

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はないこと、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、医薬品コード等などから定義）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

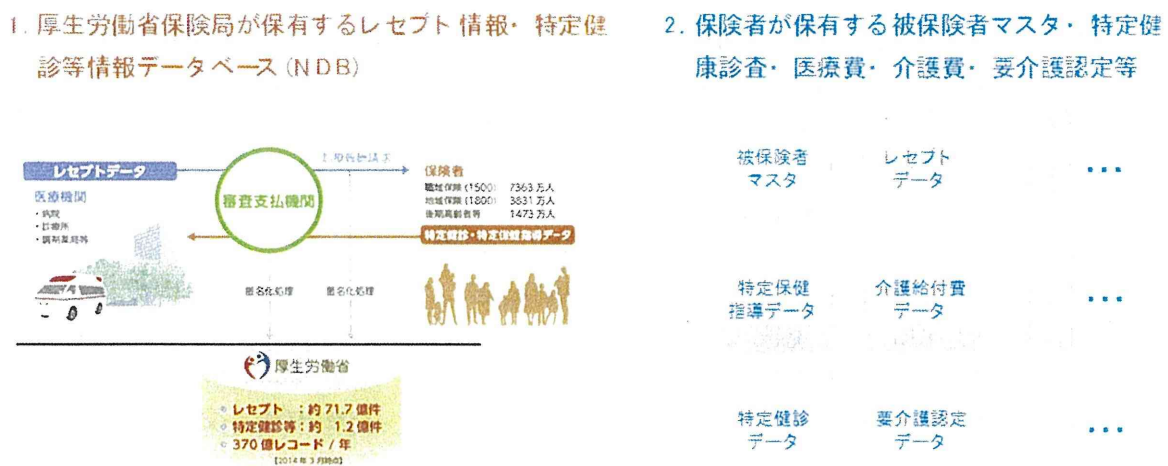
特定健診データやレセプト情報を活用した後期高齢者の支援金の加算・減算の評価は、当初の計画を実行できなかった。したがって、特定健診・特定保健指導の評価指標や参酌標準は、保険者の保健活動を評価し、且つ医療費適正化のインセンティブを有するものに再検討する必要がある。そこで、NDB レセプト情報と特定健診データを活用した新しい1次・2次・3次予防群の確立、保険者のレセプト情報と介護レセプトを活用し訪問診療をはじめとする在宅医療の実態を明らかにするとともに、疾病予防・医療・介護の連携の実態把握を行う。

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



#### I 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。

## II 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：三重県の国民健康保険（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（神奈川県のデータは、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、従来のな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。一人の患者に関す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

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には、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



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国民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)。ID1 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 that 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)。

### 4) 多次元分析 (クリニカルクエスション)

(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Ⅱ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、臨床上の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。

連携学会	
 一般社団法人 日本糖尿病学会	一般社団法人 日本糖尿病学会
 特定非営利活動法人 日本高血圧学会 The Japanese Society of Hypertension	特定非営利活動法人 日本高血圧学会
 一般社団法人 日本動脈硬化学会	一般社団法人 日本動脈硬化学会
 一般社団法人 日本腎臓学会 JAPANESE SOCIETY OF NEPHROLOGY	一般社団法人 日本腎臓学会

## 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。

## 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。

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

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

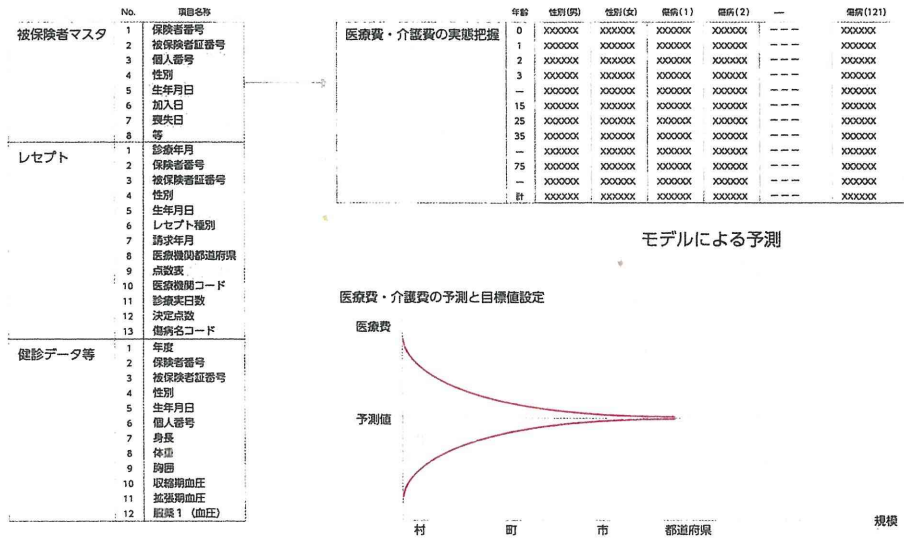
医療費・介護費の予測は、経済協力開発機構(OECD)の医療費予測モデルを踏まえて、日本版の短期予測および中長期予測モデルの3つを作成する。

- (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)



## 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。

## 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ケージし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) 一レセプト一レコード
- (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ケージしたデータセット
- (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