

特集：エビデンスに基づいたヘルスケアシステムのマネジメント—医療政策，医療計画，医療現場の新たな評価分析—

<報告>

レセプト情報・特定健診等情報データベースの概要と活用の方向性 —滋賀県特定健診データ分析事例から—

三浦克之¹⁾，宮川尚子¹⁾，村上義孝²⁾，角野文彦³⁾，岡山明⁴⁾

¹⁾ 滋賀医科大学社会医学講座公衆衛生学部門

²⁾ 滋賀医科大学社会医学講座医療統計学部門

³⁾ 滋賀県健康福祉部

⁴⁾ 結核予防会第一健康相談所

Outlines of the national database on medical expenditure and health check-up

Katuyuki MIURA¹⁾，Naoko MIYAGAWA¹⁾，Yoshitaka MURAKAMI²⁾，
Humihiko KAKUNO³⁾，Akira OKAYAMA⁴⁾

¹⁾ Department of Health Science, Shiga University of Medical Science

²⁾ Department of Medical Statistics, Shiga University of Medical Science

³⁾ Division of Health and Welfare, Shiga Prefecture

⁴⁾ First Institute for Health Promotion and Health Care, Japan Anti-Tuberculosis Association, Tokyo, Japan

抄録

「高齢者の医療の確保に関する法律」の施行により医療費適正化計画を作成することとされ、これと併せ厚生労働省が全国から収集されたレセプト情報および特定健診等情報について調査及び分析を行うことが定められた。この「レセプト情報・特定健診等情報データベース」(以下NDB)は、公益性の高い研究のためであれば個別審査を経た上で国以外の主体にデータを提供することとなった。レセプト情報は審査支払機関における一次審査分データを匿名は処理したものを国が収集、また、特定健診等情報は各保険者が所定の匿名化処理が行ったものを社会保険審査支払基金が収集し、国の保有するサーバに格納したものである。NDBデータ提供は「レセプト情報・特定健診等情報データベースの提供に関するガイドライン」に基づき、有識者会議の審査を経て実施される。データは個人情報に準じた取り扱いが行われ、「疫学研究に関する倫理指針」の遵守が必要である。成果物の公表にあたっては、特定個人・特定医療機関等の第三者によって識別される可能性をなくすため、いくつかの措置を講ずることが求められている。また、探索的研究にも道を開く匿名性が高く操作性に優れた基本データセットとして「サンプリングデータセット」も準備されている。著者らは特定健診データの模擬的データ分析のため、滋賀県特定健診データの提供を受けたが、提供データの形式変換作業およびデータクリーニングが必要であった。しかしながら、これまでになく地域住民全体の悉皆的な健康状態の把握と分析が、NDBを用いて初めて可能となった。本データベースが公益性の高い研究に広く活用されることより、国民の健康保持、疾病予防、疾病モニタリング、医療評価等、広い範囲で役

連絡先：三浦克之

〒520-2192 滋賀県大津市瀬田月輪町

Seta Tsukinowa-cho, Otsu City, Shiga, 520-2192, Japan.

Tel: 077-548-2191

Fax: 077-543-9732

E-mail: miura@belle.shiga-med.ac.jp

[平成25年1月27日受理]

立つものと期待される。

キーワード：レセプト情報, 特定健診, ナショナルデータベース

Abstract

The Japanese Government has begun to collect all data on medical expenditure and national health check-up and to make the national database for making the national medical expenditure plan. This database can be also used for researches with highly public purposes, after permission by the examination committee. The data are to be handled as private information in order not to be distinguished as a certain person. The sampling dataset which was made from the total database is also prepared. The health status of a certain population can be assessed using this database of national health check-up. The database will be very useful for public health practice, disease prevention, disease monitoring, assessment of medical practice, and so on in Japan.

keywords: Medical expenditure information, national health check-up, national database
(accepted for publication, 27th January 2013)

I. はじめに

平成18年度の医療制度改革において、「老人保健法」が改正され、「高齢者の医療の確保に関する法律」(以下「高確法」)が平成20年より施行された。その際、厚生労働大臣及び都道府県は五年ごとに全国医療費適正化計画、都道府県医療費適正化計画をそれぞれ作成することとされ、これと併せ、高確法第16条に基づき、厚生労働省が全国から収集されたレセプト情報および特定健診等情報について調査及び分析を行うことが定められた。その後、この「レセプト情報・特定健診等情報データベース」(いわゆる「ナショナルデータベース」, 以下NDB)に収集されたデータを国以外の主体が活用することについて検討が行われ、公益性の高い研究のためであれば、研究目的等について個別

審査を経た上でデータを提供する方向となった(図1) [1]。こうした経緯のもと、本来の目的外でのNDBの利用申請があった際に、データの公益性等について意見を述べることを目的とした「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)が平成22年に設立された。

著者らは本有識者会議の委員としてNDB提供に関するガイドライン作成に関わり、また、NDB利用の模擬的検討として滋賀県の特定健診データ利用を試みた経験から、NDBの概要と今後の活用の可能性について報告する。

II. レセプト情報・特定健診等情報データベース(NDB)とは

平成18年度医療制度改革において、レセプト情報のオンライン・電子媒体での請求が推進されることとなり、平成

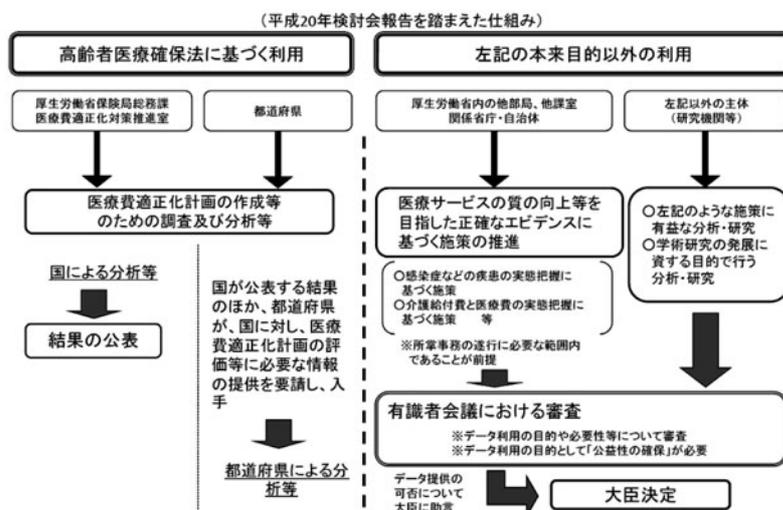


図1 レセプト情報・特定健診等情報データベースの利用

21年度から電子化されたレセプトデータが、NDBに収集されている。レセプト情報は、審査支払機関における一次審査分データに対し、所定の匿名化処理が行われたものを、国が収集し保有するサーバに格納したものである。また、特定健診等情報は、各保険者が所定の匿名化処理を行ったものを、社会保険審査支払基金が収集し、国の保有するサーバに格納したものである。平成22年8月末時点で、レセプト情報は約15億9,800万件、特定健診・保健指導情報は約2,065万件が格納されており、平成24年11月末時点で、レセプト情報は約50億件、特定健診等情報は約6,600万件が格納されている（図2）。

レセプトとは、保険診療を行った医療機関が患者一人一人の診療報酬（医療費）を、審査支払機関を経由して保険者に請求を行う際の明細書である。レセプトにおいて傷病名、診療開始日及び診療実日数等は記載事項とされているが、NDBにあるレセプト情報では、患者の氏名及び生年月日の「日」等の情報は削除されている。平成22年7月診療分で全レセプト件数のうち約8割、平成24年9月診療分で全レセプト件数のうち91%が電子化され、国が保有するサーバに格納されている。格納されていないデータは、レセプト電子化が進んでいない歯科に関するものが多い。

特定健診等情報には、特定健診の受診情報（実施日等）、保険者番号、特定健診機関情報及び受診者情報の一部に加え、健診結果・問診結果、保健指導レベル、支援形態及び特定保健指導のポイント数等の情報が記載されている。特定健診・特定保健指導機関に関する情報及び受診者の氏名等の情報は削除され、国が保有するサーバに格納されている。

複数のデータベースにおいて同一人物であることを特定できるように、固定長の擬似乱数（ハッシュ値）を付与することができるハッシュ関数を用い、レセプト情報及び特定健診等情報に、それぞれの情報から生成したハッシュ値を付与している。具体的には、レセプト情報・特定健診等情報には、①保険者番号・記号番号・生年月日・性別から生

成したハッシュ値及び②氏名・生年月日・性別から生成したハッシュ値、の2つが付与されている。転職等で保険者番号、記号・番号が変更になった場合は②のハッシュ値により、氏名の変更等がみられた場合には①のハッシュ値により突合が可能であるが、記号・番号と氏名がともに変更となった場合には、いずれのハッシュ値においても突合は不可能である。

III. データ提供の手順

NDBデータの提供に関しては、平成23年度及び平成24年度は試行期間として位置づけられ、平成23年3月に整備された「レセプト情報・特定健診等情報データベースの提供に関するガイドライン」（以下、ガイドライン）に基づき実施された。試行期間の終了後、平成25年度以降についても試行期間のレセプト情報等の提供体制を継続することが望ましいと考え、現在のガイドラインを適宜見直しながら通常運用として実施する予定となっている。データ提供は、提供依頼申出者からの申出ごとに有識者会議においてデータ提供の可否及び提供する場合の適切な提供方法について審査が行われる。

現在、NDBに格納されている患者個々のレセプト情報は、氏名等、個人を特定できる情報は匿名化（ハッシュ関数による変換）されているため、通常、単独では特定の個人を特定できず、個人情報には当たらないと考えられる。しかし、他の公知の情報などと照らし合わせることにより個人情報に当たりうるケースもある、個人立の医療機関等の機関コードは他の情報と照合することにより経営者個人の情報を識別できる、などの理由により個人情報に準じた取扱いを行っている。したがって、医療機関等が個人情報を取扱う場合等に適用される「医療情報システムの安全管理に関するガイドライン」に準じた措置を提供依頼申出者に求めており、ハード面、ソフト面での高いセキュリティ

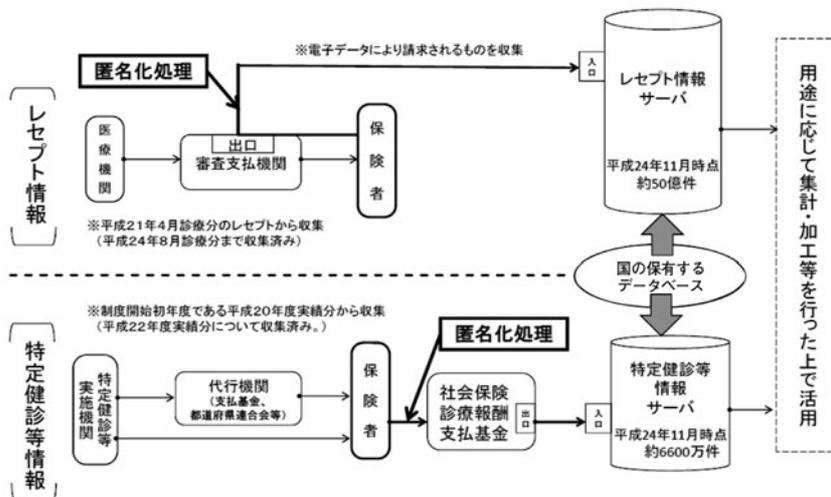


図2 レセプト情報・特定健診等情報の収集経路

要件が求められる。

一方、NDBデータはハッシュ関数による匿名化がなされているものの連結不可能匿名化とまでは言えないことから、「疫学研究に関する倫理指針」を遵守した上で活用されるべきと考えられている。

提供依頼申出者は、手数料制度及び情報漏洩等に対する法的根拠のある罰則がないこと、また、可能な限り効率的に公益性の高い研究に情報提供を行う必要があることを鑑み、提供依頼申出者を一定の範囲に限定している。公的補助金（厚労科研費等）を受けている場合を除き、営利企業は対象としないこととし、①国の行政機関、②都道府県、③研究開発独立行政法人、④大学、⑤医療保険者の中央団体、⑥医療サービスの質の向上等をその設立目的の趣旨に含む国所管の公益法人及び⑦提供されるデータを用いた研究の実施に要する費用の全部又は一部を国の行政機関から補助されている者、としており、レセプト情報等データを活用して公益性の高い研究を行う者はおおむね申出できる環境となっている。

IV. 成果物の公表

成果物の公表にあたっては、特定個人・特定医療機関等の第三者によって識別される可能性をなくすため、いくつかの措置を講ずることが求められている。

一つは最小集計単位の原則である。米国における情報提供の例も踏まえ、原則として、公表される成果物においては原則として患者等の数が10未満になる集計単位が含まれてはならないとしている。また、集計単位が市区町村（政令指定都市の場合の行政区を含む）の場合には、公表される成果物において患者等の数が100未満になる集計単位が含まれてはならないこととしている。医療機関や保険者等の個別機関については、原則として属性情報による集計によって対象となる機関が3未満に特定されてはならないとしている。

年齢区分については、原則として、5歳毎にグルーピングして集計し、85歳以上については同一のグルーピングとすることとしている。

地域区分に関しては、特定健診等情報には患者の住所地の記載があるが、原則として公表される成果物における最も狭い地域区分の集計単位は2次医療圏又は市区町村とするとしている。医療機関・保険者の所在地を集計単位に用いている場合は、原則として、公表される成果物における最も狭い地域区分の集計単位は2次医療圏又は市区町村とし、市区町村で集計した場合には保険者の特定を避けるため保険者種別でのクロス集計を公表することは、保険者の同意を得ている場合等を除き、原則認めないとしている。

V. サンプルングデータセット

探索的研究に対する研究者のニーズが高く、探索的研究にも道を開く匿名性が高く操作性に優れた基本データセッ

トを整備していく必要があると考えられたため、単月分のサンプルデータが提供依頼申出者向けに整備された。このデータは「サンプルングデータセット」と名付けられ、「医科入院」、「医科入院外」、「DPC」及び「調剤」の4種類のレセプトに対して整備が行われた。データセットはそれぞれ一定の割合で抽出され、傷病名コード等において出現頻度が少ない情報に対しては匿名化処理が施された。加えて、個人の特定可能性を下げるため、保険者番号や都道府県情報等は削除されている。

こうして整備されたサンプルングデータセットは、抽出条件の詳細な指定は求めず、研究の概略が把握できるような提供依頼申出であれば承諾することとされている。一方、セキュリティ要件は、匿名性が高められたデータとはいえ個票データであることから、基本的には通常の中出と同様のセキュリティを確保することを求めることとされた。

VI. 滋賀県特定健診データ利用の経験

筆者らはNDBの特定健診データ活用について検討するため、滋賀県の平成20年度特定健診受診者全員のデータ利用についての模擬申し出を行い、データ分析を試みたので、その利用経験について記述する。

地域の生活習慣病危険因子の現状把握においては、従来、老人保健法による住民健診データを使用することが多かったが、特定健診導入後、自治体実施の健診が国保加入者対象となったため都道府県単位の偏りのない現状把握が困難となった。一方、NDBの特定健診データでは、組合健保、協会健保を含む全国民の健診データを用いて、より偏りの少ない都道府県単位の現状分析が可能となった。そこで、特定健診NDBにおける滋賀県民のデータを用いて、滋賀県民の生活習慣病危険因子の現状を明らかにすることを目的として本利用を行った。特に、生活習慣病予防対策立案のためのエビデンスとするため、高血圧、脂質異常、糖尿病、メタボリックシンドローム等の有病率・治療率・コントロール率を、性・年齢階級別、二次医療圏別、保険者種類別（国保・組合健保・協会健保等）に分析することとした。

データ提供を依頼した詳細抽出条件は、【期間】平成20年度、【種類】特定健診データ、【抽出条件】滋賀県在住者全員のデータ、【必要な項目】受診者情報（性別、年齢、在住市町）、保険者の種類（市町村国保、協会健保、船員健保、国保組合、共済組合、組合健保）、健診結果、質問票情報、【匿名化が必要な項目】患者の生年月は受診時年齢（5歳年齢階級）に換算、在住市町は郵便番号から市町村コードに変換、として、約20万人のデータの提供を受けた。

データはCSV形式となっており、性別、年齢、保険者等が格納された「基本情報ファイル」、特定健康診査の基本的な検査項目が格納された「健診結果・問診情報ファイル」、詳細な健診項目が格納された「詳細情報ファイル」に分けて提供された。各ファイルにおいて、各個人のデータは項目数と同じ数の複数行に渡って配列されている。図3に提供データの配列例を示す。健診項目を身長、体重、

【提供する際のデータサンプル】

基本情報ファイル		
2011041115061_1h000100162011013000001000000001,00000000000000,0000000001,,00001,1,060,,,,,1,		←健診者A
2011041115061_1h000100162011013000002000000002,00000000000000,0000000002,,00002,1,055,,,,,1,		←健診者B
2011041115061_1h000100162011013000003000000003,00000000000000,0000000003,,00003,2,055,,,,,1,		←健診者C
健診結果・問診結果情報ファイル		
2011041115061_1h000100162011013000001000000001,00000010000001,,9N001000000000001,162.3,cm,,,,,,,,	←身長 ←体重 ←BMI ←腹囲	A の 個 票 情 報
2011041115061_1h000100162011013000001000000002,00000010000002,,9N006000000000001,65.5,kg,,,,,,,,		
2011041115061_1h000100162011013000001000000003,00000010000003,,9N011000000000001,25,kg/m2,,,,,,,,		
2011041115061_1h000100162011013000001000000004,00000010000004,,9N021000000000001,70,cm,,,,,,,,		
2011041115061_1h000100162011013000002000000002,00000010000001,,9N001000000000001,171.2,cm,,,,,,,,	←身長 ←体重 ←BMI ←腹囲	B の 個 票 情 報
2011041115061_1h000100162011013000002000000002,00000010000002,,9N006000000000001,67.5,kg,,,,,,,,		
2011041115061_1h000100162011013000002000000002,00000010000003,,9N011000000000001,23,kg/m2,,,,,,,,		
2011041115061_1h000100162011013000002000000002,00000010000004,,9N016160300000001,67,cm,,,,,,,,		
2011041115061_1h000100162011013000003000000003,00000010000001,002,9N001000000000001,165.3,cm,,,,,,,,	←身長 ←体重 ←BMI ←腹囲	C の 個 票 情 報
2011041115061_1h000100162011013000003000000003,00000010000002,002,9N006000000000001,62.5,kg,,,,,,,,		
2011041115061_1h000100162011013000003000000003,00000010000003,002,9N011000000000001,21,kg/m2,,,,,,,,		
2011041115061_1h000100162011013000003000000003,00000010000004,002,9N021000000000001,65,cm,,,,,,,,		

図3 提供される特定健診データのデータ形式(例)

BMI, 腹囲の4項目とすると、図のような配列になっている。

「基本情報」の「ID1」が個人に特有のID(ハッシュ関数)、「通番2」はレコード固有のIDで、ほぼ個人ID(まれに同年度に同じ人が2回受診)であり、3つのファイルの個人データを繋げるキーは「通番2」であった。

そこで一個人のデータが1行に並ぶように、CSVファイルの各行に付与されている「通番2」をキーにして結合する必要があったが、そのためにVisual Fox Proを用いたプログラムを作成し、滋賀医科大学のコンピューターにて結合作業を行った。

特定健診データは、保険者が提出する際、必ずしも形式が整っていなかったり、入力ミスと思われる外れ値があったりしたため、データクリーニングが必要であった。クリーニングは厚労省保険局「標準的な健診保健指導プログラム(確定版)」が定めたデータ規定に従って実施した。すなわち測定値の範囲チェックによる外れ値の修正、また、ある検査項目について複数の測定方法や複数回の測定値がある場合の測定値の一元化などの作業が必要であった。クリーニング作業にはSASを用いたが、518変数、約20万件の巨大なデータであったため、変換など簡単な作業でもコンピューターによる処理にかなり時間を要した。

以上の作業を経て、計画した分析を実施した。分析結果については別途論文投稿を予定しているが、わが国ではこれまでにない地域住民全体の健診データの実態を知ることが可能であった。滋賀県では特定健診受診率が約40%であり選択バイアスの存在の可能性は否定できないものの、県民の健康状態把握を目的としたデータベースとしては破格の規模であり、今後の有効活用が期待される。

VII. おわりに

平成23年度から試行期間として目的外使用のためのデータ提供が始まった「レセプト情報・特定健診等情報データベース」について、その内容、提供のガイドラインと手順などについて解説した。また、模擬申し出によってデータ提供を受けて分析した滋賀県特定健診データの利用経験について報告した。

「高齢者の医療の確保に関する法律」により健診実施が保険者に義務づけられるという大きな変化に伴って、全ての保険者が含まれる包括的な健診データの収集が可能になった点は、わが国において、あるいは国際的に見ても画期的なことである。また、このことはIT技術の進歩と普及によって初めて可能となったもので、国民の健康状態の把握は新しいステージに入ったと言える。このような変化は世界の一部の国々でもすでに認められている。本データベースは、国による行政施策の評価のみでなく、公益性の高い研究に広く活用されることより国民の健康保持、疾病予防、疾病モニタリング、医療評価等、広い範囲で役立つものと考えられる。

文献

[1] 厚生労働省保険局, レセプト情報等の提供に関する有識者会議. レセプト情報・特定健診等情報データの第三者提供の在り方に関する報告書. 2013.