

2014.01.02/B

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~26 年度 総合研究報告書

研究代表者 大江 和彦

平成 27 (2015) 年 3 月

## 目 次

### I. 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の利活用に関する研究・・・1  
大江 和彦

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  
今中 雄一

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・・・17  
満武 巨裕

参考資料 1・・・・・・・・・・26

II. 研究成果の刊行に関する一覧表・・・・・・・・・・II-1

III. 研究成果の刊行物・別刷・・・・・・・・・・III-1

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））  
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いるが、研究者による利用は必ずしも十分な件数に拡大しておらず、また利用にあたって研究者からみた課題も少なからずある。利用申請や審査に関わる手続き上の利便性等の課題はさておき、研究の質の確保の観点から利活用環境にどのような課題があるかについて検討し今後の検討課題を抽出し改題解決の方向性を提示することが本研究の目的である。そこで本研究では①「申出者対応部門の充実」、②「データ精度管理および利活用推進環境の検討」を検討する。

【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。

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。さらに二次活用を推進するためのデータセット抽出時の病名情報に関する条件指定のあり方の検討、二次活用の具体的なあり方の検討を行った。

【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。性別・年齢区間による患者単位の層化系統抽出を行った事前作成データセット HIRA-NPS を作成し研究者に提供していた。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年時点でそれぞれ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。

二次活用を推進するためのデータセット抽出時の病名情報に関する条件指定のあり方の検討においては、生活習慣病におけるレセプトにおける傷病名出現頻度の検討により、当然含まれるべき傷病名が必ずしも含まれない場合などがあった。これは分析の立場によってデータ抽出条件設定が異なることにより研究対象となる疾患名に大きなズレがおこる可能性を示唆していた。

レセプト情報・特定健診等情報データベースの活用の可能性については、レセプト情報・特定健診等情報の近畿圏限定のデータにより、急性心筋梗塞の性・年齢階層別の院内死亡率を算出し、別途算出された DPC データに基づく算出値と比し、同程度である可能性が示唆された。症例の抽出方法に最新の注意をする必要がある。また、国レベルのみならず、実際の地域レベルで医療のパフォーマンスを示すことができることが示された。一方で、DPC データが整備されている病院でのデータが約 8.5 割を占めると推定され、このような疾患の状況が DPC データでどれほど把握できるかの一参考値を得ることができた。

【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た。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。さらに、日本の NDB レセプト情報等データベースの利用促進のために、汎用性の高い基本データセットの設計と作成が検討され、今年度から試行提供が始まったため、HIRA-NPS は日本の基本データセットを考える上で有益である。

レセプト情報等の活用の際には、データの特性と限界を踏まえて分析を行えば、社会的に価値ある多くのアウトプットが算出できる。また、本データベースの活用環境の改善に向けては、特に生活習慣病において、分析の立場によってデータ抽出条件設定が異なることにより研究対象となる疾患名に大きなズレがおこる可能性があり、データ抽出条件設定における ICD-10 コードによる選択や、臨床的な関連性によるカテゴリー追加などにおいて、なんらかの標準的な傷病カテゴリー選択を指南するリストの作成などが必要であることが強く示唆された。また前年度研究で指摘された生涯通じての一意の個人 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

福祉協会 医療経済研究機構 副部長

#### 研究協力者:

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

佐藤大介・東京大学医学部附属病院企画情報運  
営部助教

満武巨裕・一般財団法人医療経済研究・社会保険

加藤源太・京都大学医学部附属病院診療報酬セ

ンター准教授

猪飼宏・京都大学大学院医学研究科医療経済学  
分野講師

佐々木典子・京都大学大学院医学研究科医療経  
済学分野助教

國澤進・立命館大学 生命科学部生命医科学科  
助教

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

## 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「データ精度管理および利活用推進環境の検討」を目的としている。

## B. 研究方法

① 分担研究者の満武らにより、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。

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。

さらに二次活用を推進するためのデータセット抽出時の病名情報に関する条件指定のあり方の検討、二次活用の具体的なあり方の検討を行った。前者では入院レセプトのうち、診

療年月が2013年3月である電子レセプトのレコードを用いて、下記5種の代表的な治療薬のそれぞれについて、少なくとも1種以上が算定されているレセプトを選択し、その傷病名を集計して分析した。後者ではレセプト情報・特定健診等情報の近畿圏2府6県限定の入院・外来データ(2010年4月から2012年3月診療分)により、急性心筋梗塞の性・年齢階層別の院内死亡率を算出した。さらに、このような疾患の状況がDPCデータでどれほど把握できるかの一参考値として、DPCデータが整備されている病院でのデータの捕捉割合を算出した。

また、別途行われた研究を二次的に考察対象として、データ利活用の可能性を検討した。

### 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。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。さらに、韓国のHIRAが作成したHIRA-NPSデータセットの調査を実施し、韓国国内において1年間に医療機関を利用した全患者対象(約4600万名)を母集団として、性別・年齢(5歳単位)区間による患者単位の層化系統抽出を行ったデータセットであり、1)健康保険に関する診療情報の利活用の増大、2)情報公開を促進させると同時にHIRAのデータ提供に要する業務の効率化、を目的としていた。事前に作成したHIRA-NPSを研究者に提供することで、前述の目的を達成しようとしていた。

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。

二次活用を推進するためのデータセット抽出時の病名情報に関する条件指定のあり方の検討においては、生活習慣病におけるレセプトにおける傷病名出現頻度の検討により、各薬剤について上位100病名が選択されたが、当然含まれるべき傷病名が必ずしも含まれない、対象傷病名の近接ICD10コードを条件に含める必要がある場合などがあった。これは分析の立場によってデータ抽出条件設定が異なることにより研究対象となる疾患名に大きなズレがおこる可能性を示唆していた。

レセプト情報・特定健診等情報データベースの活用の可能性については、レセプト情報・特定健診等情報の近畿圏限定のデータにより、急性心筋梗塞の性・年齢階層別の院内死亡率を算出し、別途算出されたDPCデータに基づく算出値と比し、同程度である可能性が示唆された。症例の抽出方法に最新の注意をする必要がある。また、国レベルのみならず、実際の地域レベルで医療のパフォーマンスを示すことがで

ることが示された。一方で、DPC データが整備されている病院でのデータが約 8.5 割を占めると推定され、このような疾患の状況が DPC データでどれほど把握できるかの一参考値を得ることができた。

## 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た。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。さらに、日本の NDB レセプト情報等データベースの利用促進のために、汎用性の高い基本データセットの設計と作成が検討され、今年度から試行提供が始まったため、HIRA-NPS は日本の基本データセットを考える上で有益である。

レセプト情報等の活用には、データの特性と限界を踏まえて分析を行えば、社会的に価値ある多くのアウトプットが算出できることから利活用の視野、範囲をさらに広げることが前提に体制を整備していくべきである。

本データベースの精度に関しては、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る。また、本データベースの活用環境の改善に向けては、生涯通じて一意の個人 ID の実装が強く望まれるほか、研究者が少ない労力で質の高い研究が実施できるようにするため、マスターテーブルやコードリストなど病名カテゴリーの標準的なセットの作成など、様々な観点からの NDB 利活用支援のための環境整備が必要である。

## G. 研究発表

### 1. 論文、書籍発表

- 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) Naohiro Mitsutake, The use of the national healthcare database for research in Japan, 25th European Medical Informatics Conference, Aug 31st - Sep 3rd, 2014
- 3) 加藤源太、平野景子、赤羽根直樹. レセプト情報・特定健診情報等情報データベースの利活用について—これまでの経緯を踏まえて—. 統計. 2014 年 10 月号 pp. 8-13.
- 4) 今中雄一. 脳梗塞 t-PA 治療の経済評価と均てん化施策. 脳と循環 2014. Sep Vol. 19, No. 3. pp. 49-52.
- 5) 今中雄一. 混合診療と医療改革—新制度 広くデータ評価を—. 経済教室 日本経済新聞 2014 年 7 月 25 日.

## 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)
- 3) 猪飼宏, 今中雄一. DPC, NDB データを用

- いた医療の質の評価. 第52回日本医療・病院管理学会学術総会: 東京, 2014年9月13日-14日. (抄録: 日本医療・病院管理学会誌51Supplement: p25, 2014.)
- 4) 今中雄一. 質の可視化から地域システムの協創へ. 第15回介護保険推進全国サミットinくまもと: 熊本, 2014年10月30日-31日.
- 5) Imanaka Y. "How to balance the health care needs and the regional system." OECD Reviews of Health Care Quality: JAPAN(2014/11/5).
- 6) Imanaka Y. "The importance of databases." OECD Reviews of Health Care Quality: JAPAN(2014/11/5).
- 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））  
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  
総合研究報告書

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 加藤 源太 京都大学医学部附属病院 診療報酬センター 准教授  
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 講師  
研究協力者 佐々木 典子 京都大学大学院医学研究科 医療経済学分野 助教  
研究協力者 國澤 進 立命館大学 生命科学部生命医科学科 助教  
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、同データベースのデータや環境に関する検討を行った。

格納されたデータの質については：

第一に、データの生成過程を文獻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。○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。○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。○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。○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。○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。○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%まで大きく異なっていたが、2型糖尿病に限定した場合は71.5%~94.3%であった。

利活用の推進環境については：

二次活用を推進するための病名情報のあり方と利活用の具体的なあり方について検討した。

（1）二次活用を推進するための病名情報のあり方

異なる解析者間で、同じ病名でも症例の抽出条件・定義が微妙に異なり、一研究を超えた統合的な分析や異なる研究者のアウトプットの比較が困難となる。そこ

で、病名などのマスターデータ等のあり方やコード指定の考え方を実データ分析も踏まえて検討した。

## (2) 二次活用の具体的なあり方

レセプト情報・特定健診等情報データベースの活用の可能性について検討した。レセプト情報・特定健診等情報の近畿圏限定のデータにより、急性心筋梗塞の性・年齢階層別の院内死亡率を算出し、別途算出された DPC データに基づく算出値と比し、同程度である可能性が示唆された。症例の抽出方法に最新の注意をする必要がある。また、国レベルのみならず、実際の地域レベルで医療のパフォーマンスを示すことができることが示された。一方で、DPC データが整備されている病院でのデータが約 8.5 割を占めると推定され、このような疾患の状況が DPC データでどれほど把握できるかの一参考値を得ることができた。

レセプト情報等の活用の際には、データの特性と限界を踏まえて分析を行えば、社会的に価値ある多くのアウトプットを算出する。当該データベースのデータ精度に関しては、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る。また、本データベースの活用環境の改善に向けては、生涯通じて一意の個人 ID の実装のほか、マスターテーブルやコードリストなど病名情報のあり方含め、様々な観点からの環境整備が望まれる。

## 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、データ精度に関しては、下記の 2 点を目的に研究を実施した。

(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。

また、レセプト情報・特定健診等情報データ

ベースの二次利用については、個人単位での情報を利活用できる点、地域の包括的な情報が得られる点などの利点が挙げられる。実際に、病名情報などの取扱を推進するにはどうしたらよいか、また、どのような基本的なアウトプットを出しうるのか（の一例）が、明らかになると、利活用を推進する一つの礎となる。そこで、以上を踏まえ、データ精度に関しては、下記の二点について、実際のデータを取り扱いながら、課題と可能性の検討を行った。

(3) 二次活用を推進するための病名情報のあり方、標準病名リストの検討

(4) 二次活用の具体的なあり方についての検討

## B. 研究方法

### データ精度のついて

(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

## 二次利活用の推進環境について

(3) 二次活用を推進するための病名情報のあり方、標準病名リストの検討

異なる解析者間で、同じ病名でも症例の抽出条件・定義が微妙に異なり、一研究を超えた統合的な分析や異なる研究者のアウトプットの比較が困難となる。そこで、病名などのマスターデータのあり方を実データ分析も踏まえて検討した。

入院レセプトのうち、診療年月が2013年3月である電子レセプトのレコードを用いて、下記5種の代表的な治療薬のそれぞれについて、少なくとも1種以上が算定されているレセプトを選択し、その傷病名を集計した。各区分の薬剤は、別表の薬効分類コードを用いてH24年度版のレセプト電算コードで定義した。

医薬品：

薬価基準収載医薬品コードの頭が、それぞれ下記に該当するもの。

〈糖尿病〉

- ・2492 (インスリンおよびインスリンアナログ)
- ・396 (糖尿病用剤)

〈高血圧〉

- ・213 (利尿剤)
- ・214 (血圧降下剤)

〈高脂血症〉

- ・218 (高脂血症用剤)

病名選択方法：

ICD-10コード表から選択した「糖尿病」「高血圧」「高脂血症」各病名の傷病名リストに、(イ)の上位100病名の中から下記の条件に応じて選

択された病名のみフラグを付した。

「糖尿病」「高血圧」「高脂血症」の各病名をテキスト病名に含むもの、上記レコードに隣接するICD-10コードを含むもの、さらに、臨床的な類似性から各カテゴリに含めるべき可能性があるもの。

集計方法：

上記レセプトのうち、下記の条件をとともに満たす傷病レコードの件数上位100件について、H24年度版の傷病名マスタを用いて変換したテキスト病名・ICD-10コードを添えて出力対象とした。

傷病は医科の「主傷病」、又はDPCの「主傷病又は入院時併存症」に限る。

いずれの場合も「疑い」の記載がある傷病は除く。

また、病名のコード指定のあり方について病名リスト案を作成しつつ考察した。

即ち、①主傷病名については、主傷病名が複数付与されている事例や、ひとつも付与されていない事例などがあり、その際には特定のロジックを用いて主傷病名を一つに定める手法を採用してはいるものの、主傷病名情報が必ずしも患者の疾病動向を正確に反映するものではなく、診療識別情報についても、細目内の細かい情報にまで踏み込んだ情報とはなっていないこと。②一方で、「社会医療診療行為別調査」等々、既存の公的統計においては、レセプトが抱えるこうした限界に対して、特段の処理や修正を加えることなく、そのまま集計作業を行っていること。これらの①や②の間のdiscrepancyについて等、データ提供の際にどういった姿勢で臨むことが必要なのか、といった点について、考察し課題を整理した。

(4) 二次活用の具体的なあり方についての検討

レセプト情報・特定健診等情報データベースの活用の可能性について二つの側面から、検討した。

まずは、レセプト情報・特定健診等情報の近畿圏2府6県限定の入院・外来データ(2010年4月から2012年3月診療分)により、急性心筋梗塞の性・年齢階層別の院内死亡率を算出した。さらに、このような疾患の状況がDPCデータでどれほど把握できるかの一参考値として、DPCデータが整備されている病院でのデータの捕捉割合を算出した。

また、別途行われた研究を二次的に考察対象として、データ利活用の可能性を検討した。

## C. 研究結果

### データ精度について

#### (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。

#### (ア) 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。

#### (イ) 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

### データ精度について

#### (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。

[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不正確な解析に繋がる可能性があり、要注意である。

### 利活用推進環境について

#### (3) 二次活用を推進するための病名情報のあり方、標準病名リストの検討

レセプト情報・特定健診等情報データベース(通称：NDB)は医療費適正化計画作成に用いられるのみならず、特定健診・保健指導の医療費適正化効果の検証や、その他の研究目的への道が開かれるなど、生活習慣病を中心に医療・保健に関わる幅広い視点からの活用が期待さ

れている。一方、生活習慣病の疾患概念は一様ではなく、分析の立場によって研究対象となる疾患名に大きなズレがあると、分析結果の活用や比較検討においても齟齬をきたす可能性もありうる。

今後も分析者の裾野が広がることが予想される中で、「生活習慣病の ICD-10 上の範囲を NDB としてどう定めて分析するか」標準的な手法の開発が求められている。入院主病名における出現頻度のみで傷病名候補を絞り込むことは当然ながらできない。しかし、出現頻度によるリストがあると、病名リストを専門家の眼を入れて作成する時に、漏れがないか、別の視点がないか、等の参考情報になりうる。引き続き、ICD-10 コードによる選択や、臨床的な関連性による追加など、分析関係者のコンセンサスを得ながら標準傷病名リストを作成することが求められる。一方で、目的に応じて、病名リストの内容が変わってくることに注意しておくべきである。

また、病名のコード指定のあり方について、実際のリスト案を作成するに当たり、以下の如く今後の課題について整理された。

- ・20,000 以上にもおよぶ傷病名の区分は容易ではない。対象疾患群を選定してはどうか。
- ・ロードマップを作成し、ある程度目標を定めて作業をすすめる必要があるのではないか。
- ・保険診療上のルールでは、「傷病名は医師が付与するもの」とされている。コード区分の作成は、ある意味で「医師が付与した（≒オーソライズした）傷病名に対し、解釈を加える」作業である。
- ・そのことの限界については、周知しておく必要があるのではないか。
- ・こうして作成された「区分」は、どのような「場」において活用されるべきか。

#### (4) 二次活用の具体的なあり方についての検討

基本的な利活用の一つとしては、国レベルなどで、診療のパフォーマンスを計測するのに役立つ。その一例として、多府県広域地域の心筋梗塞の性年齢階層別の死亡率の計測の試行ができ、また、同疾患においてレセプト情報における D P C 病院での捕捉割合を出すこともでき、今後の（限定的な D P C データではなく）国全体での D P C データ解析結果の位置づけの参考情報となる。

地域毎の急性期脳梗塞治療を可視化する研究 (Otsubo T, Imanaka Y, et al. J Stroke Cerebrovasc Dis 2015 Jan;24(1):239-51) などを見ると、地域間の顕著な差が見られ、今後の地域医療構想・医療計画に、そして新たな施策の影響をみるのに、役立つ情報が得られることがわかる。また、脳梗塞治療に従事する医師数が多い地域では、院内死亡率が低く、入院医療費が高く、医療連携が活発である地域ほど、入院医療費が低くなる傾向がみられた。さらなる検証が求められるが、今後の医療資源再配備や制度再編において、示唆に富む情報が N D B から得られうるということがわかる。

## E. 結論

レセプト情報等の活用の際には、データの特性と限界を踏まえて分析を行えば、社会的に価値ある多くのアウトプットが算出できる。本データベースの精度に関しては、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る。また、本データベースの活用環境の改善に向けては、生涯通じて一意の個人 ID の実装のほか、マスターテーブルやコードリストなど病名情報のあり方含め、様々な観点からの環境整

備が望まれる。

## G. 研究発表

### 1. 論文、書籍発表

1) 加藤源太、平野景子、赤羽根直樹. レセプト情報・特定健診情報等情報データベースの利活用についてーこれまでの経緯を踏まえてー. 統計. 2014年10月号 pp.8-13.

2) 今中雄一. 脳梗塞 t-PA 治療の経済評価と均てん化施策. 脳と循環 2014. Sep Vol.19, No.3. pp.49-52. .

3) 今中雄一. 混合診療と医療改革ー新制度 広くデータ評価を. 経済教室 日本経済新聞 2014年7月25日.

### 2. 学会発表等

1) 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)

2) 猪飼宏, 今中雄一. DPC, NDB データを用いた医療の質の評価. 第52回日本医療・病院管理学会学術総会: 東京, 2014年9月13日-14日. (抄録: 日本医療・病院管理学会誌 51Supplement : p25, 2014.)

3) 今中雄一. 質の可視化から地域システムの協創へ. 第15回介護保険推進全国サミット in くまもと: 熊本, 2014年10月30日-31日.

4) Imanaka Y. "How to balance the health care needs and the regional system." OECD Reviews of Health Care Quality: JAPAN(2014/11/5).

5) Imanaka Y. "The importance of databases." OECD Reviews of Health Care Quality:

JAPAN(2014/11/5).

## 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 |

## I. 知的財産権の出願・登録状況 (予定を含む。)

- |         |    |
|---------|----|
| 4. 特許取得 | なし |
|---------|----|

## 5. 実用新案登録

なし

## 6. 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

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。

初年度に調査対象としたのは、レセプトデータを既に研究者へ提供している韓国、台湾および米国である。韓国・台湾の調査項目は、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等である。米国は、米 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（オンサイトセンタ）について調査した。

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、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。

次年度は、台湾と同様のランダム抽出データの試行提供がはじまった韓国を対象とした。韓国は、HIRA が HIRA-NPS の開発を行った。HIRA-NPS とは、韓国国内において 1 年間に医療機関を利用した全患者対象(約 4600 万名)を母集団として、性別・年齢(5 歳単位)区間による患者単位の層化系統抽出を行ったデータセットである。初めに作成された 2009 年版のデータセットには、抽出された入院患者の約 70 万人・外来患者約 40 万人の 1 年間のすべての診療内訳と処方内訳が含まれる。開発目的は、1)健康保険に関する診療情報の利活用の増大、2)情報公開を促進させると同時に HIRA のデータ提供に要する