

図.2 レセプト情報等解析システムのソフトウェア構成

#### 参考文献

- [1] 喜連川優, 合田和生. アウトオブオーダー型データベースエンジン OoODE の構想と初期実験. 日本データベース学会論文誌. Vol.8, No.1, pp.131-136 (2009.06) .
- [2] 山田浩之, 合田和生, 喜連川優. Hadoop におけるアウトオブオーダー型並列処理系の実装に関する一考察. 電子情報通信学会第 5 回データ工学と情報マネジメントに関するフォーラム/第 11 回日本データベース学会年次大会(DEIM2013), F3-3 (2013.03).

平成 24 年度厚生労働科学研究費補助金  
(政策科学総合研究事業 (政策科学推進研究事業))

分担研究報告書

レセプト情報・特定健診等情報データベースについての研究  
(H24-政策-一般-002)

1. 研究報告まとめ	1
2. レセプト情報・特定健診等情報データベースの将来スキームについて：合同研究会とスキームの今後のあり方の検討	
(1) 第1回会合 検討内容 アウトライン案	6
(2) 議事次第	10
(3) 出席者名簿	11
(4) 座席表	12
3. レセプト情報・特定健診等情報データベースの将来スキームについて：課題リスト	
(A) 運用	14
(B) 匿名化、ハッシュ	17
(C) レセプトデータ	19
(D) 申し出	21
(E) HW -SW	23
(F) データ格納方法	25
(G) データセット	26
(H) その他	29
4. レセプト情報・特定健診等情報データベースの将来スキーム平成 24 年 9 月 24 日の合同研究会を元に考えられる今後の在り方	31
5. レセプト情報・特定健診等情報データベースの運用について (プロセス・マップ)	37
6. レセプト情報・特定健診等情報データベースの今後の運用における留意点について	41
7. レセプト等の処理システムの分析を踏まえた NDB データベースのプロジェクト管理・改善に関する提言要素 (継続的改善システムについて)	45
8. 英国 NHS による Administrative Data の全国データベースの運用と二次利用について	53

平成 24 年度厚生労働科学研究費補助金  
(政策科学総合研究事業 (政策科学推進研究事業))  
分担研究報告書

レセプト情報・特定健診等情報データベースについての研究  
(H24-政策-一般-002)

研究分担者：

今中雄一 (京都大学 大学院医学研究科 医療経済学分野 教授)  
猪飼 宏 (京都大学 大学院医学研究科 医療経済学分野 講師)  
大坪徹也 (京都大学 大学院医学研究科 医療経済学分野 助教)

研究協力者：

後藤 悦、小林大介、森島敏隆、國澤 進、佐々木典子、田中将之、宇川直人  
(京都大学 大学院医学研究科 医療経済学分野)

要旨

レセプト等ナショナル・データベースは社会にとって極めて大きな潜在的価値を有しており、これを最大限に活用するしくみを推進していく必要がある。当研究は、レセプト等ナショナル・データベースのデータ処理や活用のあり方について検討し、より効率的で効果的な運用・活用を提案し、より大きな社会的価値を生み出すようにすることを目的とする。

そのために、有識者ならびに現在システムを担当するベンダーと検討会合を持ち、データベース等に強いベンダー各社と検討会合を持ち、各種の専門家等との意見交換、文献、海外の先行事例の動向などから課題と解決策・向上策を検討した。その結果、以下が重要と考えられた。

1. 中期的なアウトプットを具体的にデザインすることで効率的なシステム設計を可能にする。
2. ベンダーに依存しないシステムを構築する。
3. 定型、非定型を分けて、同時処理を可能にし、データ処理を効率的に行う。
4. 研究に活用可能な定期集計データを集めたデータテーブルを充実、公開する。
5. NDB 運用を担当する組織を常設する。
6. 各種マスターを維持管理する組織を持つ。
7. データ形式は研究に適した DPC E/F ファイル形式で構築する。
8. 利用者の支援・教育のための組織を常設する。
9. システム担当組織の中で自らをレビューし改善していく仕組みを持つ。
10. システム担当ベンダーの外部に、当該ナショナル・データベースのシステム運用とデータ活用を継続的にレビューし課題を整理し、継続的に改善を提言する仕組みを持つ。

定期的・不定期を含め数々あるアウトプットとその出力タイミングと必要なハードとソフトを明確に設計し、特定のベンダーに依存しないことが求められる。定期・非定期のタスクのもとに、各種マスターやデータテーブル、システム全体を組織的に常時見直して継続的に完全していく仕組みの確立、などが極めて重要である。

## A. 目的

当研究は、レセプト等ナショナル・データベースのデータ処理や活用のあり方について検討し、より効率的で効果的な運用・活用を提案し、より大きな社会的価値を生み出すようにすることを目的とする。

## B. 方法

レセプト等ナショナル・データベースのデータ処理や活用のあり方について、以下の方法で、検討し提案を構築する。

1) 有識者ならびに現在システムを担当するベンダーと会合を持ち、グループでヒアリングと意見交換を行う。それを以て、解決策・向上策を検討する。

2) データベース等に強いベンダー各社からヒアリングを行い、課題と解決策・向上策を検討する。

3) 各種の専門家等との意見交換、文献、海外の先行事例の動向などから、課題と解決策・向上策を検討する。

## C & D. 結果と考察

重要点は、要約すると以下の如く整理された。ただし、断言の形で言い切っているが、実運用に適用する際には、現実的な状況に合わせて最善の手段をとる必要がある。

### 1. 「中期的なアウトプットを具体的にデザインすることで、効率的なシステム設計を可能にする」

・ NDB の特徴を生かせる定期集計(月次、年次、社会医療診療行為別調査等を含む)、政策上の要求事項、想定される非定期集計、研究領域を明確にして、それに適した新しいシステムの設計を、現行システムにとらわれることなく行う。

### 2. 「ベンダーに依存しないシステムを構築する」

・ データ処理の手順やプログラムを NDB の運営組織に引き継ぎ、また蓄積できるような仕組みづくりを行う。  
・ データ処理過程の全体の仕様において、特定のベンダーによるブラックボックス化を避けるようなオープンシステムを構築する。

### 3. 「定型、非定型を分けて、同時処理を可能にし、データ処理を効率的に行う」

・ 定型(月次処理)、非定型(NDB)に適したシステムをそれぞれ独立に構築し、並列処理を行うことで効率化を図る。  
・ 例えば、定期処理は NoSQL、非定型処理は RDB を用いる。

### 4. 「研究に活用可能な定期集計データを集めたデータテーブルを充実、公開する」

・ 全件をカバーする NDB の特徴として定期集計データだけでも研究に資するものがあり、早急な公開が望まれる。  
・ どのような集計データが有用であるかは、学会等の意見も聞き、検討していく必要がある。

### (実現には予算を伴う NDB 周辺重要事項)

### 5. 「NDB 運用を担当する組織を常設する」

・ データの将来活用計画、保守運用、利用申請への対応を厚労省の少人数の担当者で行うことは限界があるため、今後いっそうの活用を進める上では、常設の組織が NDB の運用を担当することが望まれる。

### 6. 「各種マスターを維持管理する組織を持つ」

・ 必要とされるマスターテーブルは多数あり、例えば診療報酬改定等で随時更新されるものや、市町村合併等により変更されるものもある。  
・ マスターのバージョン管理(更新および追加)に責任を持ってあたる専門組織が必要である。

・研究対象に即して最適なバージョンのマスター類が適用されることで、データ解析の精度が上がることを期待できる。

#### 7. 「データ形式は研究に適した DPC E/F ファイル形式で構築する」

・既に DPC 参加病院では作成されているファイル形式であり、E/F ファイルを厚生労働省に提出している。

・DPC に参加していない医療機関においては、レセコンにすでに E/F ファイルを作成する機能はあり、その機能を追加導入することで、ファイル作成が可能になる。

・各審査支払機関において一時的なシステム変更コストを要する一方、E/F ファイル形式は審査においても一層の効率化に資する可能性が大きい。

#### 8. 「利用者の支援・教育のための組織を常設する」

・申請者や申請を考慮している人へのサポート、データの一部を取り扱うことが可能なオンサイトセンターの運営、人材育成等を担う組織を常設する。

#### 9. システム担当組織の中で自らをレビューし改善していく仕組みを持つ。

#### 10. システム担当ベンダーの外部に、当該ナショナル・データベースのシステム運用とデータ活用を継続的にレビューし課題を整理し、継続的に改善を提言する仕組みを持つ。

が求められる。定期・非定期のタスクのもとに、各種マスターやデータテーブル、システム全体を組織的に常時見直して継続的に完全していく仕組みの確立、などが極めて重要である。

#### F. 健康危険情報

特に無し

#### G. 研究発表

特に無し

#### E. 結論

レセプト等ナショナル・データベースは社会にとって極めて大きな潜在的価値を有する。これを最大限に活用するしくみを導入していく必要がある。

そのためには、定期的・不定期を含め数々あるアウトプットとその出力タイミングと必要なハードとソフトを明確に設計し、特定のベンダーに依存しないこと

2. レセプト情報・特定健診等情報データベースの  
将来スキームについて：合同研究会とスキーム  
の今後のあり方の検討

24年度厚労科研追加交付・今中班  
第1回会合 検討内容 アウトライン案

2012年7月23日  
京都大学大学院 医学研究科  
医療経済学分野

基本方針

レセプト・ナショナルデータベース（以下NDB）の活用における課題の分析と解決策の提言を通して、客観的根拠に基づく制度・政策を推進し、医療とその提供体制の質等の向上を目指す。

1. 効率的なデータ格納について

① 送付されてきたデータが正しく格納されているかどうか

A) 全国の国保連や支払基金から提出されるデータの精度評価

1. どのテーブル、どの項目に精度のばらつきが起こりやすいか。

(ア) データの過分（返戻 含む）、不足分の件数はどの程度か、無くせないか。

(イ) 医療機関によって入力精度が異なるため、国保連・支払基金によって精度が異なるのか。

② 作業工程の見直し、精緻化など

A) 「医療機関でのデータの発生」→「審査支払基金」→「厚労省受付」→「NDB登録」→「NDBでのデータベース処理、抽出」→「利用者に届く」までの工程

1. フロー全貌の把握

2. データ精度に関する管理責任者に合わせた工程の詳細把握

3. 各工程に関する仕様の変更時期や内容、今後のスケジュール

B) データチェック機構のあり方

1. 国保連内部・NDB到着後など、各工程でのチェック内容を知りたい。

C) エラーデータの取り扱い

1. 欠損や誤りのあるデータはどのように報告・修正等行われているか。

(ア) 月またぎの入院なのに前月データが欠損している例、データ欠損なのか保険者変更なのか区別できない例あり。

③ データ精度についての懸案

A) 診療開始日はあるが発症日がないため、急性期か、発症時の症例か、判別困難。

B) 主傷病フラグ欠損が多い。

C) コメントデータの提供が望ましい。可能か。

## 2. NDBの特性を踏まえたデータの補正、実態把握

### ① 空欄をいかに埋め、分析しやすいデータベースとするか

A) 近い将来、レセプトデータの一次正規化にDPC E/F ファイルの方法を活用できないか。

### ② 匿名化の影響についての評価

A) 連結可能性の評価

1. 個人としての追跡ができなくなるケースの整理

(ア) 被保険者の異動、改姓、など？

2. 追跡可能とするための方法（案）

(ア) 生年月日・性別と特定の診療行為等でハッシュ化した第3の連結キー設定、など

B) 連結キーとなる（匿名化された）データ項目の入力精度は評価されているか。

## 3. データの妥当性についての評価

### ① 疾患ごとの評価（例：がん、感染症等における傷病名）

A) 傷病名の入力精度（主傷病フラグ、.9コード）

B) 未コード化傷病名コード0000999の使用頻度

C) 1患者が複数の都道府県で治療を受けた場合の連結可能性

### ② 返戻データの影響についての評価

A) 同一機関・同一診療内容のデータ重複が散見される。返戻レセプトなど重複データの可能性はあるか、同定可能か？

### ③ 審査を行っていることを踏まえた実態評価

A) 診療行為の実回数と請求回数との乖離は把握できるか。

B) 「支払のための審査において注視する事項」と「分析に用いる際に注視する事項」の差異の同定

C) NDB登録時に落とされる項目と分析への影響

## 4. 目的別データベースセットの確立

① 例として、性、年齢層、疾患、入院外来、医療機関（種別、規模等）、地域（都道府県、医療圏等）、保険者、死亡 を基本とし、

② 医療費、レセプト件数、日数・回数、患者数 などから適宜追加することにより

A. 受療行動の分析、B. 特定疾患のプロセス分析、などの目的に応じることを想定。

## 5. データ分析システムにおける、個別依頼によるデータ抽出過程の効率化

① データベースによる抽出 vs テキストファイルベースでの抽出



- ② データ処理速度を高めるプログラミング
- ③ 中間的データベースセット(予備的抽出)の活用
- ④ 中長期的に、使いやすいデータベースのためのレイアウトの見直し

## 6. セキュリティ要件

- ① データのセキュリティレベルを CMS の分類等も参考に分かりやすく整理したうえで、データの取扱者や研究者のセキュリティ要件を段階的に設定することができないか。  
セキュリティレベル例：1：個人識別情報を含む、2：個票データ個人識別情報が含まれない個票データ、3：集計データ

## 7. NDB プロジェクトの持続的な改善に向けた持続的・恒常的な運営支援組織のあり方

- ① データの精度や抽出過程など、定期的に点検し、改善を図る枠組み

## 参考資料

### ■NDB のあり方

医療情報化に関するタスクフォース報告書付属資料

レセプト情報等の活用による医療の効率化について

[http://www.kantei.go.jp/jp/singi/it2/iryoujyohou/dai10/siryou2\\_3.pdf](http://www.kantei.go.jp/jp/singi/it2/iryoujyohou/dai10/siryou2_3.pdf)

### ■医療機関でのデータ作成

社会保険診療報酬支払基金 レセプト電算処理システム

[http://www.ssk.or.jp/rezept/iryokikan/iryokikan\\_02.html](http://www.ssk.or.jp/rezept/iryokikan/iryokikan_02.html)

### ■NDB のシステム調達仕様

レセプト情報・特定健診情報等データベースシステム（仮称）の開発に係る調達一式

<http://www.mhlw.go.jp/sinsei/chotatu/chotatu/database-system/index.html>

### ■NDB 研究利用者への公開データ仕様

試行期間におけるレセプト情報等の提供形式について 2011年5月10日事前説明会資料

<http://www.mhlw.go.jp/stf/shingi/2r9852000001bxyg-att/2r9852000001by2c.pdf>

平成 24 年度厚生労働科学研究「汎用性の高いレセプト基本データセット作成に関する研究」内  
「レセプト情報・特定健診等情報データベースに関する研究」合同研究会

議事次第

日時：平成 24 年 9 月 24 日(月)

10 時 00 分～17 時 00 分

場所：厚生労働省 2 階 共用第 6 会議室

1 開会

2 議題

(1) 本会議の目的について

(2) NDB の全工程について

(3) データ処理の現状と問題点について

(4) NDB のデータ抽出を対象とした今後の DB 運用体制について

(5) NDB の当面の改善策について

(6) まとめ

3 閉会

配付資料

資料 1 本会議の位置づけについて

資料 2 NDB 作業マトリックス

資料 3 富士電機資料

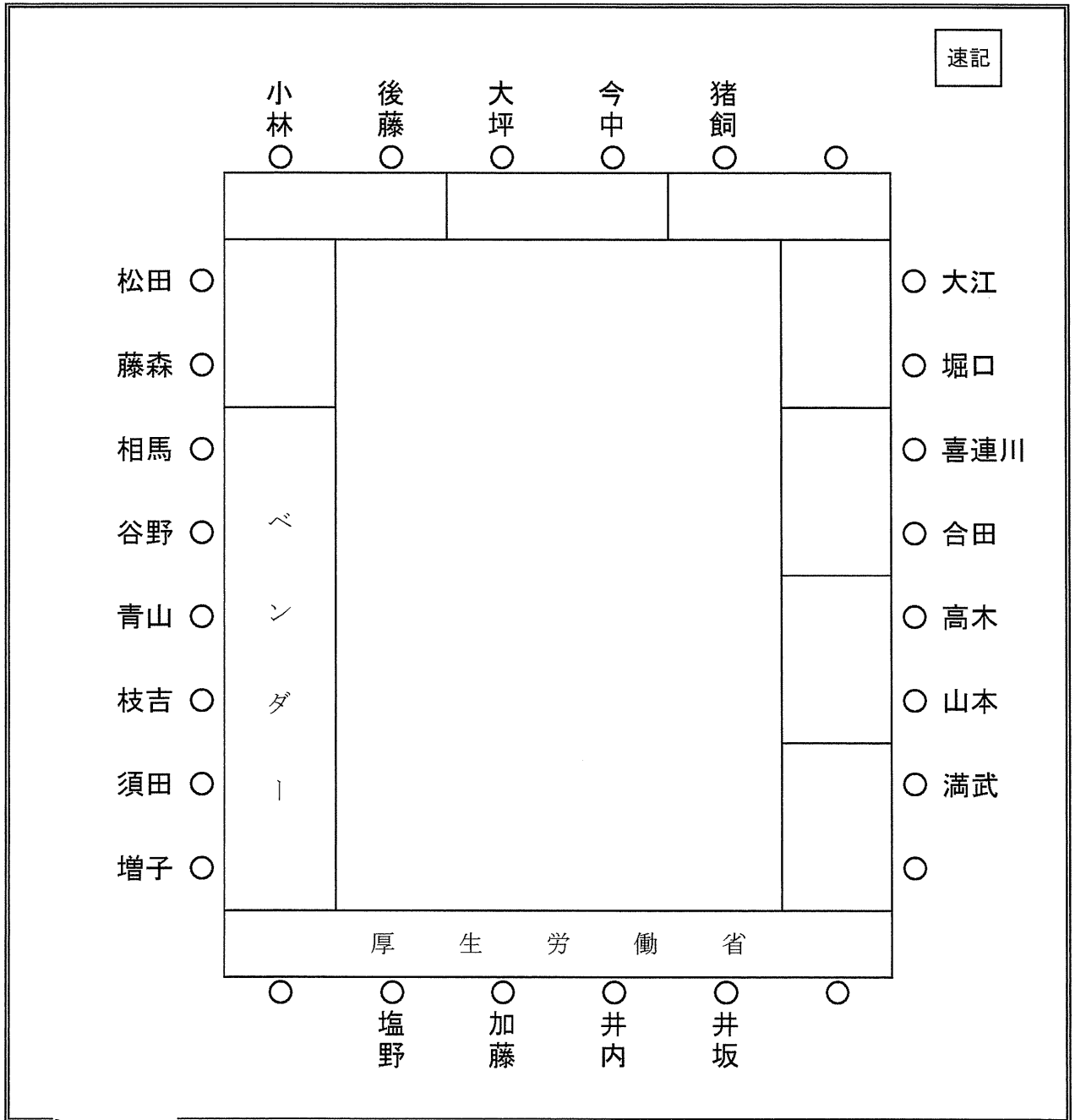
資料 4 参考資料

「汎用性の高いレセプト基本データセット作成に関する研究」内  
「レセプト情報・特定健診等情報データベースに関する研究」  
合同研究会（2012/09/24 開催）出席者名簿

氏名	所属
大江 和彦	東京大学大学院医学系研究科 医療情報経済学分野
喜連川 優	東京大学 生産技術研究所
合田 和生	東京大学 生産技術研究所
高木 利久	東京大学大学院新領域創成科学研究科 情報生命科学専攻
藤森 研司	北海道大学病院 地域医療指導医支援センター
堀口 裕正	東京大学大学院医学系研究科 医療経営政策学講座
松田 晋哉	産業医科大学 公衆衛生学教室
満武 巨裕	医療経済研究機構
山本 隆一	東京大学大学院 情報学環
ベンダー（5名） 相馬、谷野、青山、枝吉、須田、増子	
井坂 一	厚生労働省 保険局総務課 保険システム高度化推進室 室長補佐
井内 努	厚生労働省 保険局総務課 保険システム高度化推進室 室長
加藤 拓馬	厚生労働省 保険局医療課 主査
塩野 友久	厚生労働省 保険局総務課 保険システム高度化推進室 主査
今中 雄一	京都大学大学院医学研究科 医療経済学分野
猪飼 宏	
大坪 徹也	

平成24年度厚生労働科学研究「汎用性の高いレシピ基本データセット作成に関する研究」内  
「レシピ情報・特定健診等情報データベースに関する研究」合同研究会

- ・平成24年9月24日（月） 10:00～17:00
- ・厚生労働省2階 共用第6会議室



(3入口・日比谷公園側)

3. レセプト情報・特定健診等情報データベースの将来スキームについて:課題リスト

(A) 運用

ハードウェア

- 取り込みと抽出の同時作業が出来ないので、(NDBの依頼で)抽出が立て込むと取り込みが出来ない。
- 定型処理は3.5日+1日/月。再匿名さえどうにかなれば。アルゴリズムは良くなってきている。
- 2008年のターゲットと今の状況は違う。抜本的に難しい。基本設計から考える  
14Tくらいの実験システムの現状の横で走らせるのはどうか
- H26、何をするかを2008年よりは明確になっているので、現状を引きずることはないのでは。
- H26以降では、基本仕様をしっかりとすべき。
- H26以降の更新は、富士電機が一番理解しているだろう。「次はこうするんだけど…」があれば、次回以降訊いてみたい。  
(富士電機) 目的別DBがしっかり設計できるなら  
定型処理+第三者提供の範囲で良ければできるか?  
(富士電機) 既に作った定義体が300ある  
スキーマを変えないで、ということか  
(富士電機) yes
- 現状の処理プロセスを見て考えていたい。中間ファイル→outputfile

データチェック

- 国保連や支払基金からのデータをNDB取り込み時に行っているとしているチェックは具体的に? 資料P15  
(富士電機) 年齢→年齢階級等のコード置き換えが出来ているかプログラムでチェック  
IDの振り直しに時間がかかっている

研究者とハードウェア、データ

- H26では、使える人が良い研究に使える環境を

(A) 運用

- NDB運用支援組織が大事。マスター管理や研究者支援をする。  
科研費で東西の大学にセンターを置き、人材育成をふまえてDBを管理する。  
運用で大事なのは人材育成。大きなプランを  
(厚労省の)外の資源をプロジェクトに入れる。

オペレーターとデータ

- オペレーターがボトルネックになっている。増員するか、スタッフ以外が触れるようにすると流れが良くなるのでは？  
(厚労省) どこにどれだけ足りないか分析して、増員あり

研究者とデータ

- データを加工せずにそのままユーザに渡して、好きに抽出させるのはありか？  
(厚労省) 数が少ない症例は個人特定が出来てしまい、それは困る。  
一ヵ月分データを先に渡して、確認してもらっている  
特化したDB作成は準備中
- 最近ではOUTPUTが早くなってきている。  
研究者から見て容易な中間モデルの構築。  
オンサイトセンターを作るのはどうか。個票の持ち帰りは不可だが研究はしてほしいし、結果だけ持って持ち帰りできるような。
- アメリカ、オランダなどでは、データを使わせて研究者に医療政策を勝手に考えさせている。  
但し、コントロールする必要はあり。総合窓口を作り、成果も集める。
- 台湾は数年後に以前に申請されたデータセットを公開。○○先生が××の研究に用いたデータセットを見ることが出来る。  
日本でも申出書に、論文発表後数年後にデータセット公開の可否をチェックさせる欄を作ってはどうか？
- MedPerのシステムが良い。MedPerは使える人を階層的に育てる。抽出データで修士論文程度ならかける。

研究者とオペレーター

- 研究者とベンダーの直接コミュニケーションの場がない。  
研究者とベンダーで意識の乖離が大きい



(A) 運用

データ

- どのような目的別DBにするかを決めない
- 現状をどう改善するか、言葉の定義、マスター(二次医療圏、保険者単位)作成
- NDBの統計値は、国民統計としては多めに出ているのでは？ 既存統計より精度が落ちてないか？

## (B) 匿名化、ハッシュ

### ハッシュ値の問題

- ハッシュ値のつながりに不安がある。ID1、ID2でつながったりつながらなかったり  
ID1は異動、後期高齢者になったり、異字体のブレで同値にならない。5-10年を紐付るとなると不一致がひどくならないか
- レセプトデータの段階で同一人を同定することが難しいことがハッシュ値の問題をさらに複雑にしている  
匿名化されていなければ、カタカナと漢字の対応のあたりは出来るか、生データであっても100%でない  
ハッシュの前の段階で手当てした方が良いかも
- (富士電機) IDだけでカウントすると、一億四千万人くらいになる。(日本の人口より多い)
- どれくらい正しくあるべきかと、思っているのか？ 現実的な期待はどの程度？
- 今は医科だけだが、今後は介護ともつなげたい。これをつながないと高齢化社会に対応しない

### ハッシュ値の問題: 対策

- 今すぐできることと言えば、匿名化をしない
- データ収集時は匿名化しない。データを外に出す時、データセンターで匿名化を行う  
各国保連や支払基金からデータを提供すること自体が「外に出す」ことになっているから、現状は匿名化されている  
研究結果が出て、もしかして元の患者に戻って緊急に知らせることがあった場合、今の状況では元に戻せないからできない
- 特定健診は氏名がカタカナ。漢字表記の揺れをなくすため。  
H26で、レセプトにカタカナ氏名を加えられないか？
- 各国保(保険者)でデータチェックの取り組みが違う。ハイフン等、入力チェックの厳しさが違う。町村レベルだとあまりやってない

### ハッシュ値の問題: 被保険者の追跡

- 保険者が被保険者の異動を報告する仕組みがあれば良い

(B) 匿名化、ハッシュ

匿名化と処理時間

- 名寄せ→ソート→採番しなおしの過程がどうにかなったら、抽出は早くなるだろう  
これらの過程を最適化すれば時間は1/3位にもなるのでは？ 実際は不明  
匿名化の項目が多いために、時間がかかっている可能性がある。

## (C) レセプトデータ

### 匿名化で落ちる情報

- 国保連で匿名化がおこなわれている。  
コメント部分は残しておいて欲しい。せめて未コード化病名コメントは残して欲しい
- 再入院、退院日がない

### データの正しさ

- レセプトデータ連結は、ハッシュ値がそのもととなるデータの入力精度によって変わってしまう可能性あり  
審査支払からデータ受け取り段階で出来るチェックのリストアップして、とどいたデータセットは本当に正しいか、コード類は正しいか、確認  
例えば、総件数のチェックをしてはどうか
- 表記ゆれは許されない仕様では？  
(富士電機) 公開されているチェック項目はあるが、実現されているかどうかは  
二次審査データの差し替え(L3)返戻がNDBIに入ってしまった
- レセプトデータの精度管理、医療機関の入力の段階から定期的に検証する機会を作らないと
- 傷病名の付き方がいいかげん。重みづけロジックを実装してみたら？  
レセプトを統計処理して、パーセントを振るようなロジックを組み込む
- 患者数が少ない症例は、レセプトの欠点が緩和されている。病名、手術などいいかげんなデータが入りにくい。

### 返戻レセプトの扱い

- (富士電機) 匿名化の時点で返戻フラグが落ちてしまっている。記録条件仕様書にフラグの項目がない。  
返戻、L1~L4。L1,2は取り込み不可、L3(返戻),4(事務点検)は取り込み時にフラグがデータから落ちる  
氏名の漢字は、昔のJIS規格なので、その時点で漢字表記にブレが生じている可能性もある  
元データが異なっているためにID1,ID2で連結できない
- 返戻データは最長で2年かかるので、一次審査データを使うのは妥当  
返戻は帰りが紙になることもある