


スライド 1

資料2

平成27年度戦略研究
健康医療研究分野における大規模データの分析及び基盤整備に関する研究

ビッグデータ解析基盤を利用した レセプト情報・特定健診等情報の分析



一般財団法人 医療経済研究・社会保険福祉協会
医療経済研究機構
Institute for Health Economics and Policy
研究代表者：満武 巨裕

スライド 2

研究目的と成果目標

これまで構築してきた超高速レセプトビッグデータ解析基盤を利用し、レセプト情報・特定健診等情報データベース（以下、NDB）の全データ、及び保険者から申請者らが収集した500万人規模の医療・介護データを活用、医療・介護制度の質的向上を目指して以下のテーマを行う。

- 1) 次世代NDBデータ構築、FWA分析等による医療費適正化
[大規模レセプト解析IT基盤整備・運用班]
超高速レセプトビッグデータ解析基盤を活用し、NDBデータの質向上を目的とした検証。本知見に基づき次世代システム構築への提言を行う。テイル集団への詳細分析（Fraud, waste, abuse（FWA）分析）等を実施し、ビッグデータ解析基盤による日本初の医療費適正化のエビデンスを得る。
- 2) 診療エビデンスの明確化と治療方法の標準化
[診療エビデンスと標準化班]
「日本糖尿病学会、及び「日本腎臓学会」と連携し、糖尿病及び腎疾患を対象として、全国の診療バタンの実態・地域差の比較、併存症・医療費等を加味した患者群の経年変化、管理の評価指標を作成。その後、「日本高血圧学会」、「日本動脈硬化学会」を同様の研究（高血圧・高脂血症）も実施予定。
- 3) 在宅の医療（訪問看護等を含む）と介護サービスの連携強化
[医療と介護連携班]
在宅の医療（訪問看護等を含む）、介護の総合的な展開に資するため、医療レセプトと介護レセプトを名寄せし、在宅における医療・介護のサービス受給の実態を分析し、これにより、地域間の比較、時系列による比較を行う。NDBデータ（医療レセプト・特定健診）に介護保険データを含めた次世代システムプロトタイプ構築と提言。
- 4) 医療費支出目標の推計、リスク構造調整の検討
[医療・介護支出目標班] [リスク構造調整班]
当面、医療費の予測や目標設定に必要なデータ項目の整理、推計手法の研究とともに、日本におけるリスク構造調整において想定される要因の特定、調整方式、費用の推計方法の研究。

スライド 3

H27年度の内容

データソース
 NDBデータ（2009-2014年度を入手予定：有識者会議・審議分科会(7/31)において審査、2/3に承諾を得た）
 保険者（三重県の全国保(29市町)と後期高齢者：約75万人）。

ビッグデータ解析基盤への新投資は無し。本邦発のテイル集団への詳細分析を実施
 データ解析基盤整備、各班のBIツール作成

診療エビデンスと標準化班
 対象疾患を限定（糖尿病、慢性腎臓病）
 NDBデータの入手以前は、
 保険者データを活用して効率化。

医療と介護連携班
 三重県の全国保と後期高齢者をフィールドに
 医療および介護レセプト情報を連結

医療費支出目標の推計とリスク構造調整班
 当面、概念整理と分析手法の検討にともな、
 データの解析は次年度以降実施。

スライド 4

データ収集(H27年度)

厚生労働省NDBデータ
 期間： 2009年4月から2014年3月（現時点では2010年度データのみ）
 対象： 全日本（公費は除外）
 種類： 特定健診・保健指導、医療レセプトデータ

保険者データ（三重県）
 期間： 2012年4月から2014年3月
 対象： 三重県全国保(29)と後期高齢者広域連合
 種類： (NDBと同等) 特定健診・保健指導、後期高齢者基本健康診査
 (NDBと同等) 医療レセプトデータ

被保険者マスタ
 介護給付データ
 介護認定データ（一部の市町データで欠損）
 備考： 介護データは、鈴鹿亀山地区連合、紀南介護保険広域連合、紀北広域連合へ個別対応が発生

**次世代NDB
格納データとして**

スケジュール

NDBデータ利用申請 (7/31) → NDB利用承諾 (2/3) → FWA分析I
 NDB検証と改良検討 → 次世代システム構築への提言

各
班・
検
込
等

研究進捗会1 (7/21) → 研究進捗会2 (9/14) → 研究進捗会3 (9/24) → 研究進捗会4 (10/25) → 研究進捗会5 (11/30)

診療エビデンス：DM,CKD (三重県) → 診療エビデンス：DM,CKD (全NDB) → 診療パターン・地域差

データ収集 (国保+広域連合保険者) → 医療・介護費横断比較 → 地域比較 → 状態変化縦断比較 → 次世代システムプロトタイプ構築と提言

第46回戦略研究企画・調査専門検討会 (6/29) → サイトビジット (12/14) → 中間評価 (3/17)

スライド 5

我が国最大級医療BigData利活用 enabled by 世界最先端IT

超高速データベースエンジンを核とするNDB等分析

- 東大発の独自の「非順序型」ソフトウェア実行原理
- 内閣府FIRSTでIT分野唯一の採択課題
- 従来型実装に比して**1,000倍を超える高速性**を実証
- 日立製作所が製品化、**国際ベンチマーク最大クラスに世界初登録**、世界市場で認知
- オープンソース実装でも圧倒的な高速性を発揮（Hadoop, PostgreSQL, MySQL等）
- 我が国全技術領域を対象とする**全国発明表彰'21世紀発明賞**（大学、ベンチャー部門で最高峰賞）
- NDBデータ、三重県・神奈川県国保データ等に於いても、極めて高い有効性を確認済



The diagram illustrates the 'Non-sequential Database Engine' (非順序型データベースエンジン) which enables 'Non-deterministic processing order' (非決定的な処理順序) and 'Massive non-sequential input/output' (超大量非順序入出力発行). It includes a bar chart showing performance metrics and a screenshot of a data analysis dashboard.

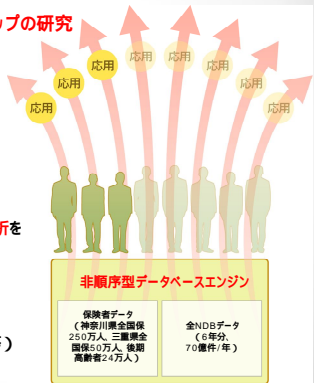
スライド 6

我が国最大級医療BigData利活用 enabled by 世界最先端IT

超高速データベースエンジンを核とするNDB等分析

世界最速のビッグデータ解析技術 → **世界トップの研究**

- 非順序型データベースエンジン（世界最先端IT）
- 目的志向のパーティカルな応用ソフトウェアの機動的立上げ（continuous dev/dep: 研究と共に日々進化するソフトウェア）
- NDB全数からのエビデンス抽出を実現
- これまでは手付かずであった**テイル集団の詳細分析**を実現（FWA分析）
- 「我が国NDBシステムの実験場」**：
本知見に基づく次世代システム構築への提言（データ管理手法、分析手法・インターフェース等）



The diagram shows the 'Non-sequential Database Engine' (非順序型データベースエンジン) being applied to various data sources, represented by arrows pointing to 'Application' (応用) boxes. Below the diagram is a table with data statistics:

非順序型データベースエンジン	
保険者データ (神奈川県全額保 250万人、三重県全 国保50万人、後継 高齢者24万人)	全NDBデータ (6年分、 70万件/年)

スライド 8

NDBピックデータ解析基盤 ---ダッシュボード---

Analytics Dashboard

・インタラクティブな解析ツールをHadoopで実現

・厚生労働省研究費に提供

スライド 11

NDB集計・分析方法に対する、腎臓学会コメント

(方法)

死亡に関する情報はレセプト情報からは抽出ができない。3ヶ月間医療機関を利用していない場合は、死亡と判断しては？

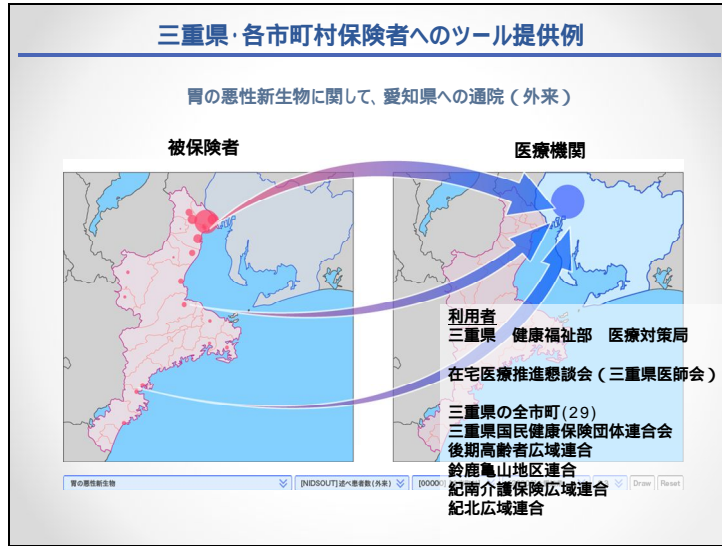
(腎臓学会コメント)

死亡の判断は、3ヶ月では時期尚早かもしれない。
例えば、国保から生活保護に保険者が変わった場合、NDBでは追跡できない可能性が高い。死亡の判断にも大きく影響する

(保険者データ分析での工夫)

保険者から被保険者マスタを入手し、
資格喪失情報から、「死亡、生活保護受給開始、国籍喪失等」を特定

スライド 12



スライド 15



医療（在宅・訪問看護等）と介護サービスの連携強化

市町名	人口 (人)	医療レセプト					介護レセプト		在宅死亡者 数(人)
		医療機関との連携 同時カンファレンス 実施件数 (後期)	往診件数 (後期)	看取り件数 (後期)	訪問診療件 数 (後期)	訪問看護提 供件数 (後期)	訪問看護指 示件数 (後期)	訪問看護 (単位数)	
1 梶名市	140,907	18	1155	62	6,489	97	977	21163	212
2 いなべ市	45,249	6	388	8	1,117	5	354	11919	64
3 木曾町	6,468		63	2	381		20	166	6
4 箕島町	25,467	6	222	17	1,100	22	152	4363	34
5 四日市市	306,107	56	3,461	234	13,273	121	2,718	69,029	663
6 蒲野町	40,325	17	411	27	913	4	251	8,218	88
7 朝日町	10,270	1	96	7	572		70	859	16
8 川越町	14,617	2	114	8	422		30	1,290	30
9 鈴鹿市	197,185	8	1,754	71	6,502	55	1,262	39,945	332
10 龜山市	50,512	1	590	14	2,508	41	590	12,790	100
11 津市	280,647	17	3,843	121	14,447	136	3,901	92,918	568
12 名張市	78,633		772	27	2,723	115	980	31,262	191
13 伊賀市	92,905		1,253	12	3,361	48	1,075	47,117	264
14 松阪市	166,099	2	2,740	67	7,617	64	2,009	68,611	295
15 多気町	14,984	2	252	9	895		164	3,797	41
16 明和町	22,560		386	11	1,797	3	369	10,365	48
17 大台町	9,786		222	7	831	1	110	2,557	35
18 大紀町	9,108		460	8	744	30	119	2,819	38
19 伊勢市	127,587	11	2,660	115	12,033	141	3,136	103,304	336
20 鳥羽市	19,643	1	293	6	1,051	9	193	14,534	56
21 志摩市	51,035	1	1,226	26	4,548	28	292	19,531	220
22 玉環町	15,517	4	159	5	970		160	6,375	24
23 栗会町	8,321		135	2	425		154	5,433	37
24 海陽町	13,054		278	13	1,267	20	294	8,812	52
25 尾鷲市	18,355		306	3	654	1	95	4,809	39
26 熊野市	18,022		374	1	803	50	339	19,455	38
27 紀北町	17,012		815	6	1,596	1	206	7,813	19
28 朝浜町	8,893		177		135	1	71	2,392	14
29 紀宝町	11,223		217	4	586	13	85	4,510	17
三重県	1,820,491	153	24,600	893	89,560	1,066	20,176	626,156	3,877

データ検証（FWA） 次世代NDB構築の提言

[検証1] IDの検証：日本の人口を上回るIDの数

- IDは、被保険者証番号に半角・全角や氏名に関しては漢字表記の揺れ等の理由で、実際には同一の被保険者にも関わらず、複数が発生している可能性がある。
 - ID1とID2の発生数は、ID1が113,970,732件、ID2が122,425,816件。
 - 外来レセプトから、18.4%が氏名変更

ID1に対するID2の数	ID1の数	割合
1	92,475,573	81.6%
2	19,808,138	17.5%
3	508,811	0.8%
4	70,252	0.1%
5	6,192	0.0%
6	1,292	0.0%
7	270	0.0%
8	75	0.0%
9	24	0.0%
10	2	0.0%
11	0	0.0%
12	0	0.0%
13	1	0.0%

[検証2] 外来・調剤レセプトの突合
外来に対応しない調剤レセが12.9%

[検証3] 特定健診データとレセプトデータの突合：約2割

	被保険者数	突合数(人数)	割合	非突合数(人数)	割合
医療レセプトID2	122,425,816	4,339,266	3.5%	118,086,550	96.5%
特定健診 ID2	22,211,086	4,339,266	19.5%	17,871,820	80.5%

NHK NEWSWEB 2013年(平成25年)11月18日(火曜日) 11:03

厚労省の診療データ 約8割が活用できず

厚生労働省が4年前からデータベースに蓄積している診療報酬明細書(レセプト)のデータのうち、約8割が活用できず、残り約2割が活用されていることが明らかになった。厚労省は、データ活用を促進するため、診療報酬明細書のデータ活用を促進する。厚労省は、データ活用を促進するため、診療報酬明細書のデータ活用を促進する。

次世代NDB構築の提言

短期課題

被保険者台帳の収集：
 資格喪失情報（死亡、生活保護受給開始、国籍喪失等）
 資格喪失年月日1～6、喪失事由なども入手

中期課題

新しいIDによるNDBデータの質向上（米国CMS Beneficiary ID）

データの入れ替え（KDB等、ただし、2013年度以降のみ）
 台湾は、初年度データ(国民健康保険導入の1995年)を提供外

介護データの組み入れ

他の研究者がデータを活用する仕組み

基本データセットの提供
 申請者らは、H24～25年度に「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（厚生労働科学研究）」を実施し、当時厚生労働省から提供された保険者（全体の30%）のデータを基に作成した。

H28年2月初旬に東京大学医学部付属病院・大江教授に提供。

平成28年3月16日
 第29回レセプト情報等の提供に関する有識者会議

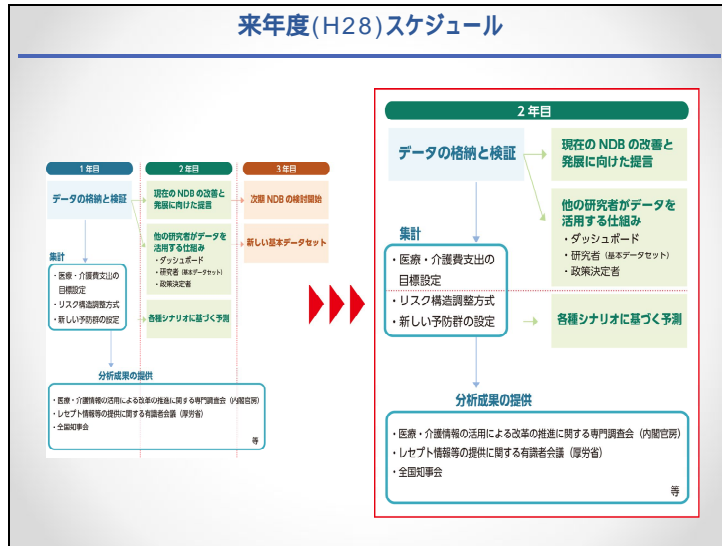
現在提供されているNDBデータの種類				
	特別抽出	サンプリングデータセット	基本データセット	集計表情報
基本的なイメージ	抽出者の属性、ICD、データセットに包含されるデータの範囲、該当する病種の特長を抽出、提供可能なデータセット	診療科研究へのニーズに応じ、抽出、提供可能な範囲でデータセットを作成する。抽出するデータの範囲、抽出方法を調整可能なデータセット	入院、外来、検査結果、処方箋情報等に関するデータセット	研究家の意向に応じ、データ抽出して作成した集計表情報とする
提供データ	病種	一部病種に限定された病種	病種、ICD、年齢	集計表
含まれているデータの種類	レセプト情報、診療報酬、病名、薬剤処方情報、ICD、ICD-10、ICD-9、ICD-10-CM、ICD-10-PCS、ICD-10-PCS、ICD-10-PCS、ICD-10-PCS	病種に限定されたデータセット	病種、年齢、性別、ICD-10、ICD-9、ICD-10-CM、ICD-10-PCS、ICD-10-PCS、ICD-10-PCS	集計表
利用にあり得る見解すべきポイント	データ抽出、提供可能な範囲でデータセットを作成する。抽出するデータの範囲、抽出方法を調整可能なデータセット	抽出するデータの範囲、抽出方法を調整可能なデータセット	抽出するデータの範囲、抽出方法を調整可能なデータセット	抽出するデータの範囲、抽出方法を調整可能なデータセット
認定される有識者	レセプト情報、診療報酬、病名、薬剤処方情報、ICD、ICD-10、ICD-9、ICD-10-CM、ICD-10-PCS、ICD-10-PCS、ICD-10-PCS	抽出するデータの範囲、抽出方法を調整可能なデータセット	抽出するデータの範囲、抽出方法を調整可能なデータセット	抽出するデータの範囲、抽出方法を調整可能なデータセット
提供実績（件数）	42件	15件	2件	7件

研究に伴う教育的要素

医学会との連携による疾病特定の定義、データクリーニング、データ抽出、データ分析のロジックは、プログラムやSQL文の履歴をシステム上に残している。（台湾では利用申請の承認を受けたデータを他の研究者でも使えるようにしている。）

他の研究者が参照可能にすることが、今後の研究発展における短期的な第1歩。

スライド 21



スライド 22

