症例数が少なすぎる」とし、RAのバリデーション研究とは異なる（“All-possible cases”の）アプローチを採用
- 3つのサブグループ（①EMRのcumulative patient profile (CPP)、②free text entries for MS-related phrases、③physician billing codes）のいずれかにMS患者が含まれることを前提に該当者943例のカルテレビューを実施しMSのケース247例を特定。
  - カルテレビューによるMSを該当集団の全ケースとみなす

## • 注意

- 「漏れ」が相当数予想される場合は“All-possible cases”は使用しない。
  - 感度が過大評価される。
- 「漏れ」が妥当性の指標に与える影響を評価する感度解析を実施。

## サンプリング方法とサンプル数（2）

（付録3文献番号 例示。網羅はされていない。）

- [2]調査票/インタビューがゴールドスタンダード
  - 陽性的中度のみを求める
  - 英国のGPへの質問票による調査では100-200例
    - 大腸がん(29)、膀胱がん(30)、乾癬性関節炎(31)、C型肝炎(33)
- 以下[3][4][5]のうち、参照が容易で、機械的に評価可能なものの症例数は数万~数十万例
- [3]電子カルテの診療録をゴールドスタンダード(9, 24)
- [4]疾患登録がゴールドスタンダード(49)
- [5]検査値/画像検査の結果がゴールドスタンダード(92)

# 妥当性の指標など

(付録3文献番号 例示。網羅はされていない。)

- 陽性的中度のみ
  - Mini-sentinel(53, 54)、THIN(30, 33)、GPRD(34)
- 感度、特異度、陽性的中度、陰性的中度(2, 8, 10-12, 15, 19, 21, 25-27, 32, 35, 36, 38, ----)
  - 4指標求めることが可能と考えられるが、一部のみを示しているもの
    - 感度と陽性的中度のみ(29, 46, 57, 63, 64, 66, 74, 77, 81, 83, 88, 97, 98, 102)
    - 感度、特異度、陽性的中度(37, 42, 43, 49, 58, 69, 72, 73, 100)
- 偽陰性・偽陽性となった理由などの詳細
  - 表などの形で表示したもの(4, 7, 8, 38, 46, 49, 51, 73, 74, 102, 103)

# 信頼性の指標

(付録3文献番号 例示。網羅はされていない。)

- 複数の人でカルテレビューを実施した時に一致度(inter-rater reliability)をKappa係数で評価(5,12,44,65,80,104,105)
- 同一の評価者の異なる時点での一致度(intra-rater reliability)をKappa係数で評価(9, 44, 71, 96)

# 指標の閾値/利用方法 (付録3文献番号 例示)

- 明確に基準を数値として示したものは少ない
  - 目標とする陽性的中度を示した論文
    - 80%以上(89)
    - 85%以上(80)
  - 得られた陽性的中度への評価
    - 70%は「低い」(67)
    - 67%未満は「低い」(81)
- 低い感度(49.8%)と高い陽性的中度 (88.4%) を示した定義への評価
  - 「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」(2)
  - 「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」(8)
- 定義の選択
  - PPVが高く、かつ感度、特異度も高い指標を選択した(9)

# 備考(レビュー担当者の感想など)

(付録3文献番号 例示)

- 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。(2)
- (Na値<135に対して) 感度1.8% (95% CI 1.7% to1.8%)は低すぎる(15)
- 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(75)
- 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。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。(87)
- 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。(103)