

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）

レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析および超高速レセプトビックデータ解析基盤の整備
(課題番号 H27 - 政策 - 戦略 - 014)

平成 27 ~ 28 年度 総合研究報告書

研究代表者 満武 巨裕

平成 29 年 (2017) 年 3 月

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

目 次

. 総合研究報告書

レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析および超高速レセプトビッグデータ解析基盤の整備

満武巨裕

1

. 分担研究報告書

1. レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析および超高速レセプトビッグデータ解析基盤の整備

満武巨裕、辻 哲夫、喜連川 優、合田 和生、伏見 清秀、辻 一郎、藤森 研司、興杢 貴英、中島 直樹、石川ハヅル、山光一、飯島 勝矢、吉江 悟、関本 美穂、印南 一路、高橋 邦彦、奥村 泰之、清水 沙友里

11

2. 超高速レセプトビッグデータ解析基盤を利用した医療需要の把握・整理

満武巨裕、辻 哲夫、喜連川 優、合田 和生、伏見 清秀、辻 一郎、藤森 研司、興杢 貴英、中島 直樹、石川ハヅル、山光一、飯島 勝矢、吉江 悟、関本 美穂、印南 一路、高橋 邦彦、奥村 泰之、清水 沙友里

18

資料 1：研究実施計画書

22

資料 2：平成 27 年度研究スライド

49

資料 3：平成 28 年度研究スライド

57

. 研究成果の刊行に関する一覧表

66

レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析
および超高速レセプトビッグデータ解析基盤の整備

研究代表者 満武 巨裕

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、副部長

研究要旨

本研究の目的は、超高速レセプトビッグデータ解析基盤を利用することにより、レセプト情報・特定健診等情報データベース（以下、NDB）のデータ（年間370億レコード）及び保険者から収集した75万人規模の医療・介護データを利用して、地域（都道府県・市町村）や保険者ごとの性別・年齢別・疾患別医療費および罹患率を把握し、都道府県・全保険者の医療・介護支出目標を精密に設定する推計方法を開発し、増大する国民医療費の適正化対策としての施策導入を目指す。また、複数年データを活用して時間経過に伴い新たに作成したデータ項目と医療・介護供給体制の情報とリンクさせて、需要および供給体制を把握し、現実のデータの変遷に合わせた予測が可能な先進的なヘルスケアビッグデータサービスを実現する。

具体的には、以下の5つとした。

- 1) 4 医学会と連携した NDB レセプトデータから生活習慣病の発症数・合併症数・重症群の定義を確立を通し、わが国の生活習慣病・癌等の複数疾患の診療の実態を明らかにし、診療エビデンス・パフォーマンスの分析・公開
- 2) 国・都道府県・市町村レベルで、医療費・介護費支出の目標を設定する方法論の確立
- 3) 日本版リスク構造調整方式の確立と財源調整施策の導入の試算
- 4) NDB レセプト情報と特定健診データを活用した新しい1次・2次・3次予防群の確立
- 5) レセプト情報と介護レセプトを活用し訪問診療をはじめとする在宅医療の実態を明らかにするとともに、疾病予防・医療・介護の連携の実態把握

採択額及び研究期間を考慮し、研究内容を重点化した。

ビッグデータ解析基盤の増強については、物理的には、東京大学構内にビッグデータ解析基盤およびソフトウェア開発用のセキュリテールームを設置した。ハードウェア構成は、サーバ装置群、専用パーティションに設置したクライアント端末構成とした。サーバ装置群は、厚生労働省のガイドラインに基づくレセプト情報等の堅牢な管理と、レセプト情報等に対する解析クエリの実行を担う。クライアント端末は、研究者へのインターフェースであり、レセプト情報等の管理ならびに解析の命令をサーバ装置群に発行し、また、その結果を取得して出力する。

NDBデータの集計・分析としては、疾患毎罹病率、医療費の傾向などの情報を都道府県・二次医療圏別に作成する。次に、患者数が多く大量の医療資源を消費している生活習慣病について、「日本腎臓学会」等と連携し、これらの疾患管理を評価するための指標を同定する。さらに、保険者から収集したデータを利用して、在宅医療（医療保険訪問看護を含む）と介護サービスがどのように併用され、サービスに要する費用等の実態を分析する。特に、在宅医療介護連携の実態を明らかにし、同じ医療や介護が必要とされる高齢者の入院や外来受診、在宅医療の地域比較を行い、現在進行中の医療介護総合改革の適切な推進等に資するエビデンスに基づく政策の推進を図る。

三重県の全ての国民健康保険と後期高齢者のデータ利用し、供給データ（医療機関・介護施設・介護サービス事業所の基本的属性）と組み合わせることで、医療と介護サービスの連携でどのような使われ方をしているか、個人単位で可視化することができた。また、ステークホルダー（主に保険者と地域医師会）用に4種類のデモを作成した。デモは、市町ごとに病名が選択でき、入外来別の患者数と医療費、どこの医療機関を受診したかを可視化した。その結果、地域により疾患別受診状況に差異があることが明らかになった。特に、二次医療圏を超える受診実態が示された。介護サービスの利用についても、地域ごとの特性が示された。本研究により、種々のデータを組み合わせて医療や介護の実態を可視化することで、現在進行中の医療介護総合改革の推進等に資するエビデンスを作成できることを示した。

分担研究者

辻 哲夫・医療経済研究機構・副所長
喜連川 優・東京大学生産技術研究所・教授
合田 和生・東京大学生産技術研究所・特任准教授
伏見 清秀・東京医科歯科大学医療政策学・教授
辻 一郎・東北大学大学院医学系研究科・教授
藤森 研司・東北大学大学院・医学系研究科・教授
興相 貴英・自治医科大学・企画経営部医療情報部・准教授
中島 直樹・九州大学病院メディカルインフォメーションセンター・教授
石川ベンジャミン光一・国立がん研究センター

がん対策情報センターがん統計研究部・室長
飯島 勝矢・東京大学高齢社会総合研究機構・准教授
吉江 悟・東京大学医学部在宅医療学拠点・特任研究員
関本 美穂・大阪府済生会吹田病院麻酔科・医長
印南 一路・慶應義塾大学総合政策学部・教授
高橋 邦彦・名古屋大学大学院医学系研究科・准教授
奥村 泰之・医療経済研究機構・主任研究員
清水沙友里・医療経済研究機構・主任研究員

A. 研究目的

本研究の目的は、超高速レセプトビッグデータ解析基盤を利用することにより、レセプト情報・特定健診等情報データベース（以下、NDB）のデータ（年間370億レコード）及び保険者から収集した75万人規模の医療・介護データを利用して、地域（都道府県・市町村）や保険者ごとの性別・年齢別・疾患別医療費および罹患率を把握し、都道府県・全保険者の医療・介護支出目標を精密に設定する推計方法を開発し、増大する国民医療費の適正化対策としての施策導入を目指す。また、複数年データを活用して時間経過に伴い新たに作成したデータ項目と医療・介護供給体制の情報とリンクさせて、需要および供給体制を把握し、現実のデータの変遷に合わせた予測が可能な先進的なヘルスケアビッグデータサービスを実現する。

B. 研究方法

本研究では、1.NDBの全データ、2.三重県下の全国民健康保険および後期高齢者の医療・介護・健康データ（75万人規模）を利用する。

NDBからは2008年度から2014年度迄の全データを取得する。（2010年度の全NDBデータ（電子レセプト：約15億件、特定健診等データ約2000万件）は取得済み）。

三重県の全市町村（国民健康保険）と後期高齢者医療広域連合2008-203年度のデータも収集する。

保険者からは、NDBと同等のデータに加えて、現在のNDBには蓄積されていない、被保険者マスタ、特定健康診査・特定保健指導の対象者ファイル、介護受給者ファイ

ル、介護給付（費）ファイルを収集する。被保険者マスタには、全加入者の基本的属性（性別・年齢）に加え、保険加入日、保険資格喪失日などが含まれ、分析対象群を設定するための必要なファイルである。

特定健康診査・特定保健指導の対象者ファイル（国保システム・特定健診結果等情報作成抽出（受診券情報）ファイル）は、健診の未受診者の基本的属性が含まれる。加えて、特定健診受診者CSVファイル、特定健診結果等情報作成抽出（健診結果情報（横展開））ファイル、特定健診結果等情報作成抽出（保健指導情報）ファイルを手に入る。

介護保険については、介護受給者ファイル（医療費の被保険者マスタに相当）に認定審査結果情報等がふくまれる。介護給付費ファイルは、レセプトに相当する。

使用するデータで示したように、本研究で対象とするデータは膨大であり、従来のデータの扱い方では十分な分析が出来ない。そのためビッグデータ解析基盤を構築する必要がある。既に内閣府Firstおよび厚生労働科学研究費補助金において基礎的な基盤を構築しているが、本研究ではデータがさらに増加するため、ビッグデータ解析基盤の増強を行う。

次に、分析の質を高めるためには原データの質を高める必要がある。一人の患者に関わるレセプトが国内約18万の医療施設（病院約8500、診療所約10万、歯科診療所約7万）および調剤薬局（約5.5万）から発生するため、それらの間の紐づけがうまくいかないことがあることが知られている。本研究においてはNDBと直接保険者から入手したデータを比較すること

等を通じて問題点を洗い出し、解決策を提示する。さらにビッグデータ解析基盤を用いてこれまでより大幅に迅速な基礎統計値を産出する体制を整える。

上記の基盤構築、データ検証を行った後に、レセプトデータを分析することで以下に取り組む。

- ・ビッグデータ解析基盤の増強
- ・全 NDB データの検証と質向上
- ・基礎統計
- ・多次元分析
- ・都道府県・二次医療圏・市町村別分析

はじめに、ビッグデータ解析基盤の増強を行う。物理的には、東京大学構内にビッグデータ解析基盤およびソフトウェア開発用のセキュリティルームを設置する。ハードウェア構成は、サーバ装置群、専用パーティションに設置したクライアント端末構成とする。サーバ装置群は、厚生労働省のガイドラインに基づくレセプト情報等の堅牢な管理と、レセプト情報等に対する解析クエリの実行を担う。クライアント端末は、研究者へのインターフェースであり、レセプト情報等の管理ならびに解析の命令をサーバ装置群に発行し、また、その結果を取得して出力する。

NDBデータの集計・分析としては、はじめに、高速保険レセプト解析基盤の実行性能比較を行った。4種類のクエリ（クエリ1：都道府県ごとの急性心膜炎患者数（対人口比率）、クエリ2：口唇の悪性新生物の都道府県ごとの平均診療費、クエリ3：呼吸器結核に関する年齢ごとのレセプト件数、クエリ4：コレラ患者の都道府県ごとの平均診療費）を実行して、検索時間を比較した。次に、薬剤の処方傾向の分析と

して、処方件数および併用薬や併用禁忌の実態を分析した。

患者数が多く医療資源を消費している生活習慣病について、「日本腎臓学会」等と連携し、患者数、疾病がどのような経過をたどるのか（時系列分析）、疾病に対してどのような診療が行われているか（診療パターン）といった分析を行い、新しい診療ガイドラインを作成するためのエビデンス、疾病を管理するための評価指標の作成につなげる。

保険者から収集したデータを利用して、在宅医療（医療保険訪問看護を含む）と介護サービスがどのように併用され、サービスに要する費用等の実態を分析する。特に、在宅医療介護連携の実態を明らかにし、同じ医療や介護が必要とされる高齢者の入院や外来受診、在宅医療の地域比較を行い、現在進行中の医療介護総合改革の適切な推進等に資するエビデンスに基づく政策の推進を図る。

C. 研究結果

採択額及び研究期間を考慮し、研究内容を重点化した。

三重県の全国健康保険（29市町）と後期高齢者医療広域連合（75歳以上）の最新の医療データ（電子レセプト）・特定健診データ、介護データはの収集が完了した。ビッグデータ解析基盤への新投資はしないで、セキュリティ環境・ソフトウェア開発、保険者データ（三重）の格納を行った。基礎統計については、可視化ツールを作成し、ビッグデータシステムに実装した。

高速保険レセプト解析基盤の実行性能比較では、従来型と比べて約 100 倍となる高速化を実現できた。

薬剤の処方傾向の分析では、主要抗うつ薬と湿布薬(ケトプロファン、ロキソプロフェナトリウム水和物処方件数)の年間点数合計と処方日数(365日を超える実態把握)、高血圧治療薬の処方件数と併用薬・併用禁忌の実態を示した。

生活習慣病の分析結果は、例えば慢性腎臓病(CKD)の患者数について約 22.5 万人を抽出した。透析患者については、診療行為名称を特定したところ、NDB から推計した全国の透析患者数は、約 22 万人(生活保護は含まず)であった。また、腎移植後に継続的に透析を行っているの患者が、298 人存在することを確認した。

三重県の全ての国民健康保険と後期高齢者のデータ利用し、供給データ(医療機関・介護施設・介護サービス事業所の基本的属性)と組み合わせることで、医療と介護サービスの連携でどのような使われ方をしているか、個人単位で可視化することができた。具体的には、複数年データを用いて、生活習慣病の高血圧患者の推移を観察し、診断済群(投薬なし、合併症なし)、服薬治療群(投薬あり、合併症なし)、合併症群(投薬なし、合併症あり、透析あり)、透析治療群(投薬あり、合併症あり、透析あり)毎に継時的変化を示した。

保険者から収集したデータを利用したサービスに要する費用等の実態を分析については、ステークスホルダー(主に保険者と地域医師会)用に 4 種類のデモを作成した。デモは、市町ごとに病名が選

択でき、入外来別の患者数と医療費、どこの医療機関を受診したかを可視化した。その結果、地域により疾患別受診状況に差異があることが明らかになった。特に、二次医療圏を超える受診実態が示された。介護サービスの利用についても、地域ごとの特性が示された。

D. 考察

我々は、最終的な研究成果の政策への反映方法として以下の 4 つを設定した。

1) 次世代 NDB データ構築と分析による医療費適正化

独自に収集した保険者データと NDB データの突合検証を行うことで、検証、本知見に基づく次世代システム構築への提言を行う。計画通り、検証を実行している。

2) 診療エビデンスの明確化と治療方法の標準化

医学会連携することで、初年度は 2 疾患の診療実態と地域差を明らかにし、患者に対する適切な医療の提供、評価指標の作成を行っている。

3) 在宅医療(訪問看護等を含)と介護サービスの連携強化

在宅の医療・介護の給付実態について、地域ごとの特性分析等を実施。将来、NDB データ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による、効率的な医療費・介護費の総合的な適正化につなげる。

4) 医療費支出目標の推計、リスク構造調整の検討として、今後の医療保険制度の在り方に関する検討に寄与する。

今回、研究の採択額及び研究期間を考慮

し、研究内容を重点化した。当初の予定とは異なり、データ収集等について想定以上に時間を要した。だが我々は、保険者から独自に収集したデータ（三重県 75 万人規模）を最初に分析、知見を得た上で、次に全日本の NDB データを分析している。これにより抽出時間、処理時間などの見積もりができ、分析の効率化が実現できていると思われる。

高速保険レセプト解析基盤の実行性能比較から、従来型と比べて約 100 倍となる高速化を実現でき、処方傾向の分析等において応用例を示すことができた。レセプト病名については、例えばステージ情報付の傷病名等を有する入力をしている医療機関が少ない、移植のレセプトは紙で請求されることが多いといった理由から、NDB を利用しても全数を把握できない可能性がある。しかし、病名と処置、処方を組み合わせて対象疾患を選定することで治療中の患者の把握率が向上する可能性がある。

今回、ビッグデータ基盤と医療政策・経済等研究を連携させ、研究者、地域医師会、地方自治体、保険者らステークホルダーが実態把握・政策提言を行い社会的にもメリットが得られるようにデータや BI ツールを開発したところ、種々のデータを組み合わせることで、現在進行中の医療介護総合改革の適切な推進等に資するエビデンスを作成できることを示すことができた。今後も、現在の関係者以外にも一定の認識が共有され、諸課題の解決に向けた社会的な合意を形成するための基盤づくりが必要であ

り、現在進行中の医療介護総合改革の推進等に資するエビデンスを作成できることを示した。

E. 結論

種々のデータを組み合わせて医療や介護の実態を可視化することで、現在進行中の医療介護総合改革等の推進等に資するエビデンスを作成できることを示した。

F. 健康危険情報

なし

G. 研究発表

- 1) 「基本データセットの提供について」第 29 回レセプト情報等の提供に関する有識者会議（平成 28 年 3 月 16 日）、<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000117367.pdf>
- 2) 満武巨裕：レセプトビッグデータ解析の現状と将来 実験医学 第 34 巻第 5 号：799-804, 2016 年
- 3) 「諸外国の医療ビッグデータ」第 2 回データヘルス時代の質の高い医療の実現に向けた有識者検討会（平成 28 年 5 月 23 日）、<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000125318.pdf>

H. 知的財産権の出願・登録状況

該当なし

レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析
および超高速レセプトビッグデータ解析基盤の整備

研究代表者 満武 巨裕

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、副部長

研究要旨

本研究の目的は、申請者が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更に発展させることにより、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（年間370億レコード）及び保険者から申請者らが収集した75万人規模の医療・介護データを利用して、全都道府県・市町村の精緻な医療・介護費支出の目標設定、財源調整施策としてのリスク構造調整方式の導入、全国民に対する新基準の1次・2次・3次予防群の設定が可能であることを明らかにすることである。こうしたプロセスを経て、現実のデータの変遷に合わせた予測が可能なヘルスケアビッグデータサービスを構築する。

本研究では第一に、地域（都道府県・市町村）や保険者ごとの性別・年齢別・疾患別医療費および罹患率を把握し、都道府県・全保険者の医療・介護支出目標を精密に設定する推計方法を開発し、増大する国民医療費の適正化対策としての施策導入を目指す。また、国民（被保険者）の健康状態・罹患状況に関して、電子レセプトデータの傷病名・診療行為・医薬品情報、介護認定・給付情報、特定健診等情報を利用してエピソード単位で把握する方法論を開発する。

第二に、集計したパラメーターを活用し、保険者の加入者（被保険者）の各種リスク状態を数値化し、予測モデルを構築する。特に、保険者の財政調整に関してエビデンスに基づくリスク構造調整方式が導入可能であることを明らかにする。また、複数年データを活用して時間経過に伴い新たに作成したデータ項目と医療・介護供給体制の情報とリンクさせて、需要および供給体制のパラメーターを変化させたシミュレーションを行なう。こうしたプロセスを経て、現実のデータの変遷に合わせた予測が可能な先進的なヘルスケアビッグデータサービスを実現する。

第三に、特定健診・特定保健指導および電子レセプトに基づく、新たな特定健診・特定保健指導の評価指標の作成、特定保健指導の対象群や医療機関との連携を促進する群等の1次・2次・3次予防群の設定が可能であることを明らかにする。

本年度は、採択額及び研究期間を考慮し、研究内容を重点化した。データソースは、NDB データ（2009-2014 年度）を入手するため、レセプト情報等の提供に関する有識者会議・審議分科会(7/31)において審査、2/3 に承諾を得た。また、三重県の全国保(29 市町)

と後期高齢者から約 75 万人規模の NDB と同等のデータに加えて、現在の NDB には蓄積されていない、被保険者マスタ、特定健康診査・特定保健指導の対象者ファイル、介護受給者ファイル、介護給付（費）ファイルを集めた。ビッグデータ解析基盤への新投資は不可能であったが、本邦発のテイル集団への詳細分析を実施し、データ解析基盤の整備として BI ツールを作成した。分析対象疾患についても、糖尿病、慢性腎臓病について限定して分析を行った。医療と介護連携については、三重県の全国保と後期高齢者を対象に、医療および介護レセプト情報を連結して分析を行った。医療費支出目標の推計とリスク構造調整については、概念整理と分析手法の検討にとどめて、データの解析は次年度以降とした。

今後も引き続き、最先端のビッグデータ基盤と最先端の医療政策・経済等研究を連携させ、研究者、厚生労働省、地域医師会、地方自治体、保険者らステークホルダーがデータや BI ツールを使い論文化・実態把握・政策提言を行い技術的にも社会的にもメリットが得られるようにする。また、現在の関係者以外にも一定の認識が共有され、諸課題の解決に向けた社会的な合意を形成するための基盤づくりについても検討を続ける。

分担研究者

辻 哲夫・医療経済研究機構・副所長
喜連川 優・東京大学生産技術研究所・教授
合田 和生・東京大学生産技術研究所・特任准教授
伏見 清秀・東京医科歯科大学医療政策学・教授
辻 一郎・東北大学大学院医学系研究科・教授
藤森 研司・東北大学大学院・医学系研究科・教授
興相 貴英・自治医科大学・企画経営部医療情報部・准教授
中島 直樹・九州大学病院メディカルインフォメーションセンター・教授
石川ベンジャミン光一・国立がん研究センターがん対策情報センターがん統計研究部・室長
飯島 勝矢・東京大学高齢社会総合研究機構・准教授
吉江 悟・東京大学医学部在宅医療学拠点・

特任研究員

関本 美穂・大阪府済生会吹田病院麻酔科・医長
印南 一路・慶應義塾大学総合政策学部・教授
高橋 邦彦・名古屋大学大学院医学系研究科・准教授
奥村 泰之・医療経済研究機構・主任研究員
清水沙友里・医療経済研究機構・主任研究員

A. 研究目的

医療分野でのビッグデータの一つに、2009 年度から日本の全患者の保険診療情報を蓄積している国家規模のデータベースがある。これはレセプト情報・特定健診等データベース（以下 NDB）と呼ばれ、このデータを分析し医療政策に利用することが医療の質の改善や医療費適正化の切り札となることが期待されている。

医療費の適正化を目的として、これまで幾多の医療制度改革が行われてきた。例えば2008年度には、特定健康診査・特定保健指導(メタボ健診)や後期高齢者医療制度が導入された。メタボ健診は生活習慣病の一次予防により将来的な医療費を削減する目的で導入されたが、これが本当に生活習慣病を予防し、医療費を削減する効果を持つのかに関するエビデンスは未だ示されていない。

エビデンスを示すことが困難であった原因の一つは、これまで診療の実態や診療行為のアウトカム、医療のコストなどを包括的に分析するための国家規模のデータが存在しなかったことである。レセプト情報は保険者に関わりなく同じフォーマットで作成されるため、この目的にかなうデータとして有望視されてきたが、全国規模のデータを集めるシステムが長らく存在しなかったため、その分析は一部の研究者や研究機関による狭いテーマに限定されてきた。NDBには、全医療機関で提供された医療をカバーするレセプト情報、特定健診情報が収載され、2013年度の社会保障制度改革国民会議の報告書(内閣官房)においても「国が保有するレセプト等データの利活用の促進にも不可欠である」と大きな期待が寄せられている。

また、今後の超高齢社会に向けて目指すべき地域包括ケアシステム構築の中で鍵となる在宅医療についても、現状ではその診療実態はほとんど明らかにされておらず、大きな課題となっている。在宅医療の質を定義する難しさの1つには、その対象となる患者において認知症や終末期など自律的判断を行えない状態にある者が多

く、また治癒を目標に据えられない慢性期医療独特の特徴があることから、診療の評価軸を一律に設定しにくいという点などがあげられるが、現状はそれ以前に、どのような疾患・状態にある患者にどのような在宅医療が提供されているかといった実態さえ明らかにされていない。ビッグデータを活用してまずは実態を確認し、特定の保険者の健診・医療・介護データから在宅医療のエビデンス・パフォーマンスを集積していく端緒を開くことが必要かつ最も効率的である。

本研究では第一に、地域(都道府県・市町村)や保険者ごとの性別・年齢別・疾患別医療費および罹患率を把握し、都道府県・全保険者の医療・介護支出目標を精密に設定する推計方法を開発し、増大する国民医療費の適正化対策としての施策導入を目指す。また、国民(被保険者)の健康状態・罹患状況に関して、電子レセプトデータの傷病名・診療行為・医薬品情報、介護認定・給付情報、特定健診等情報を利用してエピソード単位で把握する方法論を開発する。

第二に、集計したパラメーターを活用し、保険者の加入者(被保険者)の各種リスク状態を数値化し、予測モデルを構築する。特に、保険者の財政調整に関してエビデンスに基づくリスク構造調整方式が導入可能であることを明らかにする。また、複数年データを活用して時間経過に伴い新たに作成したデータ項目と医療・介護供給体制の情報とリンクさせて、需要および供給体制のパラメーターを変化させたシミュレーションを行なう。こうしたプロセスを経て、現実のデータの変遷に合わせた予測

が可能な先進的なヘルスケアビッグデータサービスを実現する。

第三に、特定健診データやレセプト情報を活用した後期高齢者の支援金の加算・減算の評価は、当初の計画を実行できなかった。したがって、特定健診・特定保健指導の評価指標や参酌標準は、保険者の保健活動を評価し、且つ医療費適正化のインセンティブを有するものに再検討する必要がある。そこで、NDB レセプト情報と特定健診データを活用した新しい1次・2次・3次予防群の確立、保険者のレセプト情報と介護レセプトを活用し訪問診療をはじめとする在宅医療の実態を明らかにするとともに、疾病予防・医療・介護の連携の実態把握を行う。

B. 研究方法

本研究では、1. NDBの全データ、2. 三重県下の国民健康保険および後期高齢者の医療・介護・健康データ(75万人規模)を利用する。

NDBからは2008年度から2014年度迄の全データを取得する。(2010年度の全NDBデータ(電子レセプト:約15億件、特定健診等データ約2000万件)は取得済み)。

三重県の全市町村(国民健康保険)と後期高齢者医療広域連合2008-203年度のデータも収集する。

保険者からは、NDBと同等のデータに加えて、現在のNDBには蓄積されていない、被保険者マスタ、特定健康診査・特定保健指導の対象者ファイル、介護受給者ファイル、介護給付(費)ファイルを収集する。被保険者マスタには、全加入者の基本的属性(性別・年齢)に加え、保険加入日、保

険資格喪失日などが含まれ、分析対象群を設定するための必要なファイルである。

特定健康診査・特定保健指導の対象者ファイル(国保システム・特定健診結果等情報作成抽出(受診券情報)ファイル)は、健診の未受診者の基本的属性が含まれる。加えて、特定健診受診者CSVファイル、特定健診結果等情報作成抽出(健診結果情報(横展開))ファイル、特定健診結果等情報作成抽出(保健指導情報)ファイルを手に入る。

介護保険については、介護受給者ファイル(医療費の被保険者マスタに相当)に認定審査結果情報等がふくまれる。介護給付費ファイルは、レセプトに相当する。

使用するデータで示したように、本研究で対象とするデータは膨大であり、従来のデータの扱い方では十分な分析が出来ない。そのためビッグデータ解析基盤を構築する必要がある。既に内閣府Firstおよび厚生労働科学研究費補助金において基礎的な基盤を構築しているが、本研究ではデータがさらに増加するため、ビッグデータ解析基盤の増強を行う。

次に、分析の質を高めるためには原データの質を高める必要がある。一人の患者に関わるレセプトが国内約18万の医療施設(病院約8500、診療所約10万、歯科診療所約7万)および調剤薬局(約5.5万)から発生するため、それらの間の紐づけ付けがうまくいかないことがあることが知られている。本研究においてはNDBと直接保険者から入手したデータを比較すること等を通じて問題点を洗い出し、解決策を提示する。さらにビッグデータ解析基盤を用いてこれまでより大幅に迅速な基礎統計

値を産出する体制を整える。

上記の基盤構築、データ検証を行った後に、レセプトデータを分析することで以下に取り組む。

- ・ビッグデータ解析基盤の増強
- ・全 NDB データの検証と質向上
- ・基礎統計
- ・多次元分析
- ・都道府県・二次医療圏・市町村別分析
- ・調整医療費の推計と地域毎医療支出目標額の設定
- ・日本版リスク構造調整方式の確立と財源調整施策の導入の試算
- ・1次・2次・3次予防および予防・医療・介護の連携の実態把握
- ・研究者用データ提供に向けた検討と実装

C. 研究結果

採択額及び研究期間を考慮し、研究内容を重点化した。

三重県の全国健康保険（29市町）と後期高齢者医療広域連合（75歳以上）の最新の医療データ（電子レセプト）・特定健診データ、介護データはの収集が完了した。

ビッグデータ解析基盤への新投資はしないで、セキュリティ環境・ソフトウェア開発、保険者データ（三重）の格納を行った。

基礎統計については、可視化ツールを作成し、ビッグデータシステムに実装した。保険者データについても作成し、以後更新を重ねる予定である。

今後、多次元分析についても同様に進める予定である。連携学会は、日本腎臓学会、

日本糖尿病学会の二つに限定した。すでに定義体の第一バージョンは作成し、保険者のデータを利用して集計を開始した。

都道府県・二次医療圏・市町村別分析、調整医療費の推計と地域毎医療支出目標額の設定、日本版リスク構造調整方式の確立と財源調整施策の導入の試算については、検討会の開催のみとして概念整理と分析手法の検討にとどめ、データの解析は次年度以降実施することとした。

1次・2次・3次予防および予防・医療・介護の連携の実態把握は、三重県の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し、連携の実態把握を行った。

研究者用データ提供に向けた検討と実装は、1件の提供を行った。現在、2件目に着手している。

D. 考察

我々は、最終的な研究成果の政策への反映方法として以下の4つを設定した。

1) 次世代 NDB データ構築と分析による医療費適正化

独自に収集した保険者データと NDB データの突合検証を行うことで、検証、本知見に基づく次世代システム構築への提言を行う。計画通り、検証を実行している。

2) 診療エビデンスの明確化と治療方法の標準化

医学会連携することで、初年度は2疾患の診療実態と地域差を明らかにし、患者に対する適切な医療の提供、評価指標の作成を行っている。

3) 在宅医療（訪問看護等を含む）と介護サービスの連携強化

在宅の医療・介護の給付実態について、地域ごとの特性分析等を実施。将来、NDB データ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による、効率的な医療費・介護費の総合的な適正化につなげる。

4) 医療費支出目標の推計、リスク構造調整の検討として、今後の医療保険制度の在り方に関する検討に寄与する。

今回、研究の採択額及び研究期間を考慮し、研究内容を重点化した。当初の予定とは異なり、データ収集等について想定以上に時間を要した。だが我々は、保険者から独自に収集したデータ（三重県 75 万人規模）を最初に分析、知見を得た上で、次に全日本の NDB データを分析している。これにより抽出時間、処理時間などの見積もりができ、分析の効率化が実現できていると思われる。

大規模データ研究の標準化や教育の充実、人材育成については、医師、医療政策研究者、医療情報研究者達からなる我が班の利用経験に加えて、行政担当者・看護師（保健師）・介護職員等にも資する標準化されたデータの利用に関する教育・人材育成を視野に入れる。H28 年度は、協力保険者の県庁職員、医師会、各市町村国保の保健師に対する教育等を計画している。

今後も引き続き、最先端のビッグデータ基盤と最先端の医療政策・経済等研究を連携させ、研究者、厚生労働省、地域医師会、地方自治体、保険者らステークスホルダーがデータや BI ツールを使い論文化・実態把握・政策提言を行い技術的にも社会的にもメリットが得られるようにする。また、現在の関係者以外にも一定の認識が共有

され、諸課題の解決に向けた社会的な合意を形成するための基盤づくりについても検討を続ける。

E. 結論

今回、研究の採択額及び研究期間を考慮し、研究内容を重点化した。当初の予定とは異なり、データ収集等について時間を要したが、今後も引き続き、最先端のビッグデータ基盤と最先端の医療政策・経済等研究を連携させ、研究者、厚生労働省、地域医師会、地方自治体、保険者らステークスホルダーがデータや BI ツールを使い論文化・実態把握・政策提言を行い技術的にも社会的にもメリットが得られるようにする。

F. 健康危険情報

なし

G. 研究発表

1) 「基本データセットの提供について」第 29 回レセプト情報等の提供に関する有識者会議（平成 28 年 3 月 16 日）、<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000117367.pdf>

2) 満武巨裕：レセプトビッグデータ解析の現状と将来 . 実験医学 第 34 巻第 5 号：799-804, 2016 年

3) 「諸外国の医療ビッグデータ」第 2 回データヘルス時代の質の高い医療の実現に向けた有識者検討会（平成 28 年 5 月 23 日）、<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/000012>

5318.pdf

H.知的財産権の出願・登録状況
該当なし

超高速レセプトビッグデータ解析基盤を利用した医療需要の把握・整理

研究代表者 満武 巨裕

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、副部長

研究要旨

本研究の目的は、超高速レセプトビッグデータ解析基盤を利用することにより、レセプト情報・特定健診等情報データベース（以下、NDB）のデータ（年間370億レコード）及び保険者から収集した75万人規模の医療・介護データを利用して、地域（都道府県・市町村）や保険者ごとの性別・年齢別・疾患別医療費および罹患率を把握し、都道府県・全保険者の医療・介護支出目標を精密に設定する推計方法を開発し、増大する国民医療費の適正化対策としての施策導入を目指す。また、複数年データを活用して時間経過に伴い新たに作成したデータ項目と医療・介護供給体制の情報とリンクさせて、需要および供給体制を把握し、現実のデータの変遷に合わせた予測が可能な先進的なヘルスケアビッグデータサービスを実現する。

はじめに、ビッグデータ解析基盤の増強を行う。物理的には、東京大学構内にビッグデータ解析基盤およびソフトウェア開発用のセキュリティルームを設置する。ハードウェア構成は、サーバ装置群、専用パーティションに設置したクライアント端末構成とする。サーバ装置群は、厚生労働省のガイドラインに基づくレセプト情報等の堅牢な管理と、レセプト情報等に対する解析クエリの実行を担う。クライアント端末は、研究者へのインターフェースであり、レセプト情報等の管理ならびに解析の命令をサーバ装置群に発行し、また、その結果を取得して出力する。

NDBデータの集計・分析としては、疾患毎罹患率、医療費の傾向などの情報を都道府県・二次医療圏別に作成する。次に、患者数が多く大量の医療資源を消費している生活習慣病について、「日本腎臓学会」等と連携し、これらの疾患管理を評価するための指標を同定する。さらに、保険者から収集したデータを利用して、在宅医療（医療保険訪問看護を含む）と介護サービスがどのように併用され、サービスに要する費用等の実態を分析する。特に、在宅医療介護連携の実態を明らかにし、同じ医療や介護が必要とされる高齢者の入院や外来受診、在宅医療の地域比較を行い、現在進行中の医療介護総合改革の適切な推進等に資するエビデンスに基づく政策の推進を図る。

三重県の全ての国民健康保険と後期高齢者のデータ利用し、供給データ（医療機

関・介護施設・介護サービス事業所の基本的属性)と組み合わせることで、医療と介護サービスの連携でどのような使われ方をしているか、個人単位で可視化することができた。また、ステークホルダー(主に保険者と地域医師会)用に4種類のデモを作成した。デモは、市町ごとに病名が選択でき、入外来別の患者数と医療費、どこの医療機関を受診したかを可視化した。その結果、地域により疾患別受診状況に差異があることが明らかになった。特に、二次医療圏を超える受診実態が示された。介護サービスの利用についても、地域ごとの特性が示された。本研究により、種々のデータを組み合わせ、医療や介護の実態を可視化することで、現在進行中の医療介護総合改革の適切な推進等に資するエビデンスを作成できることを示した。

分担研究者

辻 哲夫・医療経済研究機構・副所長
喜連川 優・東京大学生産技術研究所・教授
合田 和生・東京大学生産技術研究所・特任准教授
伏見 清秀・東京医科歯科大学医療政策学・教授
辻 一郎・東北大学大学院医学系研究科・教授
藤森 研司・東北大学大学院・医学系研究科・教授
興梠 貴英・自治医科大学・企画経営部医療情報部・准教授
中島 直樹・九州大学病院メディカルインフォメーションセンター・教授
石川ベンジャミン光一・国立がん研究センターがん対策情報センターがん統計研究部・室長
飯島 勝矢・東京大学高齢社会総合研究機構・准教授
吉江 悟・東京大学医学部在宅医療学拠点・特任研究員
関本 美穂・大阪府済生会吹田病院麻酔科・医長
印南 一路・慶應義塾大学総合政策学部・教授

高橋 邦彦・名古屋大学大学院医学系研究科・准教授

奥村 泰之・医療経済研究機構・主任研究員

清水沙友里・医療経済研究機構・主任研究員

A. 研究目的

本研究の目的は、超高速レセプトビッグデータ解析基盤を利用することにより、レセプト情報・特定健診等情報データベース(以下、NDB)のデータ(年間370億レコード)及び保険者から収集した75万人規模の医療・介護データを利用して、地域(都道府県・市町村)や保険者ごとの性別・年齢別・疾患別医療費および罹患率を把握し、都道府県・全保険者の医療・介護支出目標を精密に設定する推計方法を開発し、増大する国民医療費の適正化対策としての施策導入を目指す。また、複数年データを活用して時間経過に伴い新たに作成したデータ項目と医療・介護供給体制の情報とリンクさせて、需要および供給体制を把握し、現実のデータの変遷に合わせた予測が可能な先進的なヘルスケアビッグデータサービスを実現する。

B. 研究方法

はじめに、ビッグデータ解析基盤の増強を行う。物理的には、東京大学構内にビッグデータ解析基盤およびソフトウェア開発用のセキュリティルームを設置する。ハードウェア構成は、サーバ装置群、専用パーティションに設置したクライアント端末構成とする。サーバ装置群は、厚生労働省のガイドラインに基づくレセプト情報等の堅牢な管理と、レセプト情報等に対する解析クエリの実行を担う。クライアント端末は、研究者へのインターフェースであり、レセプト情報等の管理ならびに解析の命令をサーバ装置群に発行し、また、その結果を取得して出力する。

NDBデータの集計・分析としては、はじめに、高速保険レセプト解析基盤の実行性能比較を行った。4種類のクエリ（クエリ1：都道府県ごとの急性心膜炎患者数（対人口比率）、クエリ2：口唇の悪性新生物の都道府県ごとの平均診療費、クエリ3：呼吸器結核に関する年齢ごとのレセプト件数、クエリ4：コレラ患者の都道府県ごとの平均診療費）を実行して、検索時間を比較した。次に、薬剤の処方傾向の分析として、処方件数および併用薬や併用禁忌の実態を分析した。

患者数が多く医療資源を消費している生活習慣病について、「日本腎臓学会」等と連携し、患者数、疾病がどのような経過をたどるのか（時系列分析）、疾病に対してどのような診療が行われているか（診療パターン）といった分析を行い、新しい診療ガイドラインを作成するためのエビデンス、疾病を管理するための評価指標の作

成につなげる。

保険者から収集したデータを利用して、在宅医療（医療保険訪問看護を含む）と介護サービスがどのように併用され、サービスに要する費用等の実態を分析する。特に、在宅医療介護連携の実態を明らかにし、同じ医療や介護が必要とされる高齢者の入院や外来受診、在宅医療の地域比較を行い、現在進行中の医療介護総合改革の適切な推進等に資するエビデンスに基づく政策の推進を図る。

高速保険レセプト解析基盤の実行性能比較では、従来型と比べて約 100 倍となる高速化を実現できた。

薬剤の処方傾向の分析では、主要抗うつ薬と湿布薬（ケトプロファン、ロキソプロフェンナトリウム水和物処方件数）の年間点数合計と処方日数（365 日を超える実態把握）、高血圧治療薬の処方件数と併用薬・併用禁忌の実態を示した。

生活習慣病の分析結果は、例えば慢性腎臓病（CKD）の患者数について約 22.5 万人を抽出した。透析患者については、診療行為名称を特定したところ、NDB から推計した全国の透析患者数は、約 22 万人（生活保護は含まず）であった。また、腎移植後に継続的に透析を行っているの患者が、298 人存在することを確認した。

三重県の全ての国民健康保険と後期高齢者のデータ利用し、供給データ（医療機関・介護施設・介護サービス事業所の基本的属性）と組み合わせることで、医療と介護サービスの連携でどのような使われ方をしているか、個人単位で可視化することができた。具体的には、複

数年データを用いて、生活習慣病の高血圧患者の推移を観察し、診断済群（投薬なし、合併症なし）、服薬治療群（投薬あり、合併症なし）、合併症群（投薬なし、合併症あり、透析あり）、透析治療群（投薬あり、合併症あり、透析あり）毎に継時的変化を示した。

保険者から収集したデータを利用したサービスに要する費用等の実態を分析については、ステークスホルダー（主に保険者と地域医師会）用に4種類のデモを作成した。デモは、市町ごとに病名が選択でき、入外来別の患者数と医療費、どこの医療機関を受診したかを可視化した。その結果、地域により疾患別受診状況に差異があることが明らかになった。特に、二次医療圏を超える受診実態が示された。介護サービスの利用についても、地域ごとの特性が示された。

D. 考察

高速保険レセプト解析基盤の実行性能比較から、従来型と比べて約100倍となる高速化を実現でき、処方傾向の分析等において応用例を示すことができた。レセプト病名については、例えばステージ情報付の傷病名等を有する入力をしている医療機関が少ない、移植のレセプトは紙で請求されることが多いといった理由から、NDBを利用して全数を把握できない可能性がある。しかし、病名と処置、処方を組み合わせて対象疾患を選定することで治療中の患者の把握率が向上する可能性がある。

今回、ビッグデータ基盤と医療政策・経済等研究を連携させ、研究者、地域医師会、地方自治体、保険者らステークスホルダーが実態把握・政策提言を行い社会的にもメリットが得られるようにデータやBIツールを開発したところ、種々のデータを組み合わせることで、現在進行中の医療介護総合改革の適切な推進等に資するエビデンスを作成できることを示すことができた。今後も、現在の関係者以外にも一定の認識が共有され、諸課題の解決に向けた社会的な合意を形成するための基盤づくりが必要であり、現在進行中の医療介護総合改革の推進等に資するエビデンスを作成できることを示した。

E. 結論

種々のデータを組み合わせて医療や介護の実態を可視化することで、現在進行中の医療介護総合改革等の推進等に資するエビデンスを作成できることを示した。

F. 健康危険情報

該当なし

G. 研究発表

該当なし

H. 知的財産権の出願・登録状況

該当なし

厚生労働科学研究・戦略研究
健康医療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

レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析および超高速レセプトビッグデータ解析基盤の整備

研究実施計画書

内容

0 . 研究の名称	24
1 . 研究目的	24
2 . 研究の背景と意義	24
3 . リサーチクエスチョン：クリニカルクエスチョンと医療政策的・経済的クエスチョン	27
4 . 対象（データ）	30
変更点：NDB データは、予定通り申請済み。既に提供を受けている、本研究方法「9)研究者用データ提供に向けた検討と実装」については、データ延長手続きを行っており 2016 年末まで利用可能。既にデータセットの提供を試行提供（東京大学）しており、今後 2 件目（早稲田大学）に対応する。	31
5 . 研究の方法：必要性、目標、手順、人員・組織、スケジュール、成果	32
1) ビッグデータ解析基盤の増強	32
2) 全 NDB データの検証と質向上	34
3) 基礎統計	37
4) 多次元分析(クリニカルクエスチョン).....	38
5) 都道府県・二次医療圏・市町村別分析(Small Area Analysis: SAA)	40
6) 調整医療費の推計と地域毎医療支出目標額の設定	41
7) 日本版リスク構造調整方式の確立と財源調整施策の導入の試算	43
8) 1 次・2 次・3 次予防および予防・医療・介護の連携の実態把握.....	45
9) 研究者用データ提供に向けた検討と実装.....	47
6 . スケジュール（研究期間）	48

0．研究の名称

レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析
および超高速レセプト・ビッグデータ解析基盤の整備

1．研究目的

本研究の目的は、申請者がこれまで構築してきた高速レセプト・ビッグデータ解析基盤を更に発展させることにより、以下の医療の需要・供給、質、コストが国・地域・医療機関レベルで即座に可視化できるサービスを確立する。

- 1) 4 医学会と連携した NDB レセプトデータから生活習慣病の発症数・合併症数・重症群の定義を確立を通し、わが国の生活習慣病・癌等の複数疾患の診療の実態を明らかにし、診療エビデンス・パフォーマンスの分析・公開
- 2) 国・都道府県・市町村レベルで、医療費・介護費支出の目標を設定する方法論の確立
- 3) 日本版リスク構造調整方式の確立と財源調整施策の導入の試算
- 4) NDB レセプト情報と特定健診データを活用した新しい 1 次・2 次・3 次予防群の確立
- 5) レセプト情報と介護レセプトを活用し訪問診療をはじめとする在宅医療の実態を明らかにするとともに、疾病予防・医療・介護の連携の実態把握

変更点

- 1)は、4 学会から 2 学会（日本糖尿病学会、日本腎臓学会）に規模を縮小した。
- 2)は、検討会の開催のみとして、今年度中に 1 回の開催予定である。集計結果を基に検討する必要があるため、2 月初旬を予定している。
- 3)は、2)と同様である。

2．研究の背景と意義

（背景）

医療分野でのビッグデータの一つに、2009 年度から日本の全患者の保険診療情報を蓄積している国家規模のデータベースがある。これはレセプト情報・特定健診等データベース（以下 NDB）と呼ばれ、このデータを分析し医療政策に利用することが医療の質の改善や医療費適正化の切り札となることが期待されている。

医療費の適正化を目的として、これまで幾多の医療制度改革が行われてきた。例えば 2008 年度には、特定健康診査・特定保健指導（メタボ健診）や後期高齢者医療制度が導入された。メタボ健診は生活習慣病の一次予防により将来的な医療費を削減する目的で導入されたが、これが本当に生活習慣病を予防し、医療費を削減する効果を持つのかに関するエビデンスは未だ示されていない。

エビデンスを示すことが困難であった原因の一つは、これまで診療の実態や診療行為のアウトカム、医療のコストなどを包括的に分析するための国家規模のデータが存在しなかったことである。レセプト情報は保険者に関わりなく同じフォーマットで作成されるため、この目的にかなうデータとして有望視されてきたが、全国規模のデータを集めるシステム

が長らく存在しなかったため、その分析は一部の研究者や研究機関による狭いテーマに限定されてきた。NDB には、全医療機関で提供された医療をカバーするレセプト情報、特定健診情報が収載され、2013 年度の社会保障制度改革国民会議の報告書（内閣官房）においても「国が保有するレセプト等データの利活用の促進にも不可欠である」と大きな期待が寄せられている。

また、今後の超高齢社会に向けて目指すべき地域包括ケアシステム構築の中で鍵となる在宅医療についても、現状ではその診療実態はほとんど明らかにされておらず、大きな課題となっている。在宅医療の質を定義する難しさの 1 つには、その対象となる患者において認知症や終末期など自律的判断を行えない状態にある者が多く、また治癒を目標に据えられない慢性期医療独特の特徴があることから、診療の評価軸を一律に設定しにくいという点などがあげられるが、現状はそれ以前に、どのような疾患・状態にある患者にどのような在宅医療が提供されているかといった実態さえ明らかにされていない。ビッグデータを活用してまずは実態を確認し、特定の保険者の健診・医療・介護データから在宅医療のエビデンス・パフォーマンスを集積していく端緒を開くことが必要かつ最も効率的である。

（意義）

ビッグデータの利用環境が整備される中、重要課題はビッグデータの具体的な活用事例を提示することである。ビッグデータがその価値を発揮するには明確に定義された「検証すべき仮説」が必要であり、仮説思考に基づかないビッグデータは単なる大量のデータにすぎず、戦略なきデータ解析から得られる情報は往々にして利用価値の低いものになりがちである。そこで本研究ではまず、NDB やそれに準ずるビッグデータ（介護給付情報、被保険者マスタ情報等）の分析が、医療の需要と供給、質、コストに関してどのような情報を提供できるのかを検討する。ビッグデータの解析から得られた情報を社会に還元する方策について具体的事例を示し、そのリスクと便益を検証することは、NDB あるいはヘルスケア・ビッグデータの管理・運営ならびに利活用のあり方に資する有用な知見をもたらす。

申請者はこれまでに我が国国民全体を対象として年間約 400 億件のレコードをリアルタイムに解析し可視化し得る高速レセプト・ビッグデータ解析基盤を構築してきた。国全体を俯瞰可能なレセプト解析システムは、世界的に見ても他に類がなく、本研究においては、当該基盤の更なる発展と広汎な利活用を通じて、NDB 全データの精度検証情報の公開、ダッシュボード機能等を利用した診療情報の集計結果の公表、研究者用データセットの整備、更なる研究利用の振興と裾野を広げることにより、我が国のヘルスケア・ビッグデータの利活用を世界トップレベルに引き上げることができるものと確信している。

本研究の政策的な意義について、以下に述べる。内閣官房（社会保障制度改革推進本部）は「医療・介護情報の活用による改革の推進に関する専門調査会」において、医療費を適正化するための施策として都道府県ごとに医療支出目標を設定することを検討している。ヘルスケア・ビッグデータの解析により地域の医療の需給バランスが明らかになれば、医療支出目標の設定並びに適切な医療資源配置について具体的な方法論を提示・確立できる。

過去に医療制度改革推進本部（厚生労働省）において性別と年齢によるリスク構造調整手法が保険者の財源調整方法が検討されたが、実際の推計は行われなかった。2015 年の

通常国会において国民健康保険の運営を市町村から都道府県へ移管する法案が提出される予定だが、移管に当たり全国知事会等は財政基盤強化策・政府部門の国費投入の規模の提示（エビデンス）を求めている。また、都道府県の運営にあたっては、医療費適正化の実施が各都道府県に課せられること、そのためにリスクを加味した適正な医療費の目安や予測の方法が求められる。すなわち現在の保険者間の財源調整は、前期高齢者では「前期高齢者」と「それ以外」の区分の加入率のみに基づく調整であり、後期高齢者では医療費が発生した後の事後的な調整に過ぎない。各都道府県が保険者として医療費の適正化に取り組むためには、保険者毎のリスクプロファイルとそれに基づく適正医療費の目安、さらには公平性の高い財政調整措置の方法論が求められる。そこで、本研究より提示する諸外国のモデルに基づいた地域カテゴリ別のリスク構造の推計結果を、全国町村会、全国市町会、全国知事会、地域保険者、都道府県国保連合会、各自治体の首長へ情報提供することで、国民健康保険広域化の検討材料とする。

理論的には医療費は患者の属性と医学的な要因だけで決定されるべきだが、実際には医療費は医師による診療パターンの違いや医師密度、医療機関の利用のしやすさ、医療機器や設備の配置などさまざまな要因の影響を受ける。また既存の研究により、医療費のバラツキには、医学的な要因よりもその他の医療提供側の要因や社会的要因の寄与がはるかに大きいことが示されている。このような場合、医療費を適正化するには医療費が地域により異なることを示すだけでは不十分で、そのバラツキが何から生じるのかが明らかにされない限り適正化の対策はあり得ない。また、在宅医療との連携という今後の政策転換においても、在宅医療の実態は明らかになっていない。即ち、ビッグデータの分析の範囲は医療費に限らず、個々の疾患、診療行為、あるいは診療パターンと地域の医師密度や医療資源との関係など多角的な分析が必要となる。ヘルスケア・ビッグデータとさまざまな政府統計や医療機関の持つデータなどと組み合わせることにより、より具体的な医療費・介護費適正化の対策が立てられ、特にわが国における診療の標準化の取り組みは緒に就いたばかりであり、今後診療パターンや治療の標準化の実態を地域レベル、医療機関レベルで評価することは、医療の質を改善し、在宅医療をはじめとする医療資源を将来に効果的に配分するための欠かせない取り組みと言える。

診療の質の評価を改善に結びつける方策は大きく2つある。ひとつは選択による改善であり、診療情報の分析結果を公表し、患者が医療機関を選択する際の参考としてもらうやり方で、質の低い医療機関が淘汰されることによる質の改善である。もうひとつは医療者による自主的な改善であり、これは全ての医療者の自主的な努力により全体的に質が改善されることを目的とする。医療の質の評価と結果の公表が古くから行われてきた米国をはじめとする欧米諸国の研究によると、診療パフォーマンスの公表が患者が医療機関を選択する際に利用されることはほとんどないが、医療者は公表結果を非常に気にするため、結果的に診療の標準化や質の改善等がみられるという。わが国では診療パフォーマンスの評価は始まったばかりであり、データの利用可能性の点からこれまで専らDPCデータによる入院診療を対象とした評価が行われてきた。これに対して医療費の約半分を費やす外来診療の分析は、これまでほとんど行われてこなかったが、生活習慣病の管理を中心とするプライマリ・ケアおよび在宅医療における診療の標準化は喫緊の課題であり、これらの診療を評価する方法論の開発や、結果のフィードバックを通じた診療の質の改善などは、わが国の診療の質の改善と政策転換に大きな威力を発揮するはずである。

ただし、NDB データを使って医療機関を特定することは禁止されているために、提供単

位としては県・二次医療圏レベルの情報提供（フィードバック）となる。ただし、本研究では500万人規模の保険者をフィールドとして、医療・介護・健康データの取得を済んでいる。保険者には、市町村および医療機関別の情報提供（フィードバック）を行うものである。

特定健診・特定保健指導は、メタボリックシンドロームが生活習慣病の大きな原因であるという学説に基づき、内臓脂肪を減らすことで生活習慣病を予防し、将来的に医療費削減につなげるという目論見で2008年に開始された。特定健診・特定保健指導の実施状況は、保険者の生活習慣病予防に対する取り組みとして評価され、「高齢者の医療の確保に関する法律」により各保険者が分担する後期高齢者支援金の調整に利用されることが規定されていた。制度導入当初は特定健診実施率、特定保健指導実施率、メタボリックシンドロームの該当者・予備軍の減少率の評価に基づいて、支援金の加算・減算が行われる予定であったが、2013年度の支援金の加算・減算にこれらの評価は利用されていない。現在、保険者へのペナルティの適切性や、疾病予防や医療費削減の効果に関して、特定健診・保健指導への疑問が呈されている。したがって、特定健診・特定保健指導の評価指標や参酌標準は、保険者の保健活動を評価し、且つ医療費適正化のインセンティブを有するものに再検討する必要がある。そこで第三として、新たな特定健診・特定保健指導の評価指標の作成、特定保健指導の対象群や医療機関との連携を促進する群等の1次・2次・3次予防群の設定を行う。

本研究の成果は報告書およびWebページで公開する。特に、データの公開が医療に与える影響としては、医療機関は公表結果に注意を払い、結果的に診療に一定の変化がみられるとの報告がある。そのために、NDBを利用した診療行為実績の公表による医療費適正化効果についても検討する。

当然のことながら、本研究を進展によって、NDBデータの解析による社会的な便益が広く理解されることにより、現状においては認められていないNDBデータの解析に更なる理解が得られることも想定できる。

申請者のリーダシップの下で本研究班は、厚生労働省・厚生科学課、厚生労働省・保険局・医療介護連携政策課保険システム高度化推進室、戦略研究企画・調査委専門検討会と定期的なモニタリングと評価を頂きながら、機動的かつ柔軟な研究推進に努める次第である（NDBデータを使用した結果の公表は、必ず厚生労働省・保険局・医療介護連携政策課保険システム高度化推進室の許可を得る）。

改善点：本研究で利用するNDBデータの申請は、第5回レセプト情報等の提供に関する有識者会議審査分科会（7月31日開催）にて申請を行い、審査継続として座長の承認後、データ授受の予定である。

3. リサーチクエスチョン：クリニカルクエスチョンと医療政策的・経済的クエスチョン

NDBのレセプトデータの特徴は、DPCデータとは異なり、異なる医療機関を受診してもデータのリンケージが可能であり、転院した場合や複数医療機関を受診している場合も追

跡調査が可能である。また、NDB は DPC 病院に限らず全ての医療機関、調剤薬局からのデータを含んでいる。一方で病名が十分整理されておらず臨床的な観点からの信頼性が高くなく、死亡イベントの把握がしづらいこと等の限界もある。特定健診では血液検査データなども含まれるが、受診者は全被保険者の一部である。本研究では、こうした限界を踏まえつつ、レセプトデータの特徴を生かして様々なリサーチクエスチョンに答えていく。

クリニカルクエスチョン

I. 生活習慣病患者の実態とアウトカム

これまでコホート研究などの個別の研究において生活習慣病の正確な患者数の把握、治療の試行件数（例えば透析者）は推計できているものの、毎年の全日本データについては明らかになっていない。そこで、生活習慣病を有する患者グループを同定する方法論（傷病名、診療行為、医薬品コードなどから定義）を確立し、同じリスク・プロファイルを持つ患者に対する診療パターンや医療費が、地域や医療機関の特性によりどのように異なるかを検討する。地域の医師密度や医療機関密度、医療設備の配置、診療の密度が患者アウトカムや生活習慣病における治療成績（疾患コントロールの状態や合併症予防）との間に関係の実態を経年的に把握し、診療の適切性や日本の医療資源配分の適切性を明らかにする。

II. 心臓カテーテル治療を受けた患者のアウトカム

これまでコホート研究などの個別の研究において虚血性心疾患の内服薬とアウトカムとの関係をみた研究は報告されている (Circ J. 2007 Dec;71(12):1835-40.)。また、DPC データを用いて新しいデバイスが治療方法に与えた影響をみた研究はある (Circ J. 2010 Aug;74(8):1635-43.)。しかし、全国規模で心臓カテーテル治療(PCI)を行った患者に対して追跡調査を行った例はない。レセプトデータには病名の不確実性、死亡の把握が出来ないなどの問題はあるが、再入院やデバイスを用いた治療、手術の把握は可能であり、そうしたイベントをアウトカムとして PCI を施行された患者を縦断的に追跡することが可能である。全国的な実態把握および、アウトカムにどのような因子が影響を与えるかを分析し、利益のある治療方法について、さらに医療費分析を行い、より少ない医療費でよいアウトカムに結びつけるにはどのようにすればよいかについても知見を得る。

III. 人工透析患者のシャントに対する PTA 施行の実態把握

慢性腎不全患者に対する人工透析治療が発達し、長期生存する患者が増加した。それに伴い狭窄や血管荒廃などのシャントトラブルも増加している。従来はこうした場合には外科治療しか保険適応が認められていなかったが、2012年に既存のシャントに対してカテーテルを用いた拡張術(PTA)が保険適応となり、現在施行件数が大幅に増えてきている。しかし、全国的な PTA 施行の実態やそれに及ぼす因子、かかる医療費の分析はされていない。レセプトデータを用いて分析を行い、こうした実態や因子を明らかにすることにより、適切な PTA 施行および最適な費用対効果が得られる治療方法を明らかにする。

IV. 脳梗塞患者に対して t-PA(アルテプラゼ)を使用している患者数の把握および治療後のアウトカムの関係

日本においては2005年から発症早期(4時間30分以内)脳梗塞患者に対してt-PA血栓溶解療法が使えるようになった。しかし、病院側に治療を行える体制があるかということを含めて実際に行うことができるかどうかには複数の要因が関わってくる。レセプトデータを分析することで、まず脳梗塞に対するt-PA治療の実態および、施行にかかわる因子を明らかにし、よりよいアウトカムを得るための要因を推定する。

V. 種々の癌に対する治療方法の実態把握およびアウトカムの関係

近年、癌に対する治療方法は高度化、多様化している。具体的には、放射線治療分野ではたとえば癌の位置を高精度に測定した上で、照射も癌のみに限局して行えるような高度な機器を導入して行うようになっている。また抗癌剤は従来の化学療法に加えて、分子標的治療が次々に出てきている状況である。さらに手術も低侵襲手術の施行が拡大している。まず、レセプトデータを分析することでこうした癌治療の全国的な実態把握を行い、さらに、再治療等をアウトカムとして従来治療法と新規治療法を比較し、費用対効果分析を行うなど様々な分析を行う。

医療政策的・経済的クエスチョン

VI. 都道府県・二次医療圏・市町村の医療計画及び医療費適正化計画に資する情報の確立

衛生統計学、疫学などの公衆衛生の分野では、地域の状況を表す指標としては、年齢分布の違いを調整した、年齢調整罹病率、標準化罹病率などが良く利用されている。しかし、従来の指標は、人口の地域変動に基づく標本誤差の影響を強く受け、人口の小さい地域の指標のバラつき、稀少疾患のわずかな変化が見かけ上の指標を大きく変化させる問題がある。従来の統計指標の問題の解決策として、空間疫学・経験的ベイズ推定量を導入して、都道府県・二次医療圏・市町村の医療計画及び医療費適正化計画に実際に活用できる提供情報を確立する。

VII. 医療費の予測モデルの構築

増え続ける日本の医療費を適正化するための施策として、都道府県ごとに医療支出目標を設定することが検討されている。また、2015年の通常国会へ法案提出を目指している国民健康保険(市町村)の運営を都道府県へ移管するにあたり、全国知事会等が財政基盤強化策・政府部門の国費投入の規模の提示(エビデンス)を求めている。そこで、NDBデータを利用して外来および入院医療費に影響を及ぼす因子(年齢・性別・併存疾患等)医療供給体制などの外部のセカンダリーデータを活用して地域後の医療費の予測モデルを構築し、全国の都道府県および市町村へのデータ提供を行う。

VIII. 日本版リスク構造調整方式の確立と財源調整施策の導入の試算

現在の保険者間の財源調整は、前期高齢者では「前期高齢者」と「それ以外」の区分の加入率のみに基づく調整であり、後期高齢者では医療費が発生した後の事後的な調整に過ぎない。したがって、保険者単位のリスク構造を基に各保険者が実態を把握し医療費適正化に取り組む誘因をもった、財政調整措置等の問題を検討しなければならない。NDBおよび保険者の医療費データを用いて、日本版リスク構造調整方式の確立と財源調整施策の導入の試算を行う。

IX. 新しい1次・2次・3次予防群の確立、疾病予防・医療・介護の連携の実態把握

特定健診データやレセプト情報を活用した後期高齢者の支援金の加算・減算の評価

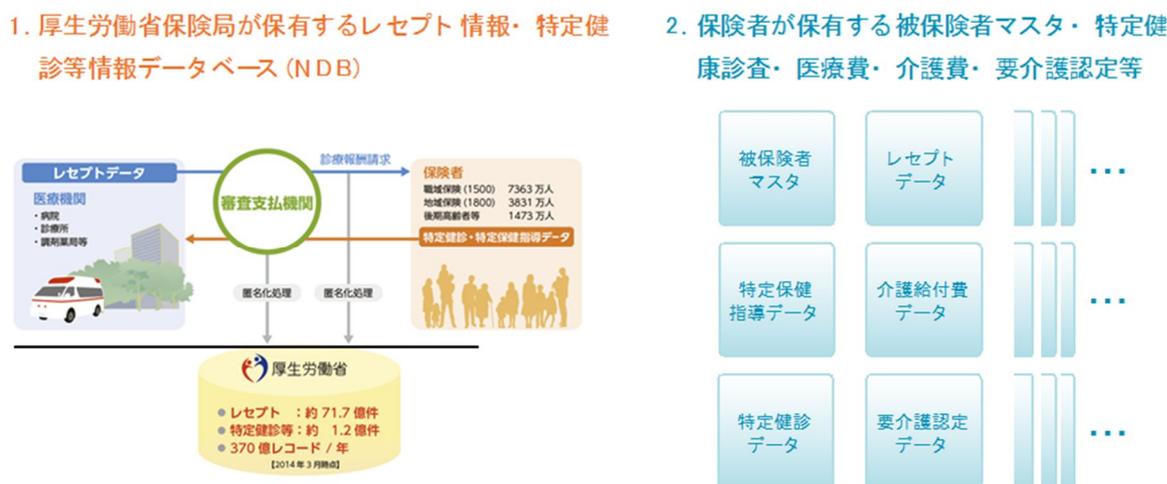
は、当初の計画を実行できなかった。したがって、特定健診・特定保健指導の評価指標や参酌標準は、保険者の保健活動を評価し、且つ医療費適正化のインセンティブを有するものに再検討する必要がある。そこで、NDB レセプト情報と特定健診データを活用した新しい1次・2次・3次予防群の確立、保険者のレセプト情報と介護レセプトを活用し訪問診療をはじめとする在宅医療の実態を明らかにするとともに、疾病予防・医療・介護の連携の実態把握を行う。

変更点：「VIII.日本版リスク構造調整方式の確立と財源調整施策の導入の試算」については、本年度は行わないものとする。「VII. 医療費の予測モデルの構築」については、検討のみとする。

4. 対象（データ）

本研究では、1.NDB の全データ、2.神奈川県および三重県下の全国健康保険の医療・介護・健康データ（500万人規模）を利用する。

図表. 利用する二つのデータソース



レセプト情報・特定健診等情報データベース(NDB)

NDBからは2008年度から2014年度迄の全データを取得する。(2010年度の全NDBデータ(電子レセプト：約15億件、特定健診等データ約2000万件)は取得済み)

NDBのデータは、「高齢者の医療の確保に関する法律」の第16条を2根拠として収集され、2009年度から2014年度までの全電子レセプトデータ、2008年度からの特定健診・特定保健指導データが含まれる。2014年10月時点では、レセプトは約83.5億件、特定健診・特定保健指導は約1.2億件分となっている。

電子レセプトデータは業務データであり、医療費機関から保険者に送付される形式は、保険診療報酬点数情報が羅列されたCSVファイルである。NDBではこのCSV形式の電

子レセプトを、複数のレコードに分割して保管している。具体的には、医科レセプト情報、DPCレセプト情報、調剤レセプト情報、歯科レセプト情報の4つである。医科レセプト情報は、医療機関情報レコードファイル(IR)、レセプト共通レコードファイル(RE)、保険者レコードファイル(HO)、公費レコードファイル(KO)、国保連固有情報レコードファイル(KH)、傷病名レコードファイル(SY)、診療行為レコードファイル(SI)、医薬品レコードファイル(IY)、特定器材レコードファイル(TO)、コメントレコードファイル(CO)、日計表レコードファイル(NI)、症状詳記レコードファイル(SJ)、臓器提供医療機関情報レコードファイル(TI)、臓器提供者レセプト情報レコードファイル(TR)、臓器提供者請求情報レコードファイル(TS)の15に分割されている。DPCレセプト情報、調剤レセプト情報、歯科レセプト情報についても同様に分割されており、これらを総じて、総レコード数は年間約400億レコードにのぼる規模を誇る。

特定健康診査・特定保健指導は、2008年4月から40歳以上の全国民を対象に導入された制度であり、データ項目は、生活習慣等に関する問診票の項目、腹囲、BMI、血圧、血糖・脂質といった血液検査の結果情報、特定保健指導に関連する情報である。特定健診情報は、基本情報レコードファイル、セクション情報レコードファイル、健診結果・問診結果情報ファイル、詳細情報レコードフォーマットに分割されている。保健指導情報は、基本情報レコードファイル、券面情報レコードファイル、セクション情報レコードファイル、エントリー情報レコードファイル、保健指導結果情報レコードファイルに分割されている。

変更点：NDBデータは、予定通り申請済み。既に提供を受けている、本研究方法「9」研究者用データ提供に向けた検討と実装については、データ延長手続きを行っており2016年末まで利用可能。既にデータセットの提供を試行提供（東京大学）しており、今後2件目（早稲田大学）に対応する。

保険者が保有するデータ（500万人規模を既に収集）

神奈川県および三重県の全市町村（国民健康保険）と後期高齢者医療広域連合の協力が確定しており、約500万人規模である。データは、既に2008-2012年度が収集済みである。そのために2013年以降のデータも収集する。また、神奈川県と三重県以外の保険者に対しても対象を広げる。現時点では、静岡県、奈良県、兵庫県等と交渉を開始している。

保険者からは、NDBと同等のデータに加えて、現在のNDBには蓄積されていない、被保険者マスタ、特定健康診査・特定保健指導の対象者ファイル、介護受給者ファイル、介護給付（費）ファイルを収集している。

被保険者マスタには、全加入者の基本的属性（性別・年齢）に加え、保険加入日、保険資格喪失日などが含まれ、分析対象群を設定するための必要なファイルである。

特定健康診査・特定保健指導の対象者ファイル（国保システム・特定健診結果等情報作成抽出（受診券情報）ファイル）は、健診の未受診者の基本的属性が含まれる。加えて、特定健診受診者CSVファイル、特定健診結果等情報作成抽出（健診結果情報（横展開））ファイル、特定健診結果等情報作成抽出（保健指導情報）ファイルを入手する。

介護保険については、介護受給者ファイル（医療費の被保険者マスタに相当）に認定審

査結果情報等がふくまれる。介護給付費ファイルは、レセプトに相当する。

変更点：三重県の全国健康保険（29市町）と後期高齢者医療広域連合（75歳以上）の最新の医療データ（電子レセプト）・特定健診データは収集が完了。介護データは、12月28日をデータ授受日（於：三重県国保連合会）としている（神奈川県データは、予算的な制約のため介護データを新たに入手することはない）。

5. 研究の方法：必要性、目標、手順、人員・組織、スケジュール、成果

使用するデータで示したように、本研究で対象とするデータは膨大であり、従来のデータの扱い方では十分な分析が出来ない。そのためビッグデータ解析基盤を構築する必要がある。既に内閣府 First および厚生労働科学研究費補助金において基礎的な基盤を構築しているが、本研究ではデータがさらに増加するため、ビッグデータ解析基盤の増強を行う（1）ビッグデータ解析基盤の増強）。

次に、分析の質を高めるためには原データの質を高める必要がある。一人の患者に関わるレセプトが国内約18万の医療施設（病院約8500、診療所約10万、歯科診療所約7万）および調剤薬局（約5.5万）から発生するため、それらの間の紐づけ付けがうまくいかないことがあることが知られている（文献：平成24年度～平成25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」、研究代表者・満武巨裕）。本研究においてはNDBと直接保険者から入手したデータを比較すること等を通じて問題点を洗い出し、解決策を提示する。さらにビッグデータ解析基盤を用いてこれまでより大幅に迅速な基礎統計値を産出する体制を整える（2-3）全NDBデータの質向上および基礎統計値の計算）。

上記の基盤構築、データ検証を行った後に、レセプトデータを分析することでさまざまな医療経済、医療政策また臨床的なクエスチョンに答えていく（4）以降）。

1) ビッグデータ解析基盤の増強

（必要性と目標）申請者がこれまでに構築してきた「高速レセプトビッグデータ解析基盤」を更に発展・進化させた「超高速レセプトビッグデータ解析基盤」を構築する。超高速レセプトビッグデータ解析基盤は、従来の6倍規模（1億2700万人規模を6年分）となるNDB（現在2010年度分のみ保持）、保険者から500万人規模のデータを縦横無尽に集計・分析することができるITプラットフォームとなる（協力保険者の増加も想定し1000万人以上でも対応できる基盤とする）。

（手順）物理的には、東京大学構内に約30平方メートルビッグデータ解析基盤およびソフトウェア開発用のセキュリティルームを設置する。ハードウェア構成は、サーバ装置群、専用パーティションに設置したクライアント端末構成とする。サーバ装置群は、厚生労働省のガイドラインに基づくレセプト情報等の堅牢な管理と、レセプト情報等に対する解析クエリの実行を担う。クライアント端末は、研究者へのインターフェースであり、レセプト情報等の管理ならびに解析の命令をサーバ装置群に発行し、また、その結果を取得して出力する。

サーバ装置群は常設のサーバ装置を以って構成し、増強したハードウェア資源を以ってレセプト情報等の管理ならびに解析を行う。多数の解析クエリを短期間に機動的に処理していく需要が見込まれたケースでは、機動的に研究室の備えるサーバ装置（現時点で最大128台）を追加投入し、解析処理能力を柔軟に増強できるシステムとする。

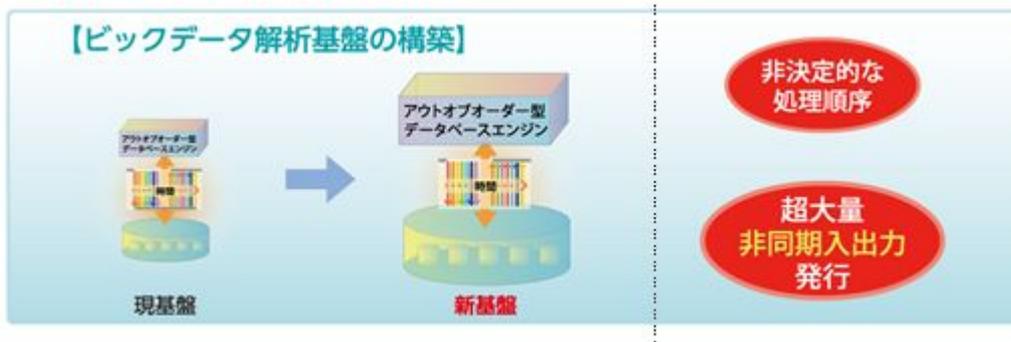
レセプト情報等の管理・解析のためのデータベースエンジンとしては、研究分担者である喜連川らが内閣府最先端研究開発支援プログラムにおいて開発した超高速データベースエンジンを用いる。

超高速レセプトビッグデータ解析基盤のソフトウェア構成は、レセプト情報等に対する解析処理を機動的に行い、最先端の情報技術を投入する。また、研究者による解析クエリの発行ならびに結果の確認を容易とするために、レセプト情報等の解析のためのビジネスインテリジェンス(BI)ツールを複数開発し、グラフィカルインターフェース上で解析クエリの構成と発行、実行状況の確認、グラフや表による結果の表示とダウンロード等、一連の作業を統合的に実施できることにする。当該ツールの開発に関しては、研究班において保険医療分野の研究者と情報技術分野の研究員の密接な連携の下、機能追加を機動的に行う。

超高速レセプトビッグデータ解析基盤の構築・運用におけるレセプト情報等の取り扱いは、厚生労働省のNDBガイドラインに従い、運用管理規定ならびに内部監査規定を設け、情報セキュリティの確保と個人情報の保護に努める。サーバ装置群については、研究室へのICカード扉錠による立ち入り制限に加えて、研究室内に「制限区画A」と称する保安区画を設け、専用セキュアラック扉錠によるアクセス制限を行うほか、論理的なセキュリティ手段として、IPアドレス認証と公開鍵方式のユーザ・端末認証を組み合わせたアクセス制限やアクセス監査ログ取得を施す。また、クライアント端末については、同じく研究室へのICカード扉錠による立ち入り制限に加えて、研究室内に「制限区画B」と称する保安区画を設け、パーティションにより物理的に窃視を防ぐ措置を実施した他、共通鍵方式のユーザ認証、セキュリティソフトウェアの導入、出力デバイスの制限、アクセス監査ログ取得を施す。制限区画Aと制限区画Bを接続するネットワークについては、保守作業時を除き、他とは独立させることとしている。このように、多重的に物理的ならびに論理的な保安措置を行う。

超高速レセプトビッグデータ解析基盤のハードウェア・ソフトウェアは、研究利用を想定して構築し、これらの基盤の利用を広汎な研究者に展開する方法についても検討する。

図表 2. 超高速レセプトビッグデータ解析基盤の構築



(人員・組織) 喜連川、合田、研究員 1、研究員 2 が東京大学構内で行う。

(スケジュール)

超高速レセプトビッグデータ解析基盤構築 : 5 か月 (2015 年 4 月 ~ 8 月)

超高速レセプトビッグデータ解析基盤構築 (増強) : 5 か月 (2016 年 4 月 ~ 8 月)

解析応用アルゴリズムソフトウェア開発 (研究員): 研究開始日から研究終了時

可視化ソフトウェア開発 (研究員): 研究開始日から研究終了時

既に構築を行った高速レセプト・ビッグデータ解析基盤を基として、新たな超高速レセプトビッグデータ解析基盤の構築と、応用ソフトウェアの開発を並行して行うことにより、研究が絶え間なく進展することができる計画を整備している。応用ソフトウェア開発を担当する研究員、研究員については、合田の指導の下、保険医療分野の研究者と密に連携して、研究開始時から研究終了時まで機動的なソフトウェア開発を進める。なお、IT 分野ではハードウェアの高性能化が極めて速く進展することから、市場動向を丁寧に見ながら、戦略的にハードウェアを段階的導入する計画とし、これにより、プロジェクト年限全体での投資効果の大幅な向上を目指す。

変更点 : ビッグデータ解析基盤への新投資はしない。研究員 1、研究員 2 は予算制約のため雇用できず。

2) 全 NDB データの検証と質向上

(必要性) これまで、全 NDB データの精度検証は行われておらず、公表されていない(ただし、レセプトと特定健診データに関するリンケージ率の低さを指摘した論文が近年公表された)。精度検証が行われなかった原因は、厚生労働省の本課題の認識不足に加え、ビッグデータを縦横無尽に解析する基盤の不在にある。将来 NDB データを使った研究活動を更に発展させるためにも、多面的なデータ検証が必須である。

ただし、厚生労働省から提供を受けた NDB データに対する集計をしても、実際のデータソースまでさかのぼらないと、根本的な問題解決には至らない。そのために、研究者の研究利用目的の枠組みにおいて禁止されている、NDB データを外部データと比較を行う。具体的には、神奈川県保健福祉局保健医療部と三重県国民健康保険連合会がそれぞれ厚生労働省に対する申請を行い、本来目的である医療費適正化分析の利用申請として、神奈川県

と三重県の全国健康保険のデータとそれに該当する NDB データの比較を日本で初めて実施する。

(目標) 本検証により、我々は NDB データを利用した研究の信頼性を担保する。

・診療所のレセプトと突合できない調剤レセプトが存在し、調剤医療費が少なく算出される問題

・特定健診データとレセプトデータを突合できない保険者が存在する問題

・ID1 と ID2 には欠点があり、ユニークな番号数が日本国民の人数を超えてしまう問題等を列挙する。NDB データの利点と欠点を明確にし、問題の解決方法を提示する。

(手順) 我々は、2010 年度のデータを使い、外来レセプトに紐づかない調剤レセプトの存在、レセプトと特定健診データのマッチング率の低さを指摘した。今回は全データ(2009-2014 年度)に対して実施する。

1. レコード件数、被保険者数、医療費などの推移を層別(レセプト種別、年月別、保険者別、地域別等)に示す。
2. 異常値の検出として、各データフィールドにおける定義以外の入力値の検証する(例えば、性別に男性(1)や女性(2)以外の値が入力、特定健診の血圧に 4 ケタ以上の値や文字情報が入力されている等)。
3. 個人を特定する ID の検証を行う。レセプト種別ごとに、NDB の ID1 と ID2 の一意(ユニーク)の件数を調べる (ID1 は保険者番号・記号番号・生年月日、性別から生成したハッシュ値、ID2 は、氏名・生年月日・性別から生成したハッシュ値である)。ID 1 に対する ID2 の重複件数、ID2 に対する ID1 の重複件数を外来、入院、調剤、DPC、歯科毎に調査する。
4. ID のリンケージ(突合)の検証を行う。外来診療に付随して発生する調剤レセプトの性質を利用しマッチング(リンケージ)を行い、病院・診療所別、保険者別、都道府県別等の層別集計を行う。さらに、特定健診データとレセプトデータとの突合を行う。厚生労働省の国民医療費や患者調査、各保険者が公表している事業報告等の既存統計との比較を行う。
5. 神奈川県および三重県の国民健康保険から提供を受けたデータと NDB データの上記項目との比較を行い、問題点を明らかにする。また、国民健康保険連合会職員に対するヒアリング調査を行い、NDB データの今後の精度向上の提言をまとめる。

(成果) 厚生労働省厚生科学課、データ提供の責任部署である厚生労働省・保険局・医療介護連携政策課保険システム高度化推進室には、検証結果を外部公表する前に伝える。特に、保険システム高度化推進室よりデータ公表の許可を得なければならない。

また、進捗は戦略研究企画・調査委専門検討会のスケジュールに沿ったモニタリング・助言を定期的に受ける。

本データ検証の結果は、Journal of Epidemiology(Japan Epidemiological Association) に 2015 年 10 月までに査読付き英語論文として投稿する予定である。

国内学会では、日本医療情報学連合大会、日本公衆衛生学会で発表する。

国際学会は、米国医療情報学会 (AMIA) を検討中である。

変更点：国内雑誌として「実験医学 3 月号」に掲載予定。

「レセプト情報などの有効活用に関する有識者会議」の構成員である印南は、研究結果および成果が実際の政策に反映されるよう、当該会議に諮り、改善を求める役割を担う（2015 年度内）。

変更点：今年には実施しない。

3) 基礎統計

(必要性)現在、厚生労働省は NDB データに関する情報は、断片的にしか公開していない。これまでの日本の医療統計は、抽出調査や検証が主だったため、データ公表が諸外国よりも 2 年遅く、集計単位も年であった。そのために、超高速レセプトビッグデータ解析基盤を用いて全数を対象に、迅速且つ定期的な情報提供機能が必要とされている。

(目標)個人情報特定されない範囲で、疾患数(罹病率)、医療費の傾向などの情報を都道府県・二次医療圏別に提供する。また、研究範囲外の一般研究者が、研究デザインを企画・立案できる方式について検討する。具体的には、インタラクティブに情報提供する機能(ダッシュボード)を作成する。

米国 CMS では、ウェブサイトにおいてパラメータを指定することでブラウザ上に、対象とする疾患の人数などが表示されるダッシュボード機能を提供している。わが国でも、(個人情報を含まない)性別、年齢、疾病、地域等のパラメータを選択し、医療費、人数、入院日数などをダッシュボードの形で提示する機能を備えることを目標とする。

(手順) NDB データに対して、ヒストグラム、要約統計量、散布図などの可視化手法を適用し、要約・層別・可視化を行う。

データ項目は、年月、ID1、ID2、性別、年齢、入院年月日、入院年月日、保険者番号、保険者法別番号、保険者区分 1(社会保険、地域保険)、保険者区分 2(健保、国保、共済、後期、協会)、都道府県(保険者)、都道府県(医療機関)、医療機関コード、診療日数、点数、入外来区分、傷病名コード(主病名)、診療開始日、診療行為コード、医薬品コード、ICD10 大分類、ICD10 中分類、121 分類疾病、DPC 病名である。

層別は、レセプト区分(入院外・入院・DPC・調剤・歯科)、保険者、医療機関コード、病名区分、診療行為コード等とする。

変更点:可視化ツールは、既に実装済み。保険者データについても作成し、以後更新を重ねる予定。NDB データ収集後、NDB 版可視化ツールの開発を進めたい。

(成果)厚生労働省厚生科学課、データ提供の責任部署である厚生労働省・保険局・医療介護連携政策課保険システム高度化推進室には、検証結果を外部公表する前に伝える。特に、保険システム高度化推進室よりデータ公表の許可を得なければならない。また、進捗は戦略研究企画・調査委専門検討会のスケジュールに沿ったモニタリング・助言を定期的に受ける(以上については、以降の研究についても同様である)。

本成果におけるダッシュボード機能については、Methods of Information in Medicine に 2016 年 3 月までに査読付き英語論文として投稿する。

国内学会では、日本医療情報学連合大会で発表する。

国際学会は、米国計算機学会データ管理国際会議(SIGMOD2015)を検討中である。

変更点:国内学会では、2015 年 11 月に医療情報学連合大会で発表済み(シンポジスト(満武))。国際学会における発表は、予算の制約から海外出張をカット。

4) 多次元分析(クリニカルクエスチョン)

(必要性)レセプト情報は、診療報酬点数の記録であり臨床データが無いことから、利用について限界があるといわれてきた。また、病名についても、保険病名の信頼性の問題についても指摘をうけてきた。しかし特定の疾患や患者グループでは、レセプトから得られる情報により詳細で深い分析が可能である。ここでは生活習慣病の外来診療に焦点を当てて、レセプト情報の利用可能性について検討する。

(目標)今回、NDBのデータ項目を多次元的に組合せて、日本国全体での疾患別の有病率・他疾患併存の状況・疾患の重症度、疾患コントロールの状況と医療費の関係を明らかにする。また4医学会の協力を得て、レセプト情報の基にした発症群・合併症群・重症群の定義を確立させ、臨床的なエビデンスに基づいた生活習慣病の診療パフォーマンスの指標を開発し、これらの指標を時系列で可視化する。これらの情報の公開や診療現場へのフィードバックが、わが国の生活習慣病の診療の質の改善にどのように寄与するのかを検討する。

(手順)患者数が多く大量の医療資源を消費している生活習慣病(高血圧、糖尿病、高脂血症)について、「日本糖尿病学会」、「日本高血圧学会」、「日本動脈硬化学会」、「日本腎臓学会」と連携し、これらの疾患管理を評価するための指標を同定する。

はじめに、生活習慣病を有する患者グループを同定する方法論(傷病名、診療行為、医薬品コード等などから定義)を確立する。例えば高血圧の場合、レセプト病名に加えて降圧剤(1)アンジオテンシン受容体拮抗薬(ARB)、(2)カルシウム拮抗薬、(3)アンジオテンシン変換酵素阻害薬(ACE阻害薬)、(4)その他)の処方の有無などから有病者を定義できるはずである。また糖尿病の場合は、「糖尿病」が傷病名として登録されている、糖尿病治療薬が処方されているなどが定義となる(現在、生活習慣病を中心に作成中であり、各学会との協議を通じて確定させていく)。

次に、各疾患の定義を用いて患者グループを同定し、患者の属性(年齢・性別)・併存症・疾患コントロールの状態などが医療費とどのように関連するかを検討する。また同じリスク・プロファイルを持つ患者に対する診療パターンや医療費が、地域や医療機関の特性によりどのように異なるかを検討する。この場合、医師誘発需要理論を念頭において、地域の医師密度や医療機関密度、医療設備の配置と、医療費や診療との関連について詳細な検討を行う予定である。さらに診療の密度が患者アウトカムや生活習慣病における治療成績(疾患コントロールの状態や合併症予防)との間に関係が見られるかも検討する。これらの知見をもとにして、診療の適切性や医療資源配分の適切性なども検討する。

次に各疾患において、罹病期間や併存症の有無、重症度、入院診療の利用などにより患者のリスク・プロファイリングを行う。発症群においても非薬剤治療者と治療者、合併症群においては併発症や四肢切断などの処理を受けた者、重症群においては透析の有無など、臨床上的視点からの精緻な実態把握を行う。今回、複数年データを利用するために、各年度で新規に医療機関を受診した患者が発症群・合併症群・重症群のどこに分類されるのか、その後の継時的な変化について明らかにする。

結果は、学会にフィードバックすることで再度指標の適切性を議論する。これらのプロセスを経て、わが国の生活習慣病の実態について問題点が明らかになるとともに、診療の標準化や質の改善に結びつけるために効果的な公表のあり方について検討する。

第二に、各学会との連携と並行して、入院・DPC・外来・調剤レセプトデータの各項目間のアソシエーション(バスケット)分析を多次元的に行う。これは、入院医療費と外来

医療費、それぞれの病名、病名と処置、病名と処方といった項目間での膨大なアソシエーション・ルール(支持度、信頼度、リフト値)について超高速レセプトビックデータ基盤を利用することで見出す。

第三に、統計的に出現頻度が低い疾患を稀少疾患と定義して、リスク構造調整時の期待損失額が負となるリスク因子等の設定を検討する。

第四に、保険者が既に重複受診や頻回受診などを推計している計算式を参考にして NDB においても同様の算出を行い、適正化可能な医療費を推定する。重複受診者の定義については同月中に同一診療科のレセプトが複数の医療機関から 4 枚以上発生、頻回受診者については同月診療科の外来レセプトの診療実日数の合計が 15 日以上を設定する。

最後に、本研究で設定した疾患定義は、NDB データベースに組み込む措置について厚生労働省と検討し、他の研究者も利用できるようにする。

次年度は、初年度の成果を基に別の学会に示し、同様の定義と実態を把握する。

(成果)

4 学会の論文誌に、学会との密接な連携による全日本の患者数データと医療費のデータについて投稿する(2015 年度末)。また、国際ジャーナルにも 4 本を投稿する(2016 年度初頭)。

国内学会では、4 学会で発表する(2015 年度中)。

国際学会は、4 学会が推奨する国際学会での発表を行う(2016 年度中)。

変更点：学会は、日本腎臓学会、日本糖尿病学会の二つに限定した。すでに定義体の第一バージョンは作成し、保険者のデータを利用して集計を開始している。年度末の成果発表を行うべく、努力している。



5) 都道府県・二次医療圏・市町村別分析(Small Area Analysis: SAA)

(必要性) 衛生統計学、疫学などの公衆衛生の分野では、地域の状況を表す指標としては、年齢分布の違いを調整した、年齢調整罹患率、標準化罹患率などが良く利用されている。しかし、従来の指標は、人口の地域変動に基づく標本誤差の影響を強く受け、人口の小さい地域の指標のバラつき、稀少疾患のわずかな変化が見かけ上の指標を大きく変化させる問題がある。

(目標) 従来の統計指標の問題の解決策として、空間疫学・経験的ベイズ推定量を導入して、都道府県・二次医療圏・市町村の医療計画及び医療費適正化計画に資する情報を提供する。

(手順) はじめに、NDB および保険者の医療費データを用いて疾病発生、診療パターン、医療費受給状況等について地域比較を行い、各地域の特性を検討する(従来の集計表を作成する)。次に、都道府県別、二次医療圏別、市区町村別に疾患別の患者属性、併存症、受診頻度、処方パターン、各検査実施頻度、医療費の実績などを比較し、その関連について検討を行う。その際、通常の住民の性・年齢構成を調整し標準化された指標(集計表)とともに、各地域の人口サイズの違いによるバラツキの影響を取り除き、小地域推定において安定した推定を行うためベイズ推定による指標についても適切なモデル化を行う(経験的ベイズ推定)。さらに周辺地域の類似性・空間相関についての検討とそれぞれの指標を作成する。

保険者(三重県)のデータを利用して、同様の疾患発生状況、治療状況と医療費の関連について医療機関(病院と診療所)の違い考慮した分析を行い、地域差や地域特性を明らかにする。

(成果) 本成果は、ダッシュボード機能に搭載する。

国内学会誌では、応用統計学に投稿する。また、日本公衆衛生学会で発表後、学会誌に査読付き論文として投稿する。

変更点: 研究員 はカット。学会は、日本腎臓学会、日本糖尿病学会の二つに限定した。すでに定義体の第一バージョンは作成し、保険者のデータを利用して集計を実施している。

6) 調整医療費の推計と地域毎医療支出目標額の設定

(必要性) 増え続ける日本の医療費を適正化するための施策として、都道府県ごとに医療支出目標を設定することが検討されている。また、2015年の通常国会へ法案提出を目指している国民健康保険(市町村)の運営を都道府県へ移管するにあたり、全国知事会等が財政基盤強化策・政府部門の国費投入の規模の提示(エビデンス)を求めている。

(目標) NDB データを利用して外来および入院医療費に影響を及ぼす因子(年齢・性別・併存疾患等)、医療供給体制などの外部のセカンダリーデータを活用して地域毎の医療費の予測モデルを構築し、全国の都道府県および市町村へのデータ提供を行う。

(手順) 多次元分析および SAA 分析の結果を基に、外来および入院医療費に影響を及ぼす因子(年齢・性別・併存疾患等)で調整したモデルを構築する。このモデルは、多変量直線回帰などより詳細なモデル作成する。

$$\text{Cost}_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Sex}_i + \beta_3 \text{Comorbid}_i + \epsilon_i \quad \text{---(1)}$$

$$\text{Cost}_{\text{adj}} = \text{Cost}_{\text{avr}} \times \text{Cost}_i / (\beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Sex}_i + \beta_3 \text{Comorbid}_i) \quad \text{---(2)}$$

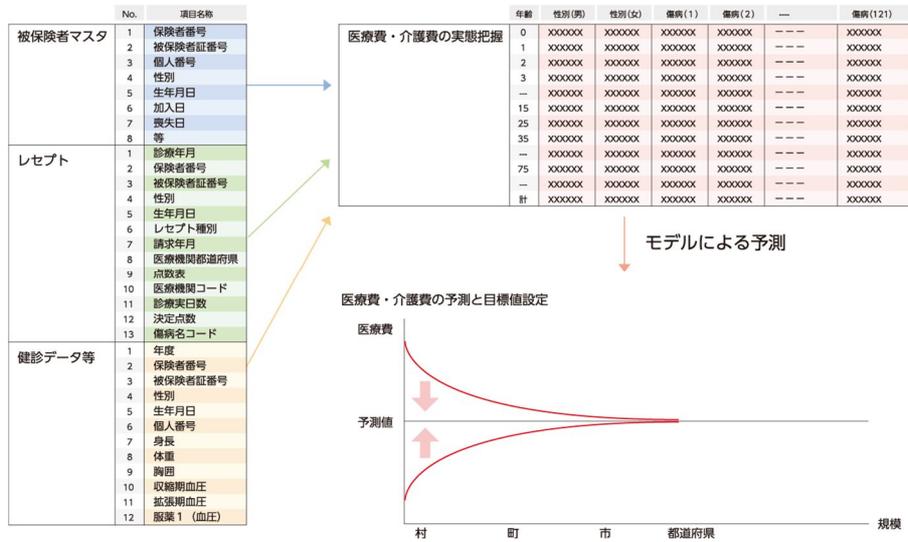
医療費・介護費の予測は、経済協力開発機構(OECD)の医療費予測モデルを踏まえて、日本版の短期予測および中長期予測モデルの3つを作成する。

- (1) 推計単位として個人に焦点を当てたマイクロ・モデル
- (2) 医療費や個人を階層化して推計を行う構成要素基礎モデル
- (3) 総医療費等を基に予測を行うマクロ水準モデルを適用する。

同時に各セカンダリーデータ(需要データ(人口、健康状態、予防プログラム等)と供給データ(医療供給体制、医療費・介護費財源町タス方法等))も組み合わせてモデルを構築する。

(成果) 本成果は、「医療・介護情報の活用による改革の推進に関する専門調査会(内閣官房・社会保障制度改革推進本部)」において、方法論と推計結果を提示し、施策への反映を促進させる。国内学会では、日本公衆衛生学会で発表する。

図表 3. 医療費・介護費の予測と目標値設定（案）



変更点：研究員 2 は予算制約のためカット。専門調査会、学会発表への働き掛けは、今年度は行わない。

7)日本版リスク構造調整方式の確立と財源調整施策の導入の試算

(必要性)現在の保険者間の財源調整は、前期高齢者では「前期高齢者」と「それ以外」の区分の加入率のみに基づく調整であり、後期高齢者では医療費が発生した後の事後的な調整に過ぎない。したがって、保険者単位のリスク構造を基に各保険者が実態を把握し医療費適正化に取り組む誘因をもった、財政調整措置等の問題を検討しなければならない。

(目標)NDB および保険者の医療費データを用いて、わが国におけるリスク構造調整の実現可能性について検討を行う。

(手順)はじめに、リスク構造モデルの構築に当たっては、諸外国(オランダ、ドイツ、ベルギー、イスラエル、スイス、アメリカなど)で実施されているリスク構造調整を参考として、日本版性別・年齢情報、地域情報、所得情報、疾病情報などのリスク項目を採用する。その上で、性別・年齢情報のみを用いた簡易的なリスク構造モデルから疾病情報を用いたより高精度のリスク構造モデルまで、複数のリスク構造モデルを構築することで、各リスク構造モデルの実現可能性について検討を行う。リスク構造調整の実現可能性については、保険集団レベル、個人レベルの二つの視点から検討を行う。

保険集団レベルでの検討に当たっては、各リスク構造モデルにおける保険集団レベルでの利潤(リスク調整交付金 - 医療費)分布と保険規模との関係をブート・ストラッピング法により定量的に評価することにより、想定される保険者規模(市区町村・二次医療圏・都道府県など)と必要とされるリスク構造モデルの精度の関係を明らかにする。あわせて、期待損失額が負となるリスク因子(腎不全などの慢性疾患患者など)を個人レベルで明らかにすることにより、各リスク構造モデルの実現可能性について個人レベルで検討を行う。

リスク因子で問題となるのは、例えば、生活習慣病の患者は複数の慢性疾患を持っていることが多く、解析対象となる生活習慣病の他に異なる併存疾患を持つ患者を、どのようにグループ分けして解析するかが課題である。医療費や受診頻度、診療内容はどのような併存疾患を持つかにより大きく異なってくるからである。そこでわれわれは、Charlson comorbidity score(index)を用いて併存症により患者リスクを層別化する方法を実施する。Charlson comorbidity score とは、心筋梗塞、うっ血性心不全、末梢血液疾患、脳血管障害、認知症、肺疾患、結合組織障害、消化性潰瘍、肝疾患、糖尿病、麻痺、腎疾患、悪性腫瘍、転移性腫瘍、重症肝疾患、HIV 感染症等 10 年以内の死亡に関連する併存症から患者の重症度を評価するためのスコアである。新しい層別データ毎に多次元分析を行い、正確な患者数、疾患毎医療費を算出する。解析対象期間中の全レセプトから Charlson comorbidity に該当する傷病名を同定することで、モデルの頑健性を確保する。

(成果)成果は、2016 年 8 月末に公表できる段階となる。その後、菊池と辻は、本研究より提示した全保険者のリスク構造、諸外国のモデルに基づいた推計結果を、全国町村会、全国市町会、全国知事会、地域保険者、都道府県国保連合会、各自治体の首長へ情報提供し、今後の国民健康保険広域化の検討材料とする(2016 年度)。

変更点：文献調査を実施している。全国町村会、全国市町会、全国知事会、地域保険者、

都道府県国保連合会、各自治体の首長へ情報提供は、今年度は行わない。来年度、データ分析ができるか検討中である。

8)1次・2次・3次予防および予防・医療・介護の連携の実態把握

(必要性)特定健診データやレセプト情報を活用した後期高齢者の支援金の加算・減算の評価は、当初の計画を実行できなかった。したがって、特定健診・特定保健指導の評価指標や参酌標準は、保険者の保健活動を評価し、且つ医療費適正化のインセンティブを有するものに再検討する必要がある。

(目標)NDBレセプト情報と特定健診データを活用した新しい1次・2次・3次予防群の確立、保険者のレセプト情報と介護レセプトを活用し訪問診療をはじめとする在宅医療の実態を明らかにするとともに、疾病予防・医療・介護の連携の実態把握

(手順)特定健診対象者の属性(年齢・性別)、特定健診受診率、医療機関受診者割合、一人当たり総医療費等を記述する。また、特定健診の受診者と未受診者の属性、医療費を比較する。次に、特定健診の受診者の年度毎の健診結果を示す。特定健診の結果により被保険者を階層化するため、「標準的な健診・保健指導プログラム(確定版)(厚生労働省)の保健指導対象者の選定を階層化」を用いる。階層化の各ステップにおいて、健診結果に基づいた対象者を分類し、基準値を超える者の人数および割合、積極的支援群および動機付け支援群の人数および割合、1人当たり医療費との関連を記述する。次に、2008年～2014年に特定健診の対象となった被保険者の分析を行う(この期間に新たに保険者(国民健康保険)に加入した被保険者、脱退した被保険者は除外する。)

特定保健指導に関しては、特定保健指導利用の有無別に翌年の医療費、受診、および高血圧・糖尿病のレセプト病名発生率がどのように変化したのか検討し、効果的な保健指導対象群の選定を検討する。

医療費適正効果の分析は、リスク有群とリスク無群の5年後の医療費を比較し、リスク要因と医療費の関連を観察し、ロジスティック回帰モデルにより生活習慣病の罹患に関する被保険者のPropensity Score(傾向スコア)を推定して、各群の被保険者の疾病罹患に関する交絡要因(性・年齢・既往歴等)を調整し、特定健診受診群と未受診群、検査値異常群と正常群、保健指導受診群と未受診群の医療費比較を行う。

1次・2次・3次予防群の設定には、複数年の特定健診・特定保健指導データおよび電子レセプトの処方情報・診療行為情報といった詳細な診療行為情報から、新しい特定保健指導の対象群、医療機関受診勧奨群等を設定する。

特定健診・特定保健指導の新しい評価指標の作成は、特定健診の受診の関連要因の分析を行い、統計的な裏付けをもとに設定する。具体的には、特定健診受診の有無や疾患等で分析群を設定し、受診関連要因を年齢、性別、前年度の特定健診受診、前年度の入院医療の利用、前年度の外来医療の有無として、次年度の健診受診に対するオッズ比を計算する。次に、受診回数で被保険者を層別化して、被保険者の属性(年齢・性別)および医療費・累積医療費を群間比較し、ロジスティック回帰分析(被説明変数に特定健診の受診の有無、説明変数に性別、年齢、前年と次年の特定健診受診の結果、および医療費)を行う。これらの分析結果をもとに、新規受診率、継続受診率、脱落率、未受診率などの指標を設定する。

また、中長期的な評価の視点として、本研究で作成する特定健診・特定保健指導の評価指標が将来的な在宅医療・ケア利用に与える影響を検討する。500万人規模の在宅医療・ケアの提供実態を把握し(基礎統計および多次元分析)それ単体で質の高い地域包括ケアシステム構築のための指標作りに直接的に貢献することにつながり得るが、それに加え

て、特定健診・特定保健指導の結果が将来の在宅医療・ケアの利用パターンにどのような影響を与えているかという関連の検討をも可能とする。これにより、将来の要介護状態を未然に防ぎ健康寿命・QALY (Quality Adjusted Life Years) の延伸に資する効果的な予防や指導のあり方をさらに精緻に見出す。

(成果) 協力保険者への成果フィードバックは、2015 年度末になる。NDB データを使った結果は、国内学会では、日本公衆衛生学会で発表する。国際学会では、欧州医療情報学会 (MIE2016)、国際医療経済学会 (iHEA2016)、米国医療情報学会 (AMIA2016) での公表を検討する。

変更点：研究員 2 は予算制約のためカット。

9) 研究者用データ提供に向けた検討と実装

(必要性) 現在、NDB を研究利用する際には、厚生労働省は複雑な構造の電子レセプトを分析し易いように加工して提供しているが、提供件数に限界がある。申請者らは、平成 24～25 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)」を実施した。しかし、当時厚生労働省から提供された限られた保険者(全体の 30%)のデータを基に作成した。全データをデータソースとした研究用途における汎用性の高いレセプトデータセットの設計と作成が必要である。

(目標) 今回は検証結果(1)を基に全データを対象とした、研究用途における汎用性の高いレセプトデータセットの設計と作成を行う。特定健診・保健指導とレセプトとリンクさせたデータセットも作成する。加えて、上記データセットでは対応できない希少疾患や長期観察が必要な研究のためのデータセット(以下、特殊データセット)を作成する。

また、台湾では特殊データセットとして、癌、糖尿病、精神疾患、高額な医療費を要する疾患等の 15 種類のデータセットがある。また、米国ではメディケア加入者を対象にして 21 種類の慢性疾患患者の追跡データが可能なデータセットがある。諸外国のデータセットの構造、特徴を参考にして、日本版の特殊データセットの設計と作成を行い、一般の研究者への利用を広げる。

(手順)

基本データセット作成を、全 NDB データを使って行う。

2009 年度から 2014 年度の全レセプト・特定健診・特定保健指導データから以下を作成する。日本のレセプト(電子レセプト)は、詳細な診療行為データ(診療行為コード(SI)や医薬品コード(IY)等)が含まれる。そこで、診療行為記録を含み、且つ研究者が利用しやすい

- (1)一レセプトーレコード
- (2)一定期間のレセプトを個人 ID 毎に統合する等の処理をしたデータセット
- (3)特定健診・保健指導とレセプトとリンクさせたデータセット
- (4)希少疾患や長期観察が必要な研究のためのデータセット(特殊データセット)

特に、諸外国のデータセットの構造、特徴を参考にして、日本版の特殊データセットの設計と作成を行い、一般の研究者への利用を広げる。

米国の CMS(Centers for Medicare & Medicaid Services)では、メディケア加入者に関連が深い 21 種類の慢性疾患が指定されて、追跡調査されている。それは大きく 6 つに分類されている。(CCW データには関連するあらゆる詳細データ、すなわち診断、処置コード、診療請求タイプ、サービスの日付が入っており、それらを使ってここに上げた慢性疾患フラッグを生成している。)

- ・心臓と肺に関する疾患(Cardiopulmonary Disorders)：急性心筋梗塞、心臓細動、心不全、慢性閉塞性肺疾患、虚血性心疾患、脳卒中、一過性脳虚血発作(TIA)
- ・癌(Cancer)：5 種類の癌：結腸・大腸ガン、子宮ガン、乳ガン、肺ガン、前立腺ガン
- ・精神・神経系疾患(Mental/Neurological Disorders)：アルツハイマー病と関連する障害

または認知症をひとまとめに見ており、アルツハイマー病だけを別立ての分類に、加えて抑鬱症を指定している。

- ・ Ophthalmic Disease : 白内障と緑内障
- ・ Skeletal and Joint Disorders : 骨格および関節については、三つ。腰および骨盤骨折、骨粗鬆症・リウマチ・変形性関節炎は1分類。
- ・ Other Disorders : 上記には分類にされず、しかもメディケア集団に関連の深いふたつの慢性疾患。慢性腎臓病と糖尿病。

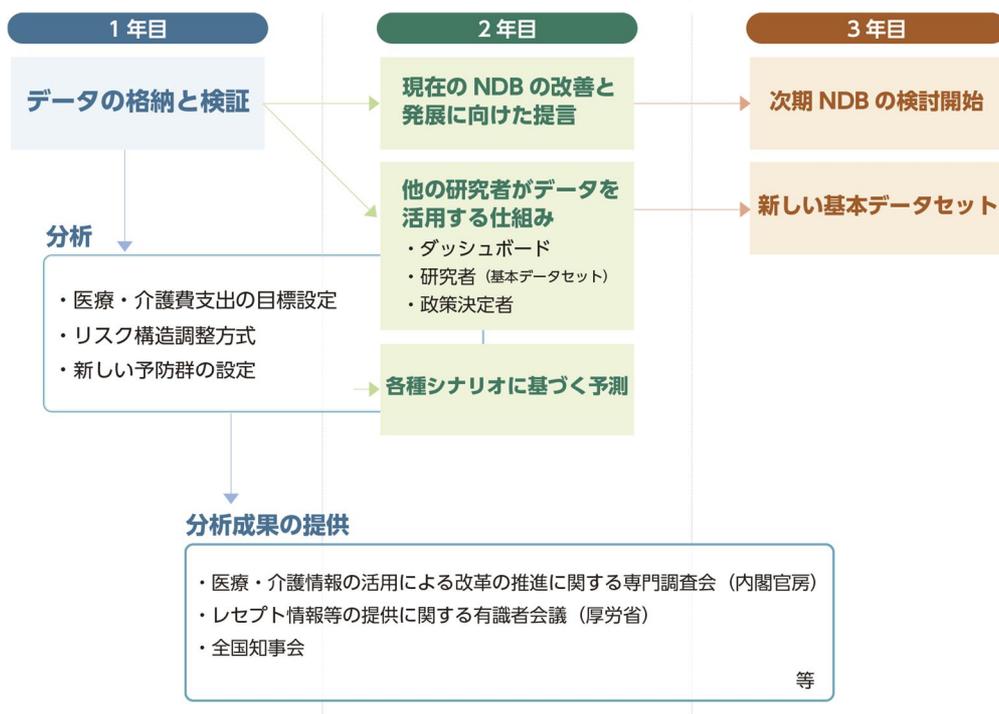
(成果) 2016年度末に、厚生労働省に上記データセットの作成アルゴリズム、フォーマット、データセット等を提供する。

変更点 : 現在、1件の提供が完了した。今後、2件目の提供を検討する。

6. スケジュール(研究期間)

2015年4月1日~2017年3月31日である。

図表 4. スケジュール(案)



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、全国の診療パタンの実態・地域差の比較、併存症・医療費等を加味した疾患群の経年変化、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

5月 6月 7月 11月 12月 1月 2月 3月

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.

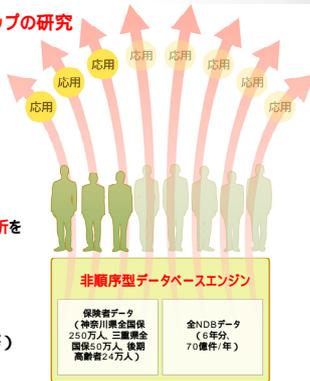
スライド 6

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**

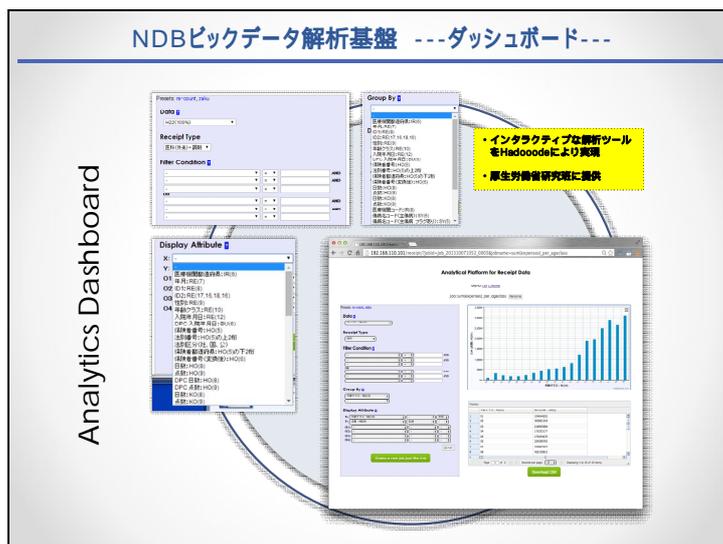
- 非順序型データベースエンジン（世界最先端IT）
- 目的志向のパーティカルな応用ソフトウェアの機動的立上げ（continuous dev/dep: 研究と共に日々進化するソフトウェア）
- NDB全数からのエビデンス抽出を実現
- これまでは手付かずであった**テイル集団の詳細分析**を実現（FWA分析）
- 「我が国NDBシステムの実験場」**：
本知見に基づく次世代システム構築への提言（データ管理手法、分析手法・インターフェース等）



The diagram shows the application of the non-sequential database engine to various data sets, with a table of 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



スライド 11

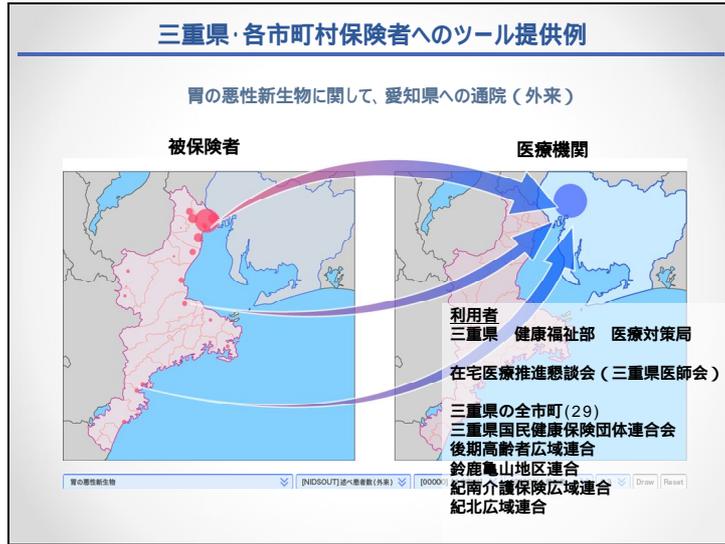
NDB集計・分析方法に対する、腎臓学会コメント

(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	386	8	1,117	5	354	11919	64
3 木曾町	6,468		63	2	361		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		366	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	6,321		135	2	425		154	5,433	37
24 海部町	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	506,811	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

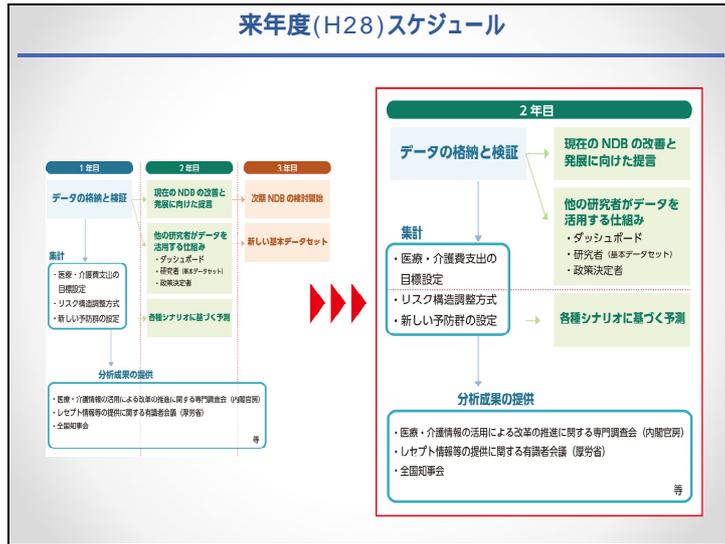
	被保険者数	突合数(人数)	割合	非突合数(人数)	割合
医療レセプトID2	122,425,816	4,339,266	3.5%	118,086,550	96.5%
特定健診ID2	22,211,086	4,339,266	19.5%	17,871,820	80.5%

NHK NEWS WEB 2018年04月13日 10:53:10 (1/1)

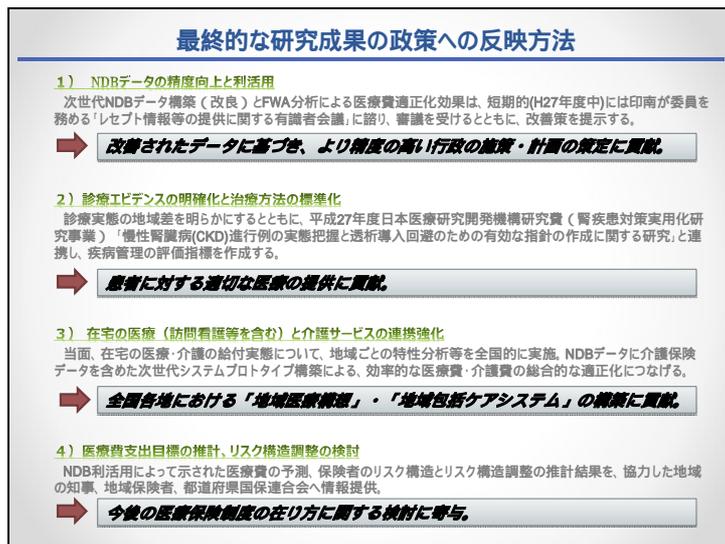
厚労省の診療データ 約8割が活用できず

厚生労働省が4年ぶりにデータからデータベースに連携している診療データ活用率の公表。厚労省が公表したデータによると、厚労省が保有している診療データのうち、約8割は活用できず、約2割は活用されていると発表された。

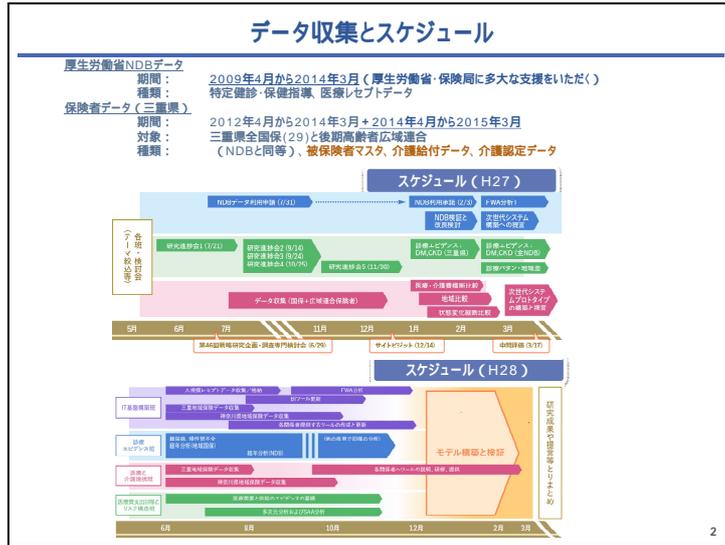
スライド 21



スライド 22



スライド 2



スライド 3



スライド 4

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）のデータ、及び保険者から申請者らが収集した医療・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、本知見に基づく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、全国の診療パターンの実態・地域差の比較、併存症、医療費等を加味した疾患群の経年変化、管理の評価指標を作成。「日本高血圧学会」、「日本動脈硬化学会」を同様の分析（高血圧・高脂血症）も実施予定。

3) 医療と介護サービスの連携

【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、医療・介護のサービス利用の実態を分析し、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

_スライド 5

世界トップクラスの医療ビッグデータと世界最速の解析技術の連携

NDBは日本の国民皆保険制度を背景とした世界でも類のない貴重なデータベースであるが、膨大なデータのため従来のデータ処理では十分な分析ができないため、ビッグデータ解析基盤の増強が課題となる。研究開発提案者らは、厚労科研・戦略研究等で、従来と比べて1,000枚を超える高速を持つ非順序型データベースエンジン(*)による高速レセプトビッグデータ解析基盤を構築し研究を行ってきた。また、レセプト情報等に対する解析処理を機動的に行うためのビジネスインテリジェンスツールとして、目的志向のパーティカルなソフトウェア開発を行ってきた。

(*) 東大発の独自の「非順序型」ソフトウェア実行原理。オープンソースベース実装でも圧倒的な高速性を発揮（Hadoop、PostgreSQL、MySQL等）



(例) 400万件のデータを用いた保険レセプト解析例（下図）

通常データベース ストレージアクセス 100回/秒 → 4億秒かかる 4600日
これでは時間がかかりすぎる。できれば解析を数分でできないだろうか



超ビッグデータ処理エンジン

ストレージアクセス 1000万回/秒 → 40秒 !! → 他の処理もいれれば数分を実現 !!

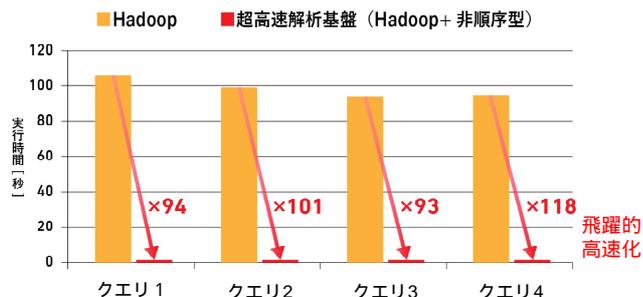
医療業界のビッグデータの活用はまだまだ発展途上の段階。だが、近年は、NDBよりも遥かに大量のデータのデータベース化と、これらのデータを組み合わせる高速処理するデータベース技術が開発されている。

5

スライド 6

高速保険レセプト解析基盤の実行性能

クエリ1：都道府県ごとの急性心膜炎患者数（対人口比率）
 クエリ2：口唇の悪性新生物の都道府県ごとの平均診療費
 クエリ3：呼吸器結核に関する年齢ごとのレセプト件数
 クエリ4：コレラ患者の都道府県ごとの平均診療費



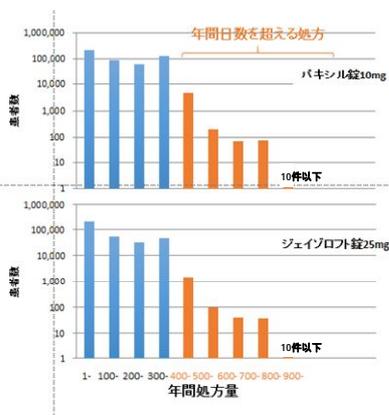
スライド 7

【大規模レセプト解析IT基盤整備・運用班】

FWA分析：薬剤の処方傾向

主要抗うつ薬

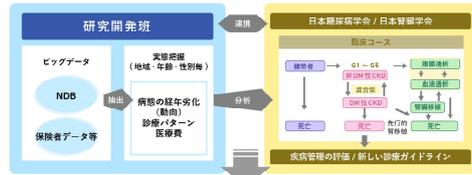
医薬品コード	年間点数合計
パキシル錠10mg	192955113
ジェイゾロフト錠25mg	147501859
パキシル錠20mg	111788152
サインバルタカプセル20mg	69196888
リフレックス錠15mg	65086479
ジェイゾロフト錠50mg	57252774
ルボックス錠2.5 25mg	40511982
レメロン錠15mg	34280097
デプロメール錠2.5 25mg	34140308
トレドミン錠25mg	34043098
...	...



医学研究の推進：時系列解析によるギャップ検証

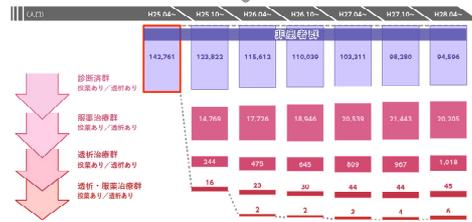
診療ガイドラインの推奨と実臨床のギャップ

「日本糖尿病学会」、「日本高血圧学会」、「日本動脈硬化学会」、「日本腎臓学会」と連携



NDBデータおよび保険者データを利用することにより、患者グループの同定、患者の属性（性別・年齢）・併存症の状態等と医療費との関係を明らかにする。

4学会に加え、連携学会を検討中。



患者数、疾病がどのような経過をたどるのか（時系列分析）、疾病に対してどのような診療が行われているか（診療パターン）といった分析を行い、新しい診療ガイドラインを作成するためのエビデンス、疾病を管理するための評価指標の作成につながる。

10

スライド 12

NDBデータを用いたCKD抽出

傷病名コード	傷病名称	ICD10	NDBし件数	NDB人数
3621014	腎性網膜症	N188	1,970	708
5859002	慢性腎不全	N189	1,291,526	194,902
5869015	尿毒症性心膜炎	N188	10件未満	10件未満
5869016	尿毒症肺	N188	10件未満	10件未満
8838554	尿毒症性多発性ニューロパシー	N188	10件未満	10件未満
8838555	尿毒症性ニューロパシー	N188	10件未満	10件未満
8841385	尿毒症性脳症	N188	10件未満	10件未満
8842116	末期腎不全	N180	34,592	7,933
8847501	赤血球造血刺激因子製剤低反応性貧血	N188	10件未満	10件未満
8847544	尿毒症性心筋症	N188	10件未満	10件未満
8847579	慢性腎臓病ステージG3	N188	10件未満	10件未満
8847580	慢性腎臓病ステージG3 a	N188	10件未満	10件未満
8847581	慢性腎臓病ステージG3 b	N188	10件未満	10件未満
8847582	慢性腎臓病ステージG4	N188	10件未満	10件未満
8847583	慢性腎臓病ステージG5	N180	10件未満	10件未満
8848103	慢性腎臓病ステージG5 D	N180	10件未満	10件未満
5859001	尿毒症	N19	10件未満	10件未満
5860004	腎性無尿	N19	10件未満	10件未満
5939017	無機能腎	N19	1,385	472
8835642	腎不全	N19	88,080	21,498
8837198	蛋白尿貧血	N19	23	12

腎不全患者は、約225,825人

（コメント）CKDは、NDBに出てくる人たちは恐らくG4-G5くらいの保存期腎不全の人たちなのでしょう。そうするとこのくらいの数字で良いのかもしれません。

NDBではCKD全体の把握は難しいでしょうね。もっと軽症者が入ってくるので900万人くらいになるかと思います。

ステージ情報（G3等）の傷病名を有する入力をしている医療機関は、皆無に近い。慢性腎不全、腎不全等の病名と処置、処方を組み合わせて対象疾患を選定する

スライド 13

NDBデータおよび三重県国保データを用いた透析患者、腎移植 透析（移植腎）

日本腎臓学会理事・山縣邦弘教授（筑波大学医学医療系腎臓内科学）と傷病名コード、診療行為コード、医薬品コードを定義中

診療行為名称	レセプト数 (NDB)	レセプト数 (三重)
人工腎臓 (その他)	65,191	312
人工腎臓 (慢性維持透析) (4時間未満)	996,343	1,434
人工腎臓 (慢性維持透析) (4時間以上5時間未満)	716,592	2,728
人工腎臓 (慢性維持透析) (5時間以上)	70,818	387
人工腎臓 (慢性維持透析) (連続)	10件未満	0

傷病名称	透析患者数 (NDB)	透析患者数 (三重)
移植後	5150	24
移植後不全	27	0
移植後、継続的透析	298	1

透析患者数 (NDB) 透析患者数 (三重)

220,497 4,861

NDBから推計した全国の透析患者数は、220,497人(生活保護は含まれない)

日本透析医学会の調査では、297,126人(医療施設調査もとに推計)

NDBを活用した患者数推計は、IDの問題や公費レセプトが除外されていることを踏まえないといけない。

(腎臓学会コメント)
 腎移植の後、機能が廃絶しても免疫抑制剤は投与し続ける。つまり
 (定義) 腎移植病名 + 免疫抑制剤で透析診療行為があれば腎移植
 (定義) 腎移植病名 + 免疫抑制剤で透析診療行為があれば透析と判断した方がよい。
 これまでの移植腎の生着率などと比較すると良い

経年的な変動で観察する必要がある。

スライド 14

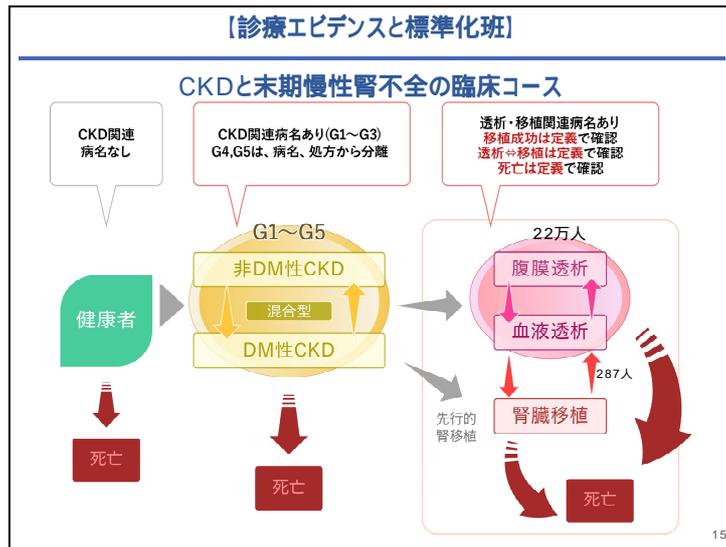
腎移植関係の病名

処置コード	処置名	NDB検索件数-レセプト
150338610	生体腎移植術	21件
150196310	同種死体腎移植術	39件
150324810	同種死体臍腎移植術	10件以下

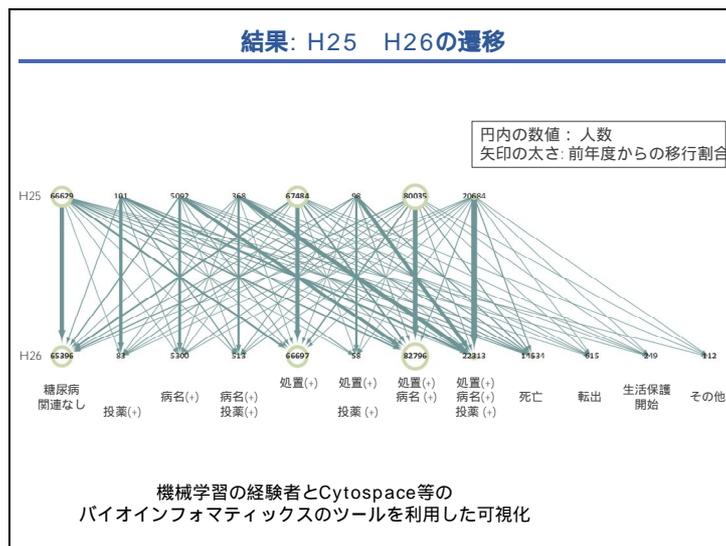
病名コード	病名	ICD10	NDB検索件数-レセプト
8835575	腎移植拒絶反応	T861	308件
8835577	腎移植不全	T861	127件
8846302	腎移植急性拒絶反応	T861	10件以下
8846303	腎移植慢性拒絶反応	T861	10件以下
8847618	死体腎移植後	Z940	10件以下
8847643	生体腎移植後	Z940	10件以下
8847671	脳死腎移植後	Z940	10件以下
9968003	腎移植後	Z940	9869件

移植のレセプトは、紙で請求されることが多いため、処置コードだけではなく、病名を用いることで把握可能性が増す。

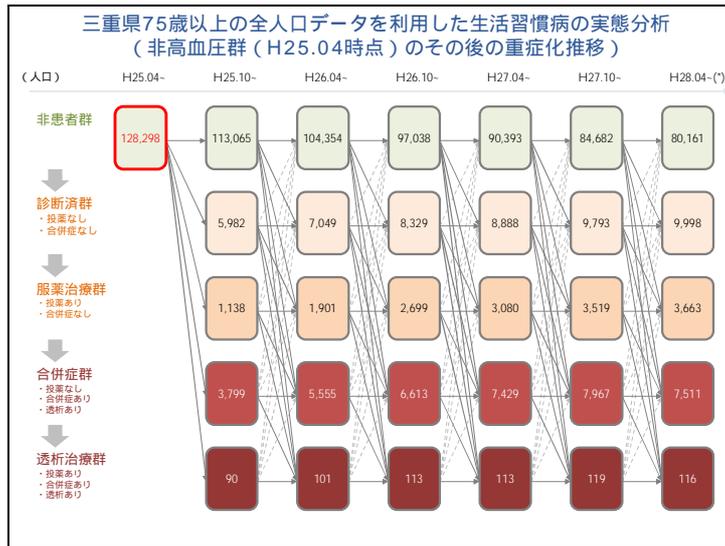
スライド 15



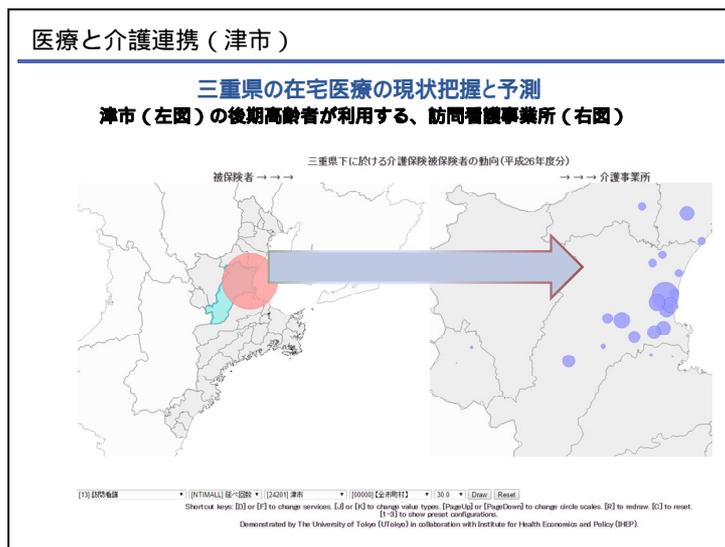
スライド 17



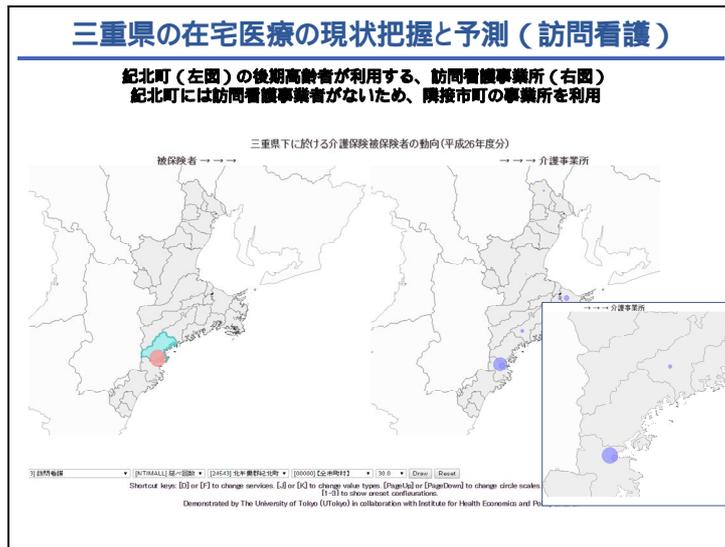
スライド 18



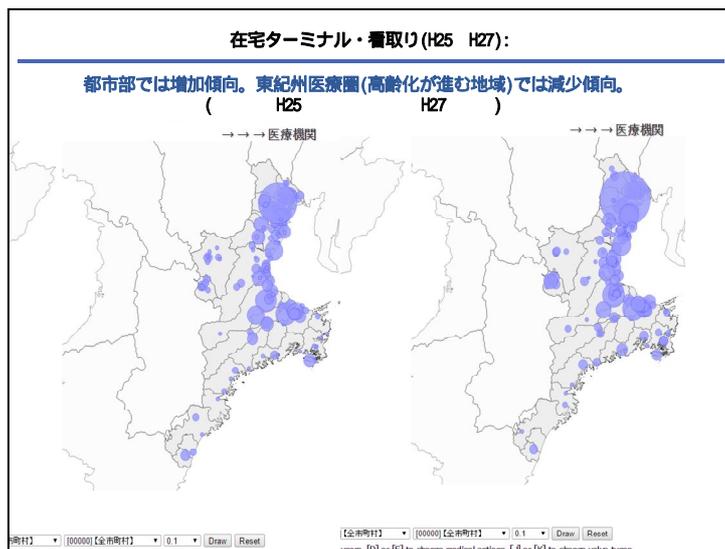
スライド 20



スライド 21



スライド 22



研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年	ページ
なし							

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
満武巨裕	レセプトビッグデータ解析の現状と将来	実験医学	第34巻第5号	799-804	2016