

201301029A

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の
利活用に関する研究

平成 25 年度 総括・分担研究報告書

研究代表者 大江 和彦

平成 26（2014）年 3 月

目 次

I. 総括研究報告書

- 1 レセプト情報・特定健診等情報データベースの利活用に関する研究-----1
大江 和彦（東京大学医学部附属病院企画情報運営部）

II. 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の利活用に関する研究-----7
（データ精度管理部門）
今中 雄一（京都大学大学院医学研究科 医療経済学分野）
- 2 レセプト情報・特定健診等データベースの申出者対応部門の充実-----15
満武 巨裕（一般社団法人 医療経済研究・社会保険福祉協会
医療経済研究機構）

- 参考資料 1. 各国別レセプト情報の二次利用状況
- 参考資料 2. 韓国 HIRA におけるレセプト情報提供の概要
- 参考資料 3. 台湾における健康保険サンプリングデータ
- 参考資料 4. 米国 AHRQ のデータセンターで利用できるデータファイル

III. 研究成果の刊行に関する一覧表 -----75

厚生労働科学研究費補助金（地域医療基盤開発推進研究事業）
総括研究報告書

レセプト情報・特定健診等情報データベースの利活用に関する研究

研究代表者 大江 和彦 東京大学医学部附属病院企画情報運営部 教授

研究要旨

【目的】レセプト情報・特定健診等情報データベースは、医療費適正化計画作成のための調査等に用いることを目的に国が収集する悉皆性の高いデータであり、研究者等への第三者提供も開始されるなど、今後の利活用が期待されている。しかし有識者会議における議論では、(1) 試行期間中に承諾された申出数は少数にとどまり、その理由の一つに、申出者が本データについて正確に把握できていないことが挙げられており、申出者との情報共有を進める必要があること、(2) 利用者にとって分析に適したデータ構造となっておらず、格納されたデータ自体にも不具合がしばしばあり、データの精度管理は不可欠の課題であるが、データ管理の手法や体制が構築されていないこと、が指摘された。そこで本研究では①「申出者対応部門の充実」、②「データ精度管理」を推進していくことを目的としている。【方法】①のための調査対象として、レセプトデータを既に研究者へ提供している韓国、台湾の提供状況、および米国のオンサイトセンタを調査した。また②ではデータリンケージの課題と提案のためデータ生成からデータベース作成に至る過程での個人IDの整合性にかかる問題点を文献的に整理した。また、レセプトデータの格納状況に関する実態調査として、研究目的の二次利用のために提供を受けた入院・外来レセプトデータの格納状況について、エラーの種類別に集計を行った。【結果】① 韓国では、外来・入院のレセプトをランダムサンプリングしたデータセットを提供している。また、小児と高齢者のレセプトのみのランダム抽出データも提供している。また研究者・行政関係者が利用できるオンサイトセンタが開設されていた。台湾では、診療報酬データ（レセプトデータ）等を、国民医療保険局が収集しデータベースを構築しており、その管理を国家衛生研究院が行っていた。2000年、2005年、2010年時点にそれぞれ20万人、100万人、100万人規模の3種類のランダム抽出データセットを作成して提供しており2013年には270件提供実績があった。米国ではAHRQがMEPSデータセットを研究者に対して提供しており、研究計画がAHRQによって承認されれば、一般公開されていない機密データ項目がオンサイトセンタを利用することで可能となっていた。② データリンケージの課題分析では、氏名や保険者証情報などでレセプト間と特定健診データとの間での突合のための障害要因が明らかになってきた。また、レセプトデータの格納状況に関する実態調査では、データ自体の不備や逸脱値はあまりないものの、レセプト審査の過程や再提出などの過程でのデータ重複や遅れにともなう見かけ上の欠損や、受療継続期間中のデータ追跡可能性に課題が一定程度存在することが示唆された。【結論】米国、韓国、台湾のデータ

提供の仕組みを調査し日本への適用可能性を検討した。韓国に関しては、HIRA に設置されているオンサイトセンタの運用、台湾の 100 万人規模のコホートデータの作成参考になると考えられる。また、米国と韓国のオンサイトセンタの機能は、来年度開設予定の日本のオンサイトセンタのサービスを考える上で有益であると考えられた。レセプト情報の活用にはデータの特性と限界を踏まえて分析を行う必要がある。また、本データベースの精度改善に向けてはデータ作成時点まで遡って各段階で改善できる可能性があり、とりわけ生涯通じて一意の個人 ID の実装が強く望まれる。

研究分担者氏名・所属機関名 役職

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

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

A. 研究目的

レセプト情報・特定健診等情報データベースは、高齢者の医療の確保に関する法律に基づき、医療費適正化計画の作成、検討のための資料を作成することを目的に国が収集しているデータである。本データベースの水準にまで網羅的に収集されたレセプト情報、特定健診等情報はわが国において他に存在しない。加えて、医療サービスの質の向上を目的とする公益性の高い研究等にも活用する観点から、本データの第三者への提供が 2011 年度より試行的に開始されており、これは我が国唯一の試みである。しかし、提供をすすめていくにつれ、現在有識者会議で以下の問題点が指摘されている。

① 第 1 回のデータ提供では 43 件の申出に対して承諾申出は 6 件にとどまるなど、申出が承諾される割合は必ずしも高いものではなかった。申出者の研究が実際の格納データでは実現困難で、データ提供にあたっての各種要件や手続きを申出者が十分に理解していないなど、データ利用に関して必要な事項を申出者が

十分に把握できていないこと、またそれに対する提供側の情報提供が不十分であったことがその理由の一つとして指摘されている。

② 本データベースに格納されている情報はレセプト電算形式をそのまま踏襲した形式となっているため、表記が省略されたままになっている箇所があるなど、利用者にとって分析に適したデータ構造とはなっていない。加えて、格納されているデータ自体に欠損、不具合がしばしば確認されており、データ精度の管理が不可欠であるが、現状ではこうしたデータ欠損、不具合は実際にデータの利用者からの偶発的な情報提供及びそれへの対応が行われているのみで、データを定常的に管理する手法や体制が構築されておらず、データそのものが不安定で活用が難しい。また、現在レセプト情報等の紐付けを主として保険者番号や被保険者記号、番号等に依拠するハッシュ値 1、および主として名前に依拠するハッシュ値 2 によって行うこととなっているが、これらハッシュ値が、名前や保険者番号等の記述の揺れなどによって必ずしも一意性が保たれていない。このため、経時的、あるいは施設横断的に患者を追いかける調査が現状では困難である。

本研究は、こうした課題に対応するべく①「申出者対応部門の充実」、②「データ精度管理」を推進していくことを目的としている。

① 申出者へのレセプトデータ提供サービス

機能をすでに有している諸外国の仕組みを調査することで、今後の日本における適切な体制作りについて検討する。

② 下記の2点を目的に本研究を実施した。

- i. レセプトデータならびに特定健診データの二次利用に際し、両データにおいて個人に関する主キーの整合性が破られる原因を整理する。その上で、データベース内の個人IDの整合性とデータベース間の個人IDの整合性について構造上の課題を整理する。
- ii. レセプトデータにおけるデータリンケージの精度を実データに基づいて検証し、現段階での本データベースの限界を明らかにする。

B. 研究方法

① 分担研究者の満武らにより、以下の方法で実施した。調査対象国としたのは、レセプトデータを既に研究者へ提供している韓国、台湾および米国のオンサイトセンタである。前二者の調査項目は、レセプトデータベース名称および担当機関、提供データの概要、データ提供開始時期、民間提供の有無、個人情報保護に関する取扱い根拠、漏洩等の不適切利用に対する対応およびその根拠、の6項目を共通項目とした。である。米国のオンサイトセンタとしてDOH(U.S. Department of Health & Human Services)傘下のAHRQ(Agency for Healthcare Research and Quality)が保持する医療費関連データ MEPS(The Medical Expenditure Panel Survey)を研究者に対してAHRQ施設内での利用環境を提供しているデータセンタを調査した。

② 分担研究者の今中らにより、以下の方法で実施した。データリンケージの課題と提案のため、データ生成からデータベース作成に至る過程での、個人IDの整合性にかかる問題点を文

献的に整理した。また、レセプトデータの格納状況に関する実態調査として、研究目的の二次利用のために提供を受けた、2010年4月から2012年3月診療分までの8府県の全ての保険者に関する入院・外来レセプトデータの格納状況について、エラーの種類別に集計を行った。

C. 研究結果と考察

① 韓国では、外来・入院のレセプトをランダムサンプリングしたデータセットを提供している。また、小児と高齢者のレセプトのみのランダム抽出データも提供している。この他に申し出者の要望に応じてDWから独自にデータ抽出を行い、データ提供を行うサービスも有する。民間に対しては、提供していない。加えて、HIRAではレセプトデータを施設内で、研究者・行政関係者が利用できるオンサイトセンタが開設されている。

台湾では、診療報酬データ(レセプトデータ)等を、国民医療保険局が収集しデータベースを構築しており、その管理を国家衛生研究院が行っている。2000年、2005年、2010年時点にそれぞれ20万人、100万人、100万人規模の3種類のランダム抽出データセットを作成して提供しており2013年には270件提供実績がある。

米国ではAHRQがMEPSデータセットを研究者に対して提供している。加えて、研究計画がAHRQによって承認されれば、一般公開されていない機密データ項目がオンサイトセンタを利用することで可能となる。

このように日本と同様の皆保険制度であり診療報酬点数制度も導入されている韓国と台湾では、研究者への年間のデータ提供件数が100件を超えている。また、オンサイトでは、一般公開されていないデータ項目も利用が可能(全てのデータ項目ではなく、審査承認も必要である)というメリットもあると同時に、オンサイトセンタ内のデータコーディネーターに更新プログラムを送付することで、データコーディネーターが研究

者の代わりに実行するサービスもあることが特徴的である。韓国 HIRA においては、研究者が直接来訪して利用しなければならないというアクセス性および便宜性に関する指摘もされている。米国と韓国の知見は、来年度開設予定の日本のオンサイトセンタのサービスを考える上で有益であると考えられた。

② データリンケージの課題分析においては、まずレセプト情報については、社会保険診療報酬支払基金、国民健康保険中央会で共通のチェックロジックに沿って、データの精度が管理されている。医療機関間での個人氏名データの整合性は必ずしも担保されるわけではない。その他の項目については、性、生年月日など、個人での首尾一貫性はチェック対象ではないため、入力ミスのまま格納されている可能性もある。

また、特定健診情報については、氏名はレセプトデータと異なり、全角カタカナ入力が規定されており、データベース内での整合性は担保されている。しかし、その他の項目についてはレセプトデータと同様の事務点検のレベルであるといえる。

また、ハッシュ値生成の元情報となる保険者証記号番号はレセプトと特定健診で分割方法が異なり、レセプト情報では記号と番号の間に「- (ハイフン)」が入る可能性を残すなど、単純にレセプト情報と特定健診情報で記号番号が一致するとは考えがたい状況にある。また、生年月日についてはレセプト情報では暦年で入力されるのに対し、特定健診情報では西暦で入力されるため、両データベース間での整合性が損なわれる可能性がある。ID②に関しては、氏名の入力規則が両データベースで異なるため、突合可能なレコードは限られることが危惧される。なお、採用している仮名化の方法に限界があることも公表されており、氏名の記載ルールが各データで異なるためにデータベース間の連携が困難であることも公表されている。

レセプトデータの格納状況に関する実態調査

においては、1) 性別や年齢の不備など、入力規則から逸脱したレコードは認められなかった。

(2) 医療費がマイナスのレセプトは DPC レセプトに認められるものの、異常値を疑わせるレコードは認められなかった。(3) 返戻前後のレセプトの重複登録は、都道府県・保険者種別・年度と関係なく、全体の約 0.05%存在した。(4) 診療月の翌々月までに 99.3%が審査された一方、審査に 2 年を要したケースも散見された。(5) 24 ヶ月間の入院症例のうち、退院月まで連続して毎月のデータを追跡しえたのは 89.9%であった。

(6) 12 ヶ月後まで追跡しうる被保険者の個人 ID は、保険者間で 47.0%~91.6%まで大きく異なったが、2 型糖尿病に限定した場合は 71.5%~94.3%であった。

D. 結論

米国、韓国、台湾のデータ提供の仕組みを調査し、日本への適用可能性を検討した。韓国に関しては、HIRA に設置されているオンサイトセンタの運用、台湾の 100 万人規模のコホートデータの作成参考になると考えられる。また、米国と韓国のオンサイトセンタの機能は、来年度開設予定の日本のオンサイトセンタのサービスを考える上で有益であると考えられた。

レセプト情報の活用にはデータの特性と限界を踏まえて分析を行う必要がある。また、本データベースの精度改善に向けてはデータ作成時点まで遡って各段階で改善できる可能性があり、とりわけ生涯通じて一意の個人 ID の実装が望まれる。

G. 研究発表

1. 論文、書籍発表

なし

2. 学会発表

1. 満武巨裕（参考人）、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」の進捗状況について（資料 3-1）、第 17 回レセプト情報等の提供に関する有識者会議，平成 25 年 12 月 19 日
2. 猪飼 宏, 大坪徹也, 今中雄一, 全国網羅的なレセプト情報の有効活用に資するデータベース基盤の課題と設計. 第51回日本医療・病

院管理学会学術総会 2013. 9. 28 京都（抄録：日本医療・病院管理学会誌 第50巻Supplement p. 235. 2013年）

H. 知的財産権の出願・登録状況 （予定を含む。）

なし

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の利活用に関する研究
（データ精度管理部門）

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
研究協力者 猪飼 宏 京都大学大学院医学研究科 医療経済学分野 特定講師
研究協力者 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

研究要旨

レセプト情報・特定健診等情報データベースの利活用を促進するために、同データベースに格納されたデータの質について検討を行った。

第一に、データの生成過程を文献的に調査検討することによりデータリンケージの構造上の課題を明らかにし、さらにその解決に向けた提案を行った。

第二に、研究目的の二次利用のために提供を受けた、2010年4月から2012年3月診療分までの8府県の全ての保険者に関する入院・外来レセプトデータの格納状況について調査を行い、以下の所見を得た。（1）性別や年齢の不備など、入力規則から逸脱したレコードは認められなかった。（2）医療費がマイナスのレセプトはDPCレセプトに認められるものの、異常値を疑わせるレコードは認められなかった。（3）返戻前後のレセプトの重複登録は、都道府県・保険者種別・年度と関係なく、全体の約0.05%存在した。（4）診療月の翌々月までに99.3%が審査された一方、審査に2年を要したケースも散見された。（5）24ヶ月間の入院症例のうち、退院月まで連続して毎月のデータを追跡しえたのは89.9%であった。（6）12ヶ月後まで追跡しうる被保険者の個人IDは、保険者間で47.0%~91.6%まで大きく異なったが、2型糖尿病に限定した場合は71.5%~94.3%であった。

レセプト情報の活用に際してはデータの特性と限界を踏まえて分析を行う必要がある。また、本データベースの精度改善に向けてはデータ作成時点まで遡って各段階で改善できる可能性があり、とりわけ生涯通じて一意の個人IDの実装が望まれる。

A. 研究目的

レセプトデータならびに特定健診データの二次利用に際し、各データベース単独での利活用のみならず、併用することで多様な研究の実践が可

能となる。単独、併用といったいずれの活用形態にせよ、各データに共通の長所として、個人単位での情報を利活用できるという点が挙げられる。単独活用においては、経時的な個人単位の情報の捕捉が有用であり、併用においては、両データベ

ースを含めた包括的なテーブル間のリレーションを図り、情報項目の拡張を実現することが期待される。しかし、現行のレセプト情報・特定健診情報等データベースにおいては、いずれの個人単位での情報の運用をも実現困難である恐れが指摘されている。その原因は、ハッシュ関数を用いて生成された個人を同定するための主キーであるに個人 ID が、各データベースにおいて 2 つの方式で生成、運用され（表 1）、首尾一貫性がないことにある。

以上を踏まえ、下記の 2 点を目的に本研究を実施した。

1. 両データにおいて個人に関する主キーの整合性が破られる原因を整理する。その上で、データベース内の個人 ID の整合性とデータベース間の個人 ID の整合性について構造上の課題を整理する。

2. レセプトデータにおけるデータリンケージの精度を実データに基づいて検証し、現段階での本データベースの限界を明らかにする。

B. 研究方法

【1】 データリンケージの課題と提案

データ生成からデータベース作成に至る過程での、個人 ID の整合性にかかる問題点を文献的に整理した。

【2】 レセプトデータの格納状況に関する実態調査

研究目的の二次利用のために提供を受けた、2010年4月から2012年3月診療分までの8府県の全ての保険者に関する入院・外来レセプトデータの格納状況について、エラーの種類別に集計を行った。

(ア) 入力規則から逸脱したレコードの存在について

下記について入院外・入院それぞれの以下レセプトデータを集計した。

- ① 男女区分の格納状況
- ② 年齢層の格納状況
- ③ 医療費の異常値発生動向の把握
- ④ 診療実日数の異常値発生動向の把握

(イ) レセプト単位での重複データの存在について

氏名（個人番号）、医療機関コード、診療科コード、診療年月、入院/外来でレセプトが一意になるものとし、これらが同一のレセプトが重複して存在するケースを集計した。ただし、審査月として参照した期間は2年である。

(ウ) 返戻データの保管に要する月数について

診療年月が平成22年4月のレセプトについて、以降の審査月別の件数を集計し、累積割合を求めた。

(エ) 入院単位での集計が困難なケースの存在について

一入院の連続する診療月において、いずれかの診療月のデータが欠測しているエピソード数を集計した。また、入院月数分のレセプトデータのうち一部が欠測している入院エピソードについて、欠測月のレセプトが、別の個人番号にて取り扱われていないかどうかを調査した。

(オ) 個人 ID 追跡可能性について

被保険者に対し、何か月まで同一個人 ID のレセプトが発生し、追跡が可能であるかを調査した。1) 全レセプトを対象とした集計と2) 2型糖尿病患者を対象とした集計を行った。

ただし、データ参照期間内で後期高齢者への異動がある被保険者、または死亡症例は除外した。データ参照期間内で年齢の変化が無い被保険者を対象とする理由は、後期高齢者

に異動すると、個人 ID が変更となり、正確に追跡できないためである。

すなわち、平成 22 年 4 月～平成 23 年 3 月にレセプトが発生している被保険者から、下記の条件に合致する被保険者を対象とした。

- 1)年齢層が 70 歳未満の全被保険者
- 2)年齢層が 70～74 歳で、且つ平成 24 年 3 月のレセプトが存在する被保険者
- 3)年齢層が 75 歳以上で、且つ平成 22 年 4 月のレセプトが存在する被保険者

C. 研究結果

【1】 データリンケージの課題と提案

データベース内の個人 ID の整合性とデータベース間の個人 ID の整合性について整理する。

1) データベース内の個人 ID の整合性

現行の生成源項目を用いた仮名化では、個人に対して経年的に整合性を保持することが困難である。理由は大きく 2 つ挙げられる。まず、わが国の医療保険制度下において個人が保険者間を異動することや、結婚等により氏名が変更することが少なからず発生し、個人 ID は変更の経緯を完全に考慮したものとはなっていない点である。なお、個人 ID が二種作成される意図は、個人の情報の変更に対し、可能な限り個人 ID の整合性を高めることにある。二点目は、個人 ID の生成源項目の入力精度が必ずしも「レセプト情報・特定健診情報等データベース」内で個人単位での整合性を担保しうるとは限らない点である。

レセプト情報については、表 2 で示す社会保険診療報酬支払基金、国民健康保険中央会で共通のチェックロジックに沿って、データの精度が管理されている。このルールを鑑みると、被保険者証記号番号は規定文字で記録されていれば、事務点検では許容されると解釈できる。氏名では、規定文字であれば漢字、

カタカナ、英字など入力規則は医療機関の裁量で決定されるため、医療機関間での個人氏名データの整合性は必ずしも担保されるわけではない。その他の項目については、性、生年月日など、個人での首尾一貫性はチェック対象ではないため、入力ミスのまま格納されている可能性もある。こうしたレセプトデータにおける個人の情報について、支払基金では被保険者マスタを作成・管理せず、国保連合会では各連合会で独自に作成・管理するといった自助努力がなされているが、現行のレセプト情報・特定健診情報等データベースではこうした支払機関による被保険者マスタの反映は実装されていない。

一方、特定健診情報については、表 3 の通りである。氏名については、レセプトデータと異なり、全角カタカナ入力が規定されており、データベース内での整合性は担保されている。しかし、その他の項目についてはレセプトデータと同様の事務点検のレベルであるといえる。

2) データベース間の個人 ID の整合性

表 4 に示すとおり、データベース間の連携は、「レセプト情報・特定健診情報等データベース」で具備すべき機能であると掲げられている。この要件に準じ、各データベースの個人 ID は類似した生成源項目を使用している（表 1）。しかし、ID①に関する保険者証記号番号は、レセプト情報では 1 カラムとして記号番号が管理され、特定健診データでは記号と番号それぞれ 1 カラムで管理されている。したがって、レセプト情報では記号と番号の間に「-（ハイフン）」が入る可能性を残すなど、単純にレセプト情報と特定健診情報で記号番号が一致するとは考えがたい状況にある。また、生年月日についてはレセプト情報では暦年で入力されるのに対し、特定健診情報では西暦で入力されるため、両データベース間での整合性が損なわれる可能性

がある。ID②に関しては、氏名の入力規則が両データベースで異なるため、突合可能なレコードは限られることが危惧される。なお、採用している仮名化の方法に限界があることも公表されており、氏名の記載ルールが各データで異なるためにデータベース間の連携が困難であることも公表されている。⁵

【2】 レセプトデータの格納状況に関する実態調査

(ア) 入力規則から逸脱したレコードの存在について

① 男女区分の格納状況

表 5 に示した。仮名化後のデータに異常値の入力はみられない。

ただし、個人を意味する ID の生成に男女区分情報が使用されているため、同一個人での男女区分情報の整合性は評価できていない。

② 年齢層の格納状況

表 6 に示した。仮名化後のデータに異常値の入力は見られない。

ただし、個人を意味する ID の生成に生年月日情報が使用されているため、同一個人での生年月日情報の整合性は評価できていない。

③ 医療費の異常値発生動向の把握

表 7・表 8 に示した。全レセプト件数に占める医療費が異常値である件数の絶対数は極めて小さい。

医療費が異常値のレセプトの多くは合計点数の調整に関連するものまたは返戻分であると考えられる。表 8 より、合計点数がマイナスのレセプトは全て DPC レセプトであるため、調整分のレセプトであるといえる。

④ 診療実日数の異常値発生動向の把握

表 9 に示した。入院および外来いずれも、診療実日数がひと月の日数と矛盾するレセプト（日数末日超）が稀に登録されていた。診療実日数に矛盾のあった入院レセプトの診療実日数は 60 日であり、外来レセプトは診療月が 11 月であったが診療実日数は 31 日であった。返戻対象となったレセプトの可能性はある。

(イ) レセプト単位での重複データの存在について

表 10-1～表 10-3 に示した。いずれの請求年度の医科および DPC レセプトにおいても、わずかながら、レセプトの重複登録の存在が確認された。また、その発生割合においては都道府県や保険者種別による一定の傾向は見られなかった。

重複データはわずかではあるが存在するため、研究デザインによっては、重複データを管理する必要が生じる可能性がある。

(ウ) 返戻データの保管に要する月数について

表 11 に示す。審査月の翌々月までで、99.5%のレセプトが審査済みであった。

(エ) 入院単位での集計が困難なケースの存在について

表 12 に示す。約 4% (⑦/①) は一入院で集計できないことが明らかとなった。支払確定データ以外のデータを含めてすべて収集し、データベース管理者がデータクリーニングすることが望まれる。

(オ) 個人 ID 追跡可能性について

表 13・図 1 で全レコードでの月別追跡結果を保険者別に示し、表 14 で 2 型糖尿病患者に限った集計、表 15・図 2 で年齢層別の集計結果を示す。

2010 年 4 月から 2011 年 3 月において最初に確認された個人 ID について、同一個人 ID

のレセプトの発生状況を 12 か月間にわたって把握した。

ここでの個人 ID 追跡割合は、同一人物において個人 ID の変更があった場合を「追跡不可能」とみなすことに加えて、受療機会がないためにレセプトが発生せず、同一人物において個人 ID の変更があったかどうかを判定できない場合も「追跡不可能」とみなしているため、過小評価となっている可能性がある。すなわち、若年層では追跡可能であるにもかかわらずレセプトがないために追跡不可能とみなした個人 ID の割合が高齢者よりも高いという偏りが生じる可能性がある。そこで、年齢層別や二型糖尿病の記載のあったレセプトを持つ個人を対象に集計をしたところ、この指標は受療頻度の影響を大きく受けるという限界が確認された。

こうした指標上の限界はあるものの、レセプトデータの発生状況から、後期高齢者医療制度では約 9 割の被保険者は 1 年間追跡可能であることがわかった。また、社会保険や共済組合保険では約 5 割であり、保険者によって追跡可能性が異なった。

D. 考察

【1】 データリンケージの課題と提案

このように、レセプトデータおよび特定健診データが二次利用を想定して運用されてこなかったために、あるいは、個人 ID の生成源での入力規則が個人単位での整合性を担保することを意図してこなかったために、レセプト情報・特定健診情報等データベースで求められるデータリンケージは本来期待される精度を保持し得ない。データ精度の向上に向けて、下記の抜本的提案が考えられる。

〔改善案〕

1. 匿名化前に、個人 ID に係る項目を、保険者において標準化すること。(1 章参照)

2. レセプト情報と特定健診情報において、データフォーマットとチェックロジック ("ASP チェックロジック") を統一すること。保険者を超えて統一し、二次利用に耐えうるデータとするためにチェックロジックを構築することが望まれる。

3. 同一の「医療機関番号+カルテ番号」であれば、異なる個人 ID を同一人物とみなすなどにより、個人の識別の精度が高まるのではないかと。

〔データリンケージに関する抜本的な解決策〕

4. 一生涯を通じて唯一の個人番号を個人 ID として実装すること。

〔その他〕

5. 医療機関コードの変更が生じることはよくあるが、その際には入院単位での集計が困難となるため、データベース管理者は医療機関コードの変遷を把握することが望まれる。

【2】 レセプトデータの格納状況に関する実態調査

入力規則から逸脱したレコードの存在はまれであった。返戻データに起因すると考えられるデータの重複や、データ収集の遅れが

E. 結論

レセプト情報の活用の際にはデータの特性と限界を踏まえて分析を行う必要がある。また、本データベースの精度改善に向けてはデータ作成時点まで遡って各段階で改善できる可能性があり、とりわけ生涯通じて一意の個人 ID の実装が望まれる。

G. 研究発表

1. 論文、書籍発表

なし

2. 学会発表

全国網羅的なレセプト情報の有効活用に資するデータベース基盤の課題と設計. 第 51 回日本医療・病院管理学会学術総会 2013.9.28 京都 (抄録: 日本医療・病院管理学会誌 第 50 巻 Supplement p. 235. 2013 年)

H. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

参考文献

1. 厚生労働省. 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>
2. 社会保険診療報酬支払基金. 受付・事務点検 ASPに係るチェック一覧 平成24年7月版.
3. 社会保険診療報酬支払基金. レセプトオンライン請求 受付・事務点検 ASPに係るチェックロジック 平成25年5月
4. 社会保険診療報酬支払基金. 特定健診・保健指導システム エラーコード表 平成25年4月版.
5. 厚生労働省保険局総務課 保険システム高度化推進室. レセプト情報等のデータ構造について. 平成25年1月17日
<http://www.mhlw.go.jp/stf/shingi/2r985200002ss9z-att/2r9852000002ssf5.pdf>

レセプト情報・特定健診等情報データベースの申出者対応部門の充実

研究分担者 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

本報告書では、申出者へのレセプトデータ提供サービス機能をすでに有している諸外国の仕組みを調査することで、今後の日本におけるレセプト情報・特定健診等情報データベースの申出者対応部門の適切な体制作りについて検討する。

調査対象国としたのは、レセプトデータを既に研究者へ提供している韓国および台湾である。調査項目は、レセプトデータベース名称および担当機関、提供データの概要、データ提供開始時期、民間提供の有無、個人情報保護に関する取扱い根拠、漏洩等の不適切利用に対する対応およびその根拠等である。オンサイトセンタとしては、米国 AHRQ が保持する医療費関連データ MEPS を、研究者に対して AHRQ 施設内での利用環境を提供しているデータセンタを調査した。

韓国では、外来・入院のレセプトをランダムサンプリングしたデータセットを提供している。また、小児と高齢者のレセプトのみのランダム抽出データも提供している。この他に申し出者の要望に応じて DW から独自にデータ抽出を行い、データ提供を行うサービスも有する。データ提供件数は、2012 年は 69 件、2013 年は 115 件であり、DVD 媒体で提供される。民間に対しては、提供していない。加えて、HIRA ではレセプトデータを施設内で、研究者・行政関係者が利用できるオンサイトセンタが開設されている。オンサイトセンタの PC は、外部データウェアハウスと回線で接続されており、利用者はアクセスできるものの、結果を持ち帰ることはできない。分析結果は、HIRA が検閲を行い、許可された情報のみが後日、郵送あるいは e メール等で利用者にも送られる仕組みになっている。

台湾では、診療報酬データ（レセプトデータ）等を、国民医療保険局が収集しデータベースを構築しており、その管理を国家衛生研究院が行っている。データベース名称は、全民健康保険研究資料庫である。研究者は、1996 年からの全レセプトデータを利用できる。提供データは、5 種類である。その中でも、ランダム抽出データは、個票（個人単位）データであり、被保険者ファイル(ID)から、ランダムサンプリングして、2000 年データとして、20 万人を抽出し、抽出された被保険者の（1996 年～2007 年の）入院、外来、調剤レセプトデータを提供している。加えて 2005 年単位で 100 万人、2010 年単位で 100

万人のランダム抽出データが作成されたことから、現時点では 3 種類のランダム抽出データが存在する。2013 年は、上記のデータセットが研究者へ 270 件提供された。

米国では AHRQ が MEPS データセットを研究者に対して提供している。加えて、研究計画が AHRQ によって承認されれば、一般公開されていない機密データ項目がオンサイトセンタを利用することで可能となる。オンサイトセンタでは、ユーザーの申請書に基づいて AHRQ があらかじめデータセットを作成し、オンサイトセンタ内の PC にコピーされる。ユーザーはオンサイトセンタにて、そのデータセットを利用して分析を行う。データセットは、オンサイトセンタでしか利用できない。利用のためには、利用者は、初回は必ず AHRQ に来なければならない。その後、更新プログラム(SAS 等)を AHRQ に送付すれば、AHRQ 内のデータコーディネーター（常勤 1 名）がプログラムを代わりに実行する(run)ことも可能である。分析結果は AHRQ 内で審査を受けた後、ユーザーにメールで送られる。本方式によって、ユーザーは継続的に分析を続けることができる。

日本と同様の皆保険制度が導入され、診療報酬点数制度も採用されている韓国と台湾では、研究者への年間のデータ提供件数が 100 件を超えている。日本は、特別抽出の形態で、依頼に応じてデータセットを抽出・提供していたが、近年はサンプリングデータセットとして、1 ヶ月分の医科入院、DPC、医科入院外、調剤レセプトを、抽出及び匿名化処理等を行ったうえで提供している。今後、さらに提供件数を増やすためにも、韓国の小児や高齢者のみのデータセットの提供、さらに進んでいる台湾のような 5 種類のデータセットを準備しておくことも有効な手段であると考えられる。韓国では研究者専用に統計分析のためのオンサイトセンタが用意されていることに加え、HIRA に仮想 PC を接続して統計分析が行える遠隔接続サービスを提供する計画もある。米国では、レセプトデータではないが、医療や保健サービスへのアクセス・医療関連の支払いなどのデータを含む AHRQ の MEPS データをオンサイトセンタで提供している。オンサイトでは、一般公開されていないデータ項目も利用が可能（全てのデータ項目ではなく、審査承認も必要である）というメリットもあるが、加えてオンサイトセンタ内のデータコーディネーターがプログラムを研究者の代わりに実行するサービスも参考になる。韓国 HIRA のオンサイトセンタに関しては、研究者が直接来訪して利用しなければならないというアクセス性および便宜性に関する指摘もされている。米国と韓国の知見は、来年度開設予定の日本のオンサイトセンタのサービスを考える上で有益であると考えられた。

A. 研究目的

レセプト情報・特定健診等情報データベース（以下、NDB）は、高齢者の医療の確保に関する法律に基づき、医療費適正化計画の作成、検討のための資料を作成することを目的に国（担当部局：厚生労働省・

保険局・保険システム高度化推進室）が構築・管理している。NDB は、2009 年度から全ての電子レセプト、2008 年度から導入された特定健康診査（以下、特定健診）・特定保健指導のデータを収集しており、悉皆性の高いデータを保持している。

医療サービスの質の向上を目的とする公益性の高い研究等にも活用するために、本データの研究者等への第三者提供が2011年度から試行的に開始された。しかし、第一回のデータ提供に関しては、43件の申出に対して承諾件数は6件であり、申出件数および承諾割合は必ずしも高いものではない。本データの承諾審査を行う「レセプト情報等の提供に関する有識者会議（以下、有識者会議）」からは、次の点が指摘されている。(1)申出者が求めるデータ項目が実際に格納されているデータでは実現困難であった申請が存在した、(2)データ提供にあたっての各種要件や必要な事項を申出者が十分に把握していない申請が存在する、(3)提供側の情報提供が不十分であった等である。そのために、厚生労働省は、本データの提供依頼申出者からの研究の実現可能性やその具体策等の助言などを行う窓口の設置、データを利用できるオンサイトセンタの設置を検討している。

そこで、本研究では申出者へのレセプトデータ提供サービス機能をすでに有している諸外国の仕組みを調査することで、今後の日本における適切な体制作りについて検討する。

B. 研究方法

調査対象国としたのは、レセプトデータを既に研究者へ提供している韓国および台湾である。調査項目は、以下の6項目を共通項目とした。レセプトデータベース名称および担当機関、提供データの概要、データ提供開始時期、民間提供の有無、個人情報保護に関する取扱い根拠、漏洩等の

不適切利用に対する対応およびその根拠である。

オンサイトセンタとしては、米国のDOH(U.S. Department of Health & Human Services)傘下のAHRQ(Agency for Healthcare Research and Quality)が保持する医療費関連データMEPS(The Medical Expenditure Panel Survey)を、研究者に対してAHRQ施設内での利用環境を提供しているデータセンタを調査した。

C. 研究結果

韓国では、レセプトデータベースの名称をDW(データウェアハウス)と表現しており、特に決まった名称はない。管理・運営は、日本における審査支払機関に該当するHIRA(The Health Insurance Review and Assessment Service)が担当している。提供データは、外来・入院のレセプトをランダムサンプリングしたデータセットである。また、小児と高齢者のレセプトのみのランダム抽出データも提供している。この他に申し出者の要望に応じてDWから独自にデータ抽出を行い、データ提供を行うサービスも有する。データIDは、韓国の国民背番号を利用しており、抽出は1年単位である。データの提供時期は、2012年からである。データ提供件数は、2012年は69件、2013年は115件であり、DVD媒体で提供される(DVD一枚当たり30万ウォン)。民間に対しては、提供していない。ただし、HIRAの統計分析チームが公共部門とみなしてマスコミ・患者団体等の限られた組織に提供している。個人情報保護に関する取り扱いに関しては、レセプ

トは個人情報には該当しないデータとして扱われているが、個人を特定できる可能性のある情報を含んでいるために、実際の運用においては（個人情報に準ずる）「行政機関が有する個人情報の保護に関する法律」に準拠している。漏洩等の不適切利用に対する対応およびその根拠としては、申請者が利用誓約書の事項に違反した場合はあらゆる不利益および民事刑事上の処罰を受け、その後の利用制限が課せられる。

加えて、HIRA ではレセプトデータを施設内で、研究者・行政関係者が利用できる（民間は利用できない）オンサイトセンタが開設されている。入退室管理など、ISMS の理念に則ったセキュリティが課され、1 部屋に約 15 台の PC がセットされている。オンサイトセンタの PC は、外部にデータウェアハウスと回線がつながっており、利用者は分析できるものの、結果を持ち帰ることはできない。得られた結果は、HIRA が検閲を行い、許可された情報のみが後日、郵送あるいは e メール等で利用者に送られる仕組みになっている。

また、韓国の保険公団（保険者）は、HIRA とは別に研究者に対するレセプトデータ提供機能の検討を始めている。提供するデータは、2002 年からのレセプトデータを基に 100 万人をサンプリングしたコホートデータの試行版を作成した。2014 年から、特定の研究者に試行提供が行われる予定になっている。2013 年 12 月には、本データセットの利活用の第 1 回ワークショップが開催された。

台湾では、1995 年に国民皆保健制度が導入され、診療報酬データ（レセプトデー

タ）等を、国民医療保険局（Bureau of National Health Insurance: BNHI）が収集しデータベースを構築しており、その管理を国家衛生研究院（National Health Research Institutes :NHRI）が行っている。データベース名称は、National Health Insurance Research Database（全民健康保険研究資料庫）である。研究者は、1996 年からの全レセプトデータを利用できる。提供データは、5 種類である。（Ⅰ）基本資料データ（基本資料檔）であり、これは集計データである。（Ⅱ）系統抽出データ（系統抽樣檔）であり、個票（個人単位）データである。入院レセプトファイル（DD）は 5%抽出、外来レセプトファイル（CD）は 0.2%抽出となっている。（Ⅲ）特定主題データ（特定主題分檔）は、個票（個人単位）データであり、特定疾患や調査目的に沿って抽出・作成したデータセットであり、悪性新生物、糖尿病、精神疾患、交通事故、リハビリ、漢方薬（中醫藥）等の種類がある。（Ⅳ）ランダム抽出データ（抽樣歸人檔）は、個票（個人単位）データであり、被保険者ファイル（ID）から、ランダムサンプリングして、2000 年データとして、20 万人を抽出し、抽出された被保険者の（1996 年～2007 年の）入院、外来、調剤レセプトデータを提供している。加えて 2005 年単位で 100 万人、2010 年単位で 100 万人のランダム抽出データが作成されたことから、現時点では 3 種類のランダム抽出データが存在する。（Ⅴ）教育用データ（教學用資料檔）は、個票（個人単位）データであり、2000 年のレセプトデータから 1,000 人をランダムサンプリングし、教育用に無償提供している。2013 年は、上記

のデータセットが研究者へ提供された件数は 270 件である。一方、台湾は、民間への提供は、禁止している（過去に製薬会社に、治験業務サポート等の研究目的として試行的に提供していた）。個人情報保護に関する取り扱いに関しては、個人情報コンピュータ処理保護法（電腦處理個人資料保護法(1995 年)）に則った取扱いをしている。漏洩等の不適切利用に対する対応およびその根拠としては、個人情報コンピュータ処理保護法に罰則・罰金規定が定められている。同時に、誓約書などの規定に違反した研究者に対しては、その研究者の研究機関の長、研究者本人、共同研究者の全員に利用停止およびデータの返却を通知している。

オンサイトセンタとしては、米国 AHRQ のオンサイトセンタを調査した。米国では医療費に関する調査として、AHRQ が実施している MEPS(Medical Expenditure Panel Survey)がある。MEPS は家計を対象として医療や保健サービスへのアクセス、医療関連の支払い、健康状態、医療保険の加入状況等を調査しているもので、同一対象者に対して 2 年ずつ調査をし、毎年新しい対象者を加える形式をとっている。1996 年までは National Medical Expenditure Survey として 10 年おきに断面調査が実施されていたが、パネルデータの重要性から 1996 年以降は新しい対象者を加え、毎年調査する形式となっている。

AHRQ では、MEPS データセットを研究者に対して提供している。加えて、研究計画が AHRQ によって承認されれば、一般公開されていない機密データ項目がオン

サイトセンタを利用することで可能となる。また、AHRQ では、研究者の所属・組織によりデータの提供の可否を決めていない。研究プロポーザルの内容が公益性を満たすかにより提供の可否が決まる。例えば、ランド研究所などの民間の研究組織や非営利団体にもデータを提供しており、博士論文 (Dissertation) を執筆中の学生も利用できる。ただし、オンサイトセンタでは、氏名・ID 等の個人の特定コードは提供されない(他の数値に置き換えた形となる)。地域を特定する情報も提供されない。

オンサイトセンタでは、ユーザーの申請書に基づいて AHRQ があらかじめデータセットを作成し、オンサイトセンタ内の PC にコピーされる。ユーザーはオンサイトセンタにて、そのデータセットを利用して分析を行う。データセットは、オンサイトセンタでしか利用できない。利用のためには、利用者は、初回は必ず AHRQ に来なければならない。その後、更新プログラム(SAS 等)を AHRQ に送付すれば、AHRQ 内のデータコーディネーター（常勤 1 名）がプログラムを代わりに実行する(run)ことも可能である。分析結果は AHRQ 内で審査を受けた後、ユーザーにメールで送られる。本方式によって、ユーザーは継続的に分析を続けることができる。オンサイトセンタの運営コストは、フルタイムスタッフ 2 名、場所代、光熱費などで約 300,000 \$ / 年ほどだと推計される。運営コストの正確な公表資料はなく、収入は、約 100 万ドル/年の政府助成金、ユーザーから 300 ドル/人を徴収している。オンサイトセンタは、2002 年に設立・サービス提供が開始

したが、提供件数などの統計は取られていない。オンライセンタの申請件数は、約 30 件／年であり、約 90%が認可される。研究によっては、論文の査読者からのコメントによる再度の分析なども許可しているので、単年度では終わらない。常時、50 件の研究 PJ が進行している状態である。

オンサイトセンタは、セキュリティ入室管理をしている。また、利用者はデータセンタには、メモやマニュアルのみ持ち込が可能で（誓約書で規定）、USB メモリー、携帯電話、スマートフォン、タブレット端末等あらゆる IT 機器の持ち込みは禁止されている。ただし、ビデオでの監視はしていない。

D. 考察

日本と同様の皆保険制度であり診療報酬点数制度も導入されている韓国と台湾では、研究者への年間のデータ提供件数が 100 件を超えている。日本は、特別抽出の形態で、依頼に応じてデータセットを抽出・提供しているが、近年はサンプリングデータセットとして、1ヶ月分の医科入院、DPC、医科入院外、調剤レセプトを、抽出及び匿名化処理等を行ったうえで提供している。今後、さらに提供件数を増やすためにも、韓国の小児や高齢者のみのデータセットの提供、さらに進んでいる台湾のような 5 種類のデータセットを準備しておくことも有効な手段であると考えられる。また、韓国 HIRA では基本的に 1 年単位のデータセットを提供しているが、1 年単位の資料では、分析が不可能な希少疾患など長期間のフォローアップの必要性

が求められていた。そこで、個人情報保護される範囲内で複数年の標本追跡コホートデータセットの作成も始まっている。

韓国では（現時点において、研究班が得ている情報が限定的なために今度も情報収集を継続するが）、研究者専用統計分析のためのオンサイトセンタ（資料処理室 15 席）が用意されていることに加え、HIRA に仮想 PC を接続して統計分析が行える遠隔接続サービスを提供する計画もある。

米国 AHRQ では、医療や保健サービスへのアクセス・医療関連の支払いなどのデータを含む MEPS データをオンサイトセンタで提供している。オンサイトでは、一般公開されていないデータ項目も利用が可能（全てのデータ項目ではなく、審査承認も必要である）というメリットもあると同時に、オンサイトセンタ内のデータコーディネーターに更新プログラムを送付することで、データコーディネーターが研究者の代わりに実行するサービスもあることが特徴的である。韓国 HIRA においては、研究者が直接来訪して利用しなければならないというアクセス性および便宜性に関する指摘もされている。米国と韓国の知見は、来年度開設予定の日本のオンサイトセンタのサービスを考える上で有益であると考えられた。

E. 結論

米国、韓国、台湾のデータ提供の仕組みを調査し、日本への適用可能性を検討した。韓国に関しては、HIRA に設置されているオンサイトセンタの運用、台湾の 100 万人規模のコホートデータの作成参考にな