

# 医療ビッグデータ利用の 現状と課題

20181127

東京大学大学院医学系研究科

公共健康医学専攻臨床疫学・経済学

松居 宏樹

# 本日本話しする内容

## この講演の目標

- RWDとしての大規模医療データベース
  - DPC
  - NDB
  - 介護レセDB
- 研究利用における大規模医療データベースの現状と課題

# 本日本話しする内容

## 想定する対象者

- 公的な大規模医療データを利用したことがなく、今後利用を検討している方々

# 本日本話しする内容

## もくじ

- 大規模データベースを利用した研究活動
  - Administrative Claims Database とは？
  - NDBとはなにか
  - NDBのデータに含まれる情報
  - NDBの強みと弱点
- NDBデータの研究利用
  - 申請方法
- NDBデータの研究利用の課題

# 簡単な自己紹介

松居宏樹

東京大学大学院 臨床疫学・経済学教室 助教

## 主たる研究テーマ

大規模データベースを用いた臨床疫学研究

- DPCデータベース
- NDBデータベース
- 介護レセデータベース

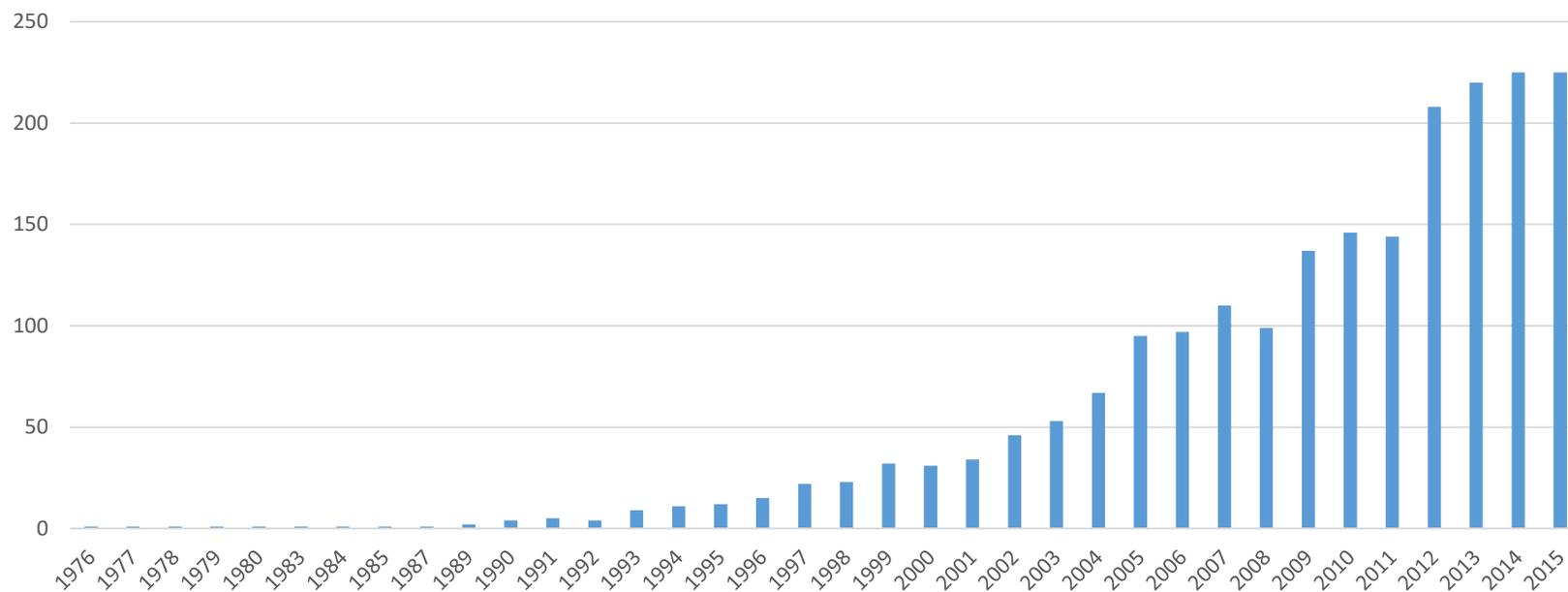
# Administrative Claims Database の研究利用

- 記述統計研究
  - 薬剤の使用状況・疾病の分布など
- 政策評価研究
  - 政策施行前後での患者受療行動の変化など
- Ecological 研究
  - 地域間の記述統計・相関解析など
- 治療やリスクファクターとアウトカムとの関連性（因果関係の検証）
  - 自然実験・操作変数などを用いた治療効果・政策効果の検証など

# Administrative Claims Database の研究利用

研究利用する試みは数多く行われている。

Administrative Claims Database でPubmed を検索した際のHit 数



検索日: 20151016.12:00

# 海外の大規模データ

## 海外の大規模臨床データベース

- 米国のCMS Medicare DataやNISデータ
- 英国のClinical Practice Research Datalink (CPRD)
- 台湾のNational Health Insurance Research Database
- 韓国のHealth Insurance Review Assessment Service

…等々

NDBの研究利用の現状

# Administrative Claims Database の比較

|          | 企業が提供する<br>レセプトデータ | NDB                    |                | NIS Data<br>(米国) | CMS data<br>(米国)   |
|----------|--------------------|------------------------|----------------|------------------|--------------------|
|          |                    | レセプト                   | 特定健診情報         |                  |                    |
| 収集元      | 国内の保険者             | 審査支払機関<br>社会保険診療報酬支払基金 |                | 収集に同意した病院        | Medicare, Medicaid |
| 収集対象     | 特定の健保加入者           | <b>全国民</b>             | <b>特定健診対象者</b> | 入院患者             | 加入者                |
| データ規模    | 数百万人               | <b>1億2000万人</b>        | <b>2400万人</b>  | 700万人            | 5,000万人 (2012)     |
| 個人の縦断的観察 | 可                  | <b>可</b>               | <b>可</b>       | 不可               | 可                  |
| 患者の重症度   | なし                 | <b>なし</b>              | <b>検査結果あり</b>  | あり (一部)          | なし                 |
| 医療機関匿名化  | 匿名化                | 匿名化                    | 匿名化            | 匿名化              | Facility Number    |
| 患者匿名化    | 匿名化                | 匿名化                    | 匿名化            | 匿名化              | SSN                |

# Administrative Claims Database の比較

|               | 市販レセプトDB           | NDB                         |                             | NIS Data (米国) | CMS data (米国)                   |
|---------------|--------------------|-----------------------------|-----------------------------|---------------|---------------------------------|
|               |                    | レセプト                        | 特定健診情報                      |               |                                 |
| 提供方法          | データを研究者に提供         | ・データを研究者に提供<br>・アクセスポイントを提供 | ・データを研究者に提供<br>・アクセスポイントを提供 | ・データを研究者に提供   | ・データを研究者に提供<br>・VPNオンラインアクセスを提供 |
| オプトアウト        | 可                  | 不可                          | 不可                          | 可             | 不可?                             |
| 利用料金          | 有料                 | 無料                          | 無料                          | 有料            | 有料                              |
| 他DBとのリンク      | 不可<br>※特定健診情報とは突合可 | 不可<br>※特定健診情報とは限定的に突合可      | 不可<br>※レセ情報とは限定的に突合可        | 不可            | 可                               |
| 被保険者台帳        | あり                 | なし                          | なし                          | なし            | あり                              |
| 傷病名の正確性担保     | なし                 | なし                          | なし                          | なし            | なし                              |
| データハンドリングサポート | あり                 | なし                          | なし                          | あり            | あり                              |

# 国内の大規模データベース

## 国内の大規模臨床データベース

- DPC (Administrative Claims Database)
- NDB (Administrative Claims Database)

はじめに

# NDB とは何か

- 正式名称：レセプト情報・特定健診等情報 データベース
- 利用目的は全国医療費適正化計画及び都道府県医療費適正化計画の作成、実施及び評価 に資するため  
(高齢者の医療の確保に関する法律 第16条)
- 保有：厚生労働大臣
- 内容
  - レセプトデータ 約87億8,900万件※
  - 特定健診・保健指導データ 約1億2,000万件※
  - Administrative Claims Database の一種

※レセプト情報等の提供に関するワーキンググループ中間報告より (H27.3.17)

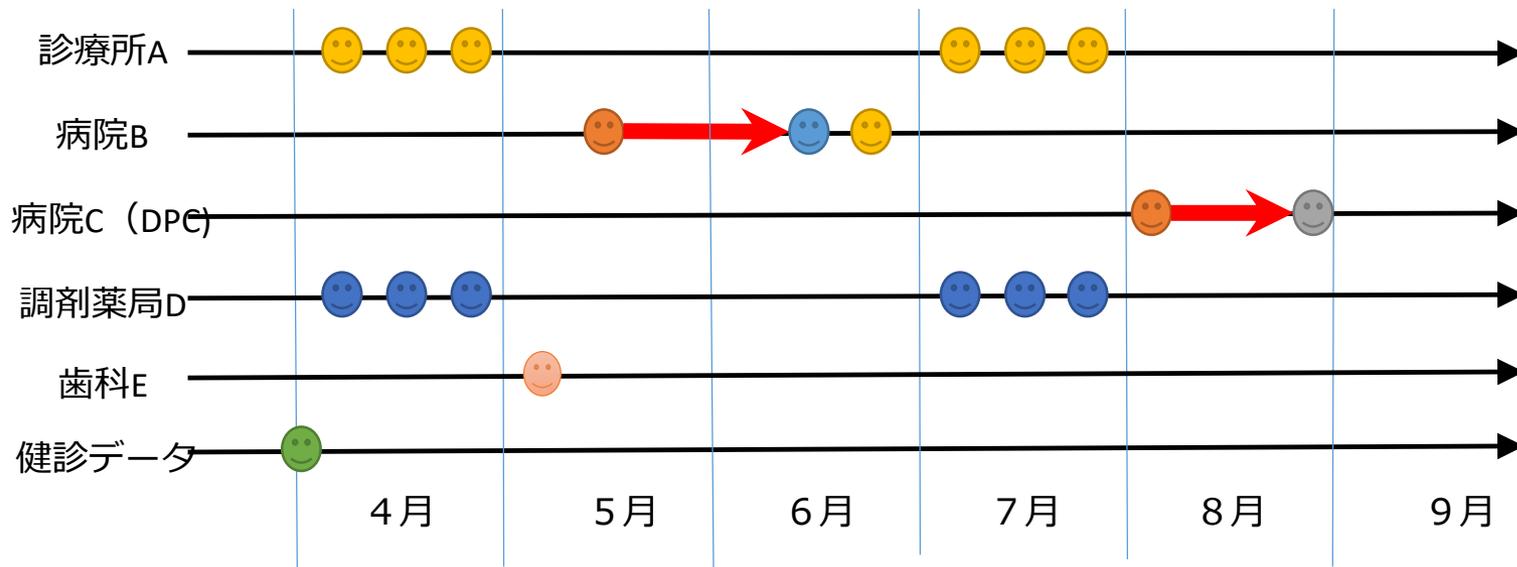
# NDBデータの大まかな構造

レセプト部分と健診情報部分がある。

- 医科(MED)
  - DPC(DPC)
  - 調剤(PHA)
  - 歯科(DEN)
  - 特定健診データ
  - 保健指導データ
- レセプト情報
- 健診情報

# NDBデータ の大まかな構造

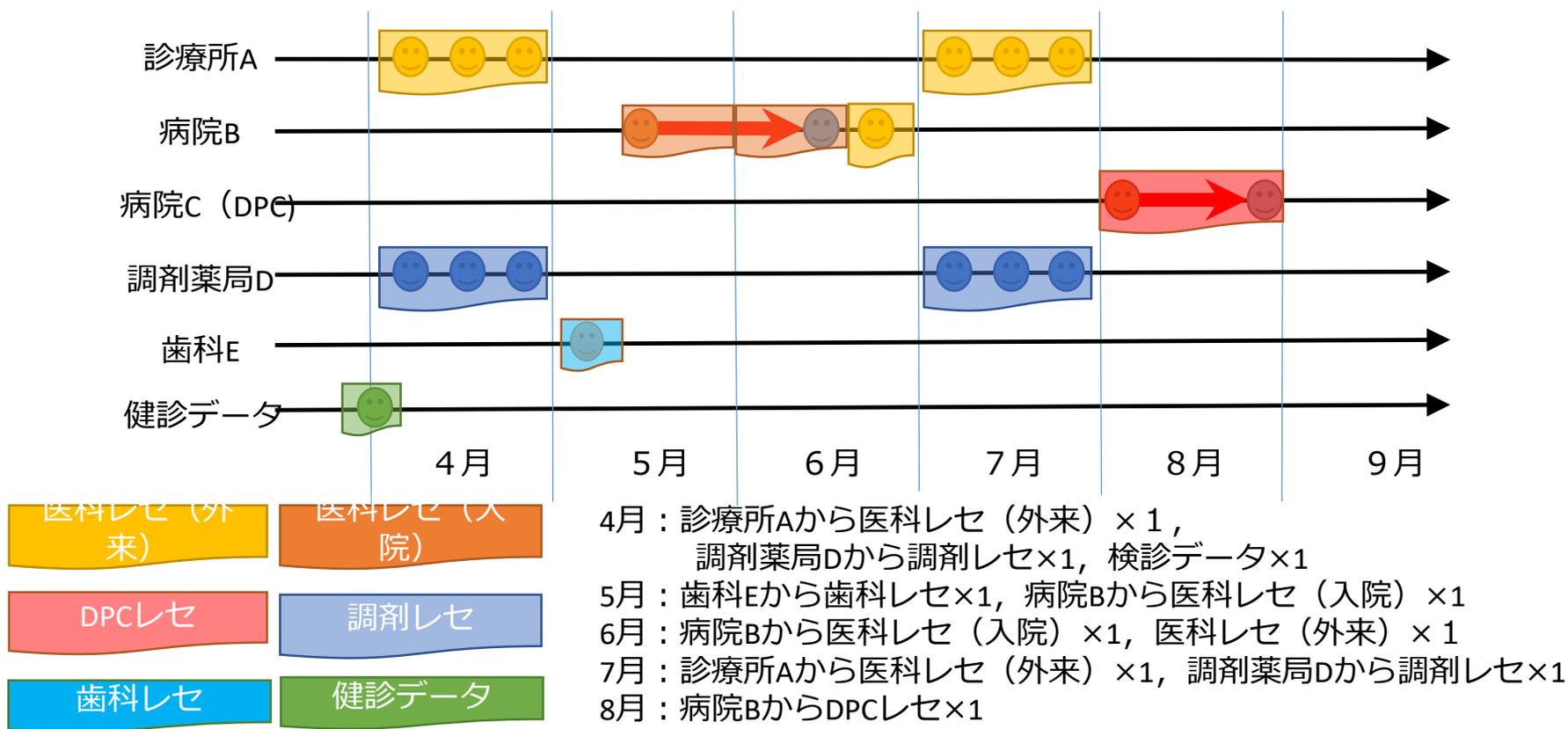
患者のエピソードで考えると以下の様になる。



- 😊 外来受診
- 😊 入院開始
- 😊 健診受診
- 😊 調剤薬局
- 😊 生存退院
- 😊 歯科受診
- 😊 死亡退院

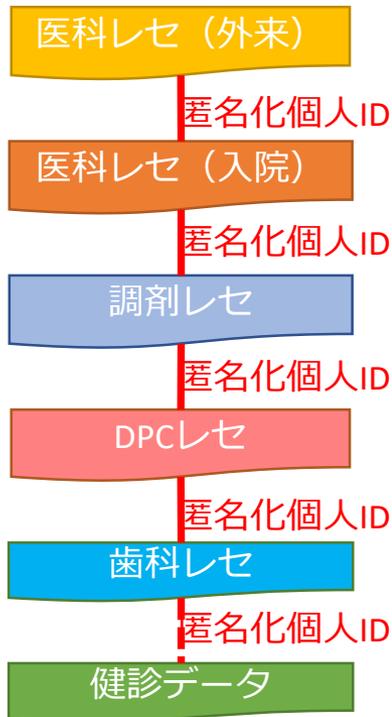
4月：診療所Aに外来受診3回，同日調剤，健診受診  
5月：歯科E受診，病院B入院  
6月：病院B退院，同病院にて外来フォローアップ  
7月：診療所Aに外来受診3回，同日調剤  
8月：病院CにDPC入院，同月死亡退院

# NDBデータ の大まかな構造



# NDBデータ の大まかな構造

匿名化された個人IDで紐付けを行う事が出来る



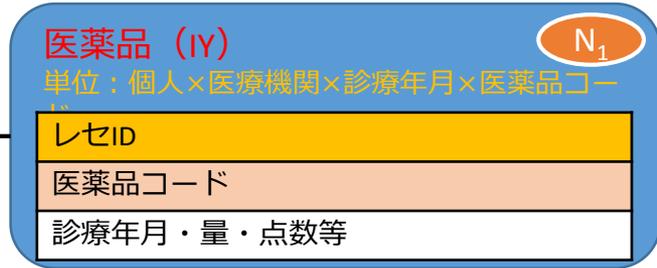
- NDB には2つのID が存在する。
- 匿名化ID1  
保険者番号、被保険者証 の記号・番号、生年月日、性別を基に作成
- 匿名化ID2  
氏名、生年月日、性別を基に作成

# NDBデータ の大まかな構造

- 複数のテーブル（右図）で1レセの情報となる。
- **匿名化個人ID** は各REに含まれる、**匿名化ID1**と**匿名化ID2**の2つ
- 他のレセプト（MED, DPC, PHA, DEN）情報（や健診情報）とは**匿名化個人ID**で接続可能
- 月単位で登録されるため、月またぎ入院や月内外来受診の処理が煩雑



※DPCレセはREが複数個



⋮

# Example of NDB study

Preoperative oral care and effect on postoperative complications after major cancer surgery.

Ishimaru M, Matsui H, Ono S, Hagiwara Y, Morita K, Yasunaga H. ***Br J Surg*** 2018 ;105:1688-96.

# 周術期口腔機能管理による術後肺炎発症予防の効果

## 背景

**ガンの手術後に発症する術後肺炎**は発症率は高くないが、発症すると死亡率を高め、入院日数が伸びてコストがかかる合併症である

術後肺炎は口腔内や咽頭にいる細菌を誤嚥してしまうことで発症すると言われており、**術前の口腔ケア**によって口腔内の細菌数は減少すると報告されている

しかし、**術前の歯科医師による口腔ケア**が実際に術後肺炎を減少させているかは未だはっきりわかっていない

# 周術期口腔機能管理による術後肺炎発症予防の効果

模擬申し出として東大が行っている研究

- P：以下の症例
  - 頭頸部がん・消化器がん・呼吸器がんで手術を行った患者
  - 弁形成・弁置換・冠動脈バイパス術を受けた患者
- E：周術期口腔管理の実施
- C：周術期口腔管理の非実施
- O：術後肺炎発症

# 研究計画-方法-

- レセプトデータを用いてガン手術施行患者を特定
- Exposure: 手術の30日前までに「周術期口腔機能管理料」を策定
- Outcome: 術後肺炎の発症、術後30日以内の死亡
- Propensity scoreを計算し、重み付け分析を行い共変量を調整
- 一般化線形回帰モデルでリスク差を推定

# 研究計画-結果-

- ❖ 509,179人の患者が解析対象
- ❖ 81,632人(16.3%)の患者が術前口腔ケアを施行
- ❖ 15,724人(3.1%)の患者に術後肺炎が発症
- ❖ 1,734人(0.34%)の患者が術後30日以内死亡

# 研究計画-結果-

|                | 曝露群   | 対照群   | リスク差                | P値     |
|----------------|-------|-------|---------------------|--------|
| <b>術後肺炎</b>    |       |       |                     |        |
| 無調整            | 3.28% | 3.05% | 0.23%(0.10-0.36)    | <0.001 |
| 調整後            | 3.28% | 3.76% | -0.48%(-0.64--0.32) | <0.001 |
| <b>術後30日死亡</b> |       |       |                     |        |
| 無調整            | 0.30% | 0.35% | -0.05%(-0.09--0.01) | 0.022  |
| 調整後            | 0.30% | 0.42% | -0.12%(-0.17--0.07) | <0.001 |

# NDB の強みと課題

## NDB の強みと弱み

### 強み

- 悉皆性
  - 一部を除き、全国民を対象とした、データベース
- 症例追跡
  - 医療機関をまたいでの症例追跡が可能
- 厚労省から個票レベル情報の提供が行われている。

### 弱み

- 患者重症度がデータベースに含まれない。
- そのほかにもいろいろな弱点がある。

# 本日本話しする内容

## もくじ

- 大規模データベースを利用した研究活動
  - Administrative Claims Database とは？
  - NDBとはなにか
  - NDBのデータに含まれる情報
  - NDBの強みと弱点
- NDBデータの研究利用
  - 申請方法
- NDBデータの研究利用の課題

# NDBデータの公的利用申請

## 大規模医療データの研究目的利用

- DPCデータは厚労省によるデータ提供が“集計表”に限られている。
- ここからは、NDBデータのデータ提供について説明します。

# 申請にあたって目を通すべき資料

申請にあたって以下の文章はすべて目を通す。

- 利用を検討している方々へのマニュアル
- 第三者提供の審査スケジュール
- **レセプト情報・特定健診等情報の提供に関するガイドライン**
- レセプトの仕様書

# 研究利用の課題

## 研究利用には課題が多い

- 時間のかかる審査体制
- 求められるセキュリティ基準
- ユーザービリティを考慮していないデータ仕様
- ユーザーのニーズとスキルのミスマッチ

# NDB研究利用の課題

## NDBの研究利用には課題が多い

- 時間のかかる審査体制
  - 徐々に迅速化？
- 求められるセキュリティ基準
  - 利用法によっては極めて多額の投資が必要
- ユーザービリティを考慮していないデータ仕様
  - DPCはある意味利用しやすい
  - 次期システムの課題
- ユーザーのニーズとスキルのミスマッチ
  - データをハンドルするための人材が必要

# NDBデータの利用の方法

## 申請方法

- 厚労省窓口（NDB+申請で検索）

[https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou\\_iryoku/iryouhoken/reseputo/index.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryoku/iryouhoken/reseputo/index.html)

- 申請できる者

- ① 国の行政機関
- ② 都道府県
- ③ 研究開発独立行政法人（PMDA含む）
- ④ 大学（大学院含む）
- ⑤ 医療保険者の中央団体
- ⑥ 医療サービスの質の向上等をその設立目的の趣旨に含む国所管の公益法人
- ⑦ 提供されるデータを用いた研究の実施に要する費用の全部又は一部を国の行政機関や研究開発独立行政法人等から補助されている者等

# NDB利用の手順

## NDB の利用には有識者会議の審査が必要

- 有識者会議は年4回
- 開催時期の1か月前くらいに申請書類提出締め切りがある。
- その前に、厚労省担当に事前相談が求められている。
- 申請種別によって、倫理申請を事前に通す必要がある。

# NDB の利用形式

NDB にはいくつかの利用形式があり必要な書類等が違う

- サンプルングデータセット

- ある特定月の横断データデータからレセプトをランダムサンプリングしたもの
- 匿名性を担保するため、コードマスキングがかかっている。
- 外来—調剤間の接続のみなされているが、患者の縦断化は不可能
- 詳細な公表形式・抽出要件などを求められない。
- **必要となるセキュリティ要件の緩和**
- 倫理審査結果の提示は不要

# NDB の利用形式

NDB にはいくつかの利用形式があり必要な書類等が違う

- 集計表の取得
  - 集計方法を指定
  - 集計を厚労省に依頼して集計結果のみを取得する。
  - 申請時に詳細な公表形式・抽出要件を指定する。
  - 必要となるセキュリティ要件の緩和（サンプリングデータセットと同程度）
  - 倫理審査結果の提示が必要

# NDB の利用形式

NDB にはいくつかの利用形式があり必要な書類等が違う

- 特別抽出

- 特定の条件でレセプト情報・特定検診情報を抽出する。
- 抽出された情報を取得解析する。
- **必要となるセキュリティ要件が高い**
- 申請時に詳細な公表形式・抽出要件を指定する。
- 倫理審査結果の提示が必要

# NDBの利用形式

## 特別抽出で必要となる書類の一例

- **様式1：レセプト情報等の利用に関する申出書**
- 様式1-1：レセプト情報等を利用した研究に関する承認書
- 別添1-1：本人確認ができるものの写し
- 別添1-2：本人所属確認ができるものの写し
  
- 研究の目的や方法などを記載

# NDBの利用形式

## 特別抽出で必要となる書類の一例

- 別添2-1：運用フロー図
- 別添2-2：リスク分析・対応表
- 別添2-3：運用管理規程
- 別添2-4：自己点検規程
  
- セキュリティ確保状況の詳細な資料
  - データの保存場所は入退管理が必要
  - データの閲覧場所は入室制限が必要
  - ルール化して文書化

# NDBの利用形式

## 特別抽出で必要となる書類の一例

- 別添7：倫理委員会承諾書の写し
  - 別添8：詳細な抽出条件
  - 別添9：詳細な公表形式
- 
- 必要となるレセプト・特定健診の情報をリストアップ
  - 公表する予定の表や図を説明

# 本日本話しする内容

## もくじ

- 大規模データベースを利用した研究活動
  - NDBとはなにか
  - NDBのデータに含まれる情報
  - NDBの強みと課題
- NDBデータの利用
  - 申請方法
- NDBデータの研究利用の課題

# NDBデータから研究論文を出すまで

研究者のスキルとニーズがずれている。



- NDBで実行可能な研究計画を建てる
- 研究計画に必要なデータを抽出する
- 研究に必要な形式にデータを成型する
- 基準に従いデータをクリーニング
- データの集計と解析
- 論文執筆

研究者はデータを抽出整形する作業のプライオリティを下げている。  
その部分のスキルを十分に有する人材が少ない。

データベースを取り扱うことのできるエンジニアや疫学者との協働が必須

# NDBデータの構造上の問題

- データは保存しやすい形式で保存されている。
  - 利用しやすいデータ形式への変換が求められる。
- 入院期間を適切に計算するだけでも多大な労力を要する。

# NDBデータの構造上の問題

- 社会保障番号が日本には存在しない。
  - 75歳のタイミングでIDが総入れ替え。
  - 名前などから作成された匿名化IDを元に情報の連結が必須。
- 死亡個票などとの連結は不可能
  - 院内死亡ならある程度追える
  - 院外死亡はルールを決めないと厳しい。
- 他のデータベースとの連結は**禁止**されている。

# ID構造と縦断データの作成

匿名化個人IDリストから、追跡可能な全ての匿名化個人IDリストを抽出する

| 匿名化ID1 | 匿名化ID2 | ID3_A | ID3_B | ID3_C | ID3_D |
|--------|--------|-------|-------|-------|-------|
| A      | 1      | B_1   | A_1   | C_2   | A_1   |
| B      | 1      | B_1   | B_2   | C_2   | B_1   |
| B      | 2      | C_2   | B_2   | C_2   | B_2   |
| C      | 2      | C_2   | C_2   | C_2   | C_2   |
| D      | 3      | D_3   | D_3   | D_3   | D_3   |
| E      | 4      | E_4   | E_4   | E_4   | E_4   |

ID3という個人IDを作成するとして、

ID3\_AはID2を基準に作成（例：名前が変更されると追跡不能だが、保険者変更を追跡可能）

ID3\_BはID1を基準に作成（例：保険者変更を追跡不能だが、名前の変更を追跡可能）

ID3\_CはID1 or ID2 を基準に作成（例：保険者の変更・名前の変更を追跡可能だが、同姓同名問題がある。）

ID3\_Dは両者を基準に作成（例：性別が同じ双子も判別可能だが、追跡率は落ちる。）

# 縦断ID作成の事例

学術論文でOnline になったIDの作成法

National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB): Outline and Patient-Matching Technique

<https://www.biorxiv.org/content/early/2018/03/29/280008.article-metrics>

# マンパワーの問題

- 医療や臨床、保険制度のことを理解している。
- データベースのハンドリングを行う技術がある。
- 疫学や統計学の知識が豊富。
  
- 上記を満たすチームとインセンティブの構築が必須

# まとめ

- NDBデータに含まれる情報を理解する。
  - レセプトと特定健診の個票レベル情報が格納されている。
- NDBデータの申請方法について理解する。
- データ利用の申請方法を理解する。
- NDBデータの構造とその課題を理解する。