

<原著>

NDB 解析用データセットテーブルの開発

福田治久¹⁾, 佐藤大介²⁾, 白岩健²⁾, 福田敬²⁾¹⁾九州大学大学院医学研究院医療経営学分野²⁾国立保健医療科学院保健医療経済評価センター

The Development of Dataset Tables for NDB Analyses

Haruhisa Fukuda¹⁾, Daisuke Sato²⁾, Takeru Shiroyiwa²⁾, Takashi Fukuda²⁾¹⁾Kyushu University Graduate School of Medical Sciences²⁾Center of Outcomes Research and Economics Evaluation for Health, National Institute of Public Health

抄録

目的: 2011年度より第三者提供が開始されたレセプト情報・特定健診等情報データベース (NDB) の研究利用が不十分な状況にある。学術研究を加速化させ、エビデンスに基づいた医療政策を推進するためには、NDBの活用可能性を高めていく必要がある。本研究の目的は、臨床疫学研究および医療経済研究を行うのに有用性が高く、かつ、データ容量の効率性が高いNDB解析用データセットを構築することである。

方法: 2009年4月から2016年12月の間の医科入院レセプトおよびDPCレセプトにおいて1度でも出現したことの解析用患者IDを全データから無作為に25%分を抽出し、当該解析用患者IDの全期間における全診療行為情報を含む全レセプトデータを格納したNDBを用いた。臨床疫学研究および医療経済研究を行うのに有用性の高い解析用データセットテーブルに必要な変数について検討した。また、医科レセプトにおいては退院年月日情報が含まれていないことから、レセプトデータを用いた補完的な退院年月日を付加する方法について検討した。本検討では、退院年月日情報が含まれるDPCレセプトを用いて、入院年月日と診療実日数を用いる場合と、診療行為発生日を用いる場合のそれぞれの方法で退院年月日を算出し、実際のDPCレセプトに記載されている退院年月日との一致状況について検証した。

結果: NDBに含まれるレコード識別情報を有機的に連結させた解析用データセットテーブルとして、以下の11テーブルを開発した: 患者 (KAN), レセプト (REC), 傷病名 (SYO), 診療行為 (SIN), 薬剤 (IYA), 特定器材 (TOK), 調剤 (TYO), 調剤加算料等 (TKA), DPC診断群分類 (BUD), 医療機関 (IRK), 入院 (ADM). 医療機関 (IRK) を除く各テーブルは解析用患者IDによって相互に突合することができる。また、医科レセプトにおける補完退院年月日は、診療行為 (SI), 医薬品 (IY), 特定器材 (TO), コーディングデータ (CD) の各レコードにおける診療行為年月日の最終日情報を用いることで、99.83%の入院症例において正しい退院年月日を付加することができた。

結論: 本研究において開発した解析用データセットテーブルを用いることで、NDBを用いた臨床疫学研究および医療経済研究を即座に実施可能な環境をもたらすことができる。

キーワード: NDB, 解析用データセットテーブル, 臨床疫学研究, 医療経済研究

Abstract

Objectives: Although the National Database of Health Insurance Claims and Specific Health Checkups of

連絡先: 福田治久

〒812-8582 福岡県福岡市東区馬出3-1-1 基礎研究B棟132

3-1-1, Maidashi, Higashi-ku, Fukuoka City, 812-8582 Japan.

Tel: 092-642-6956

E-mail: h_fukuda@hcam.med.kyushu-u.ac.jp

[平成31年3月13日受理]

Japan (NDB) was approved for third-party use in 2011, it remains underused for research purposes. There is therefore a need to improve the NDB's usability to facilitate its applications in academic research and support its contributions to evidence-based health policy. This study aimed to construct NDB analytical datasets with high storage efficiency and specific applications for research in clinical epidemiology and health economics.

Methods: The NDB in this study comprises a random sample of 25% of all patients with at least one claims record in medical claims data or diagnosis procedure combination (DPC) claims data between April 2009 and December 2016. This database stores claims data for all medical services provided to these patients throughout the study period. We investigated the variables needed in analytical dataset tables that would facilitate research in clinical epidemiology and health economics. As medical claims data do not include information on discharge dates, we also examined methods for supplementing this information using other claims data. We first calculated the possible discharge dates using admission dates and the number of actual treatment days or dates of recorded treatments; these estimated discharge dates were then compared with the actual discharge dates obtained from the DPC claims data.

Results: We developed the following 11 analytical dataset tables that were organically linked using patient identification codes available in the NDB: Patient information (KAN), claims information (REC), disease information (SYO), medical services information (SIN), pharmaceutical information (IYA), device information (TOK), drug dispensing information (TYO), drug dispensing incentives information (TKA), DPC diagnostic group classifications (BUD), medical institution information (IRK), and hospital admissions information (ADM). With the exception of the IRK table, all tables could be mutually compared using patient identification codes. In the analysis of supplementary discharge dates for the medical claims data, estimates using the final date of recorded treatments (available in the SI, IY, TO, and CD files) were able to accurately identify the correct discharge date in 99.83% of all hospitalized cases.

Conclusion: The use of the analytical dataset tables developed in this study may help to establish an environment that facilitates the rapid initiation of research in clinical epidemiology and health economics using the NDB.

keywords: NDB, analytical dataset tables, clinical epidemiological research, health economics research

(accepted for publication, 13th March 2019)

I. 緒言

平成23年度より、医療費適正化計画策定に資する目的以外でのレセプト情報・特定健診等情報データベース(NDB)の第三者提供が開始され、平成30年2月までに167研究に対してデータ提供が承諾されている[1]。NDBは、ほぼ全国民の保険診療データが網羅されているという悉皆性と患者単位での追跡可能性を有することから臨床疫学や医療経済学の学術利用のみならずエビデンスに基づいた医療政策の立案に資するかつてないデータベースである。しかし、これまでのところ、学術論文としての報告は19研究28件に過ぎず[1-2]、NDBを活用した研究成果の産出は十分であるとは言い難い状況にある。学術研究を加速化させ、エビデンスに基づいた医療政策を推進するためには、NDBの活用可能性を高めていく必要がある。

NDBは、医科、DPC、調剤、歯科のレセプトデータが含まれるが各データにはレコード識別情報があり、各識別情報に詳細な項目が格納される階層的なデータ構造となっている。NDB申出者は各々のリサーチ・クエリを解決するために、これらデータ構造を有機的に

統合させた解析用データセットを各々で構築し、その上で、必要な解析を行う必要がある。しかしながら、解析用データセットの構築にはレセプトデータに関する専門知識と高度なデータ処理力が不可欠であるため、多くのNDB申出者はこのデータセット構築の段階に多大な時間を投じていると思われる。我々は、この問題を縮小すべく、NDBを用いた臨床疫学研究および医療経済研究を行うのに適した解析用データセットについて検討することとした。

II. 方法

1. データベース

本研究で使用したデータは、レセプト情報等の提供に関する申出により抽出されたNDBである(「費用対効果評価に活用するためのナショナルデータベースを用いた費用データ基盤整備」(申出者:厚生労働省・金光一瑛))。本研究で使用したNDBには医科、DPC、調剤の各請求情報が含まれており、医療機関情報、レセプト情報、摘要情報の請求情報が含まれている。各請求情報には情報の内容に応じて、各種のレコード識別情報から

構成されているが, レコード識別情報の詳細は仕様書を参照されたい [3-5]. 本研究では, 平成21年4月から平成28年12月における医科入院およびDPCレセプトから1度でも出現したことのある解析用患者IDから無作為に25%を抽出し, 当該解析用患者IDの全期間における全診療行為情報を含む全レセプトデータを格納したデータベースを用いた. なお, ここでいう解析用患者IDとは, 後述するように, NDBにおいて提供されるハッシュ値1 (保険者番号, 被保険者証の記号・番号, 生年月日, 性別から生成) [6]あるいはハッシュ値2 (氏名, 生年月日, 性別から生成) [6]から, いずれかが一致しているものを突合せたIDを指している. 調剤レセプトの連結に際しては, レセプトデータを構成するレセプト共通 (RE) レコードのハッシュ値1と, 都道府県コード, 医療機関コード, 調剤年月 (当月と前月) をキーとして連結した.

2. 解析用データセットテーブル作成に向けた変数置換

NDBを用いた臨床疫学研究および医療経済研究を行うのに有用性が高く, かつ, データ容量の効率性が高いデータセットを構築するために, 我々はNDBに含まれるレコード識別情報を有機的に連結させた解析用データセットテーブルを検討した. これによりレセプトデータに重複する情報を削除し, 解析に適した情報のみを抽出した. 解析用データセットテーブルの作成に際して, 以下の変数を作成した.

1) ハッシュ値1およびハッシュ値2のShort ID (SID) への置換

分析の実施可能性を高めるために, NDBに当初より提供される医科およびDPCレセプトにおけるREレコードのハッシュ値1とハッシュ値2の各64バイトを抽出し, それぞれをソートして重複を排除し, 9桁の連番を振ったものをShort ID 1 (SID1), Short ID 2 (SID2) とし, REレコードのハッシュ値1とハッシュ値2に置き換えた.

2) 日付情報の西暦変換

レセプトデータにおける日付情報はGYMMの形式であり, Gは元号表記 (1: 明治, 2: 大正, 3: 昭和, 4: 平成) となっていることから, 医科レセプトにおけるREレコード (診療年月, 入院年月日), 傷病名 (SY) レコード (診療開始日), DPCレセプトにおけるREレコード (診療年月, 入院年月日), SYレコード (診療開始日), 診断群分類 (BU) レコード (今回入院年月日, 今回退院年月日), コーディングデータ (CD) レコード (実施年月日), 調剤レセプトにおけるREレコード (調剤年月), 調剤 (CZ) レコード (処方年月日, 調剤年月日), 基本料・薬学管理料 (KI) レコード (前回調剤年月日) に対して西暦表記に変換した.

3. 解析用データセットテーブル作成に向けた変数付加

1) 解析用患者IDの付加

NDBでは, 保険者番号, 被保険者証の記号・番号, 生年月日, 性別に基づいて付加されるハッシュ値1と氏名, 生年月日, 性別に基づいて付加されるハッシュ値2が提供される[6]. 前者は主に保険者の変更によりハッシュ値が変更され, 後者は主に氏名の変更や入力間違いによりハッシュ値が変更されるという特徴を有している. 一方, ハッシュ値1とハッシュ値2が同時に変更されることはさほど多くないと考えられる. そこで, 本研究では, ハッシュ値1を短縮化したSID1あるいはハッシュ値2を短縮化したSID2の何れかが一致している場合は同一患者と想定することとし, 当該患者の識別用に解析用患者IDを付加した. なお, SID1とSID2が同時に変更されたため, 同一患者と想定できない場合は, 本来は同一患者であるにもかかわらず変更後は別患者として分類されることとなる.

2) DPCレセプトへの「レセプト総括区分」の付加

DPCレセプトのREレコードには「レセプト総括区分」が記載されているものの, REレコード以外のレコードには当該情報が付加されていない. そのため, DPCレセプト, 総括レセプト, 総括対象DPCレセプト, 総括対象医科入院レセプトのいずれに属するものかを識別できるように, 「レセプト通番」, 「レコード順序」でソートして直前のREレコードの「レセプト総括区分」をREレコード以外の各レコードに付加した.

3) DPCレセプトの重複データへの「重複フラグ」の付加

DPCレセプトにおいて包括評価部分に係る診療行為の内容が分かる情報を出来高部分に係る診療行為の情報も含めCDレコードに記録することとなっている一方, 出来高情報はSIレコード, IYレコード, TOレコードに記載することになっている[4]. したがって, 出来高の診療内容をSIレコード, IYレコード, TOレコードから, 包括分の診療内容をCDレコードから集計する場合, 診療内容に重複が発生する. そのため, 本研究では, 重複データに対して重複フラグを付与した. 重複の判定方法は, 2012年3月以前は, レセプトID, レセプト電算処理コード, 診療識別が一致することを, 2012年4月以降は, レセプトID, レセプト電算処理コード, 診療識別, 実施年月日が一致することを条件とした. なお, 重複レコードを削除せずに, 重複レコードを残した上で重複フラグを付加した理由は, 重複の判定は必ずしも完全な条件である保障がなく, 本研究において適切と考えられる条件として設定しているためであり, 必要に応じて原データを参照できるようにするためである.

4) DPCレセプトのCDレコードおよび調剤レセプトのIYレコードとTOレコードへの「補完後点数」の付加

DPCレセプトのCDレコード, 調剤レセプトのIYレコード, TOレコードには明細毎の点数が記載されていないことから, レセプト電算処理システムマスター [7]を参照して「補完後点数」を付加した.

5) 医科レセプトへの「補完退院年月日」の付加

DPCレセプトにおいては診断群分類 (BU) レコード

において「今回退院年月日」データが含まれているものの、医科レセプトにおいては退院年月日を把握可能なデータが含まれていない。医科レセプトにおいて退院年月日を推測する方法として、REレコードにおける「入院年月日」データに保険者（HO）レコードにおける「診療実日数」を加算する方法（方法1）と、SIレコード、IYレコード、TOレコードにおける最終行為発生日を用いる方法（方法2）が考えられる。

「補完退院年月日」としていずれの方法に基づくデータを付加すべきかを検討するために、我々は、DPCレセプトが出現し、包括期間内のみで入院した2012年4月から2016年12月の期間における入院症例を対象に、方法1で作成した「補完退院年月日1」と方法2で作成した「補完退院年月日2」を、BUレコードから参照した「今回退院年月日」との一致度を検証し、一致度の高い方法を用いて「補完退院年月日」データを付加した。一致度の指標としては κ 統計量を使用した。なお、2012年3月以前は、SIレコード、IYレコード、TOレコードに実施年月日情報が含まれていないことから、補完退院年月日は欠測値とした。

6) エピソード別入院IDの付加

入院エピソードは、複数月にまたがることや、DPCレセプトから医科レセプトに移行することや、医科レセプトからDPCレセプトに移行することがあるため、複数のレセプトデータを突合させる必要がある。一方で、入院エピソードに関する分析は臨床疫学的にも医療経済評価的にも実施頻度が高い。そのため、我々は、入院レコードを一意に識別可能な「解析用患者ID+エピソード別入院年月日」に基づくエピソード別入院IDを作成し、REC、SYO、SIN、IYA、TOK、BUD、ADMに付加した。

このようなエピソード別入院年月日を判定するためには、REレコードに記載されている「入院年月日」データを参照する必要があるが、DPCレセプトにおいては、総括区分0および2に対しては「入院年月日」データが記載されておらず、BUレコードに「DPC今回入院年月日」データが記載されている。このとき、REレコードにおける「入院年月日」データとBUレコードにおける「DPC今回入院年月日」データを用いて入院IDを付与した場合、以下のようなケースにおいて正しくエピソード別入院IDを付与できない問題が生じる。それは、患者Aが2017年7月27日に、DPC対象病院であるものの非DPC算定病棟に入院し、DPC算定病棟に転棟し、再

度、非DPC算定病棟に転棟した場合などである。これらは実際には全て同一の入院エピソードであるものの、レセプトデータ上は異なる入院IDが2つ作成され、3回目の病棟における入院年月日には2回目の病棟における入院年月日よりも以前の日付が入ってしまう（表1）。こうした問題を回避し、適切にエピソード別入院IDを「A-20170727」と付与するために、以下の処理を行った：DPCレセプトの総括区分が0と2のレコードが出現した場合には、診療行為最終日とDPC今回入院年月日の連続性を確認し、日付が連続している場合には一連の入院エピソードとみなした。なお、診療行為最終日には「補完退院年月日2」を使用した。しかしながら、2012年3月以前は、実施日情報が欠落しているため、診療行為最終日には「補完退院年月日1」を使用した。

3. 結果

1) 解析用データセットテーブル

我々は、NDBに含まれるレコード識別情報を有機的に連結させた解析用データセットテーブルとして、表2に示す11テーブルを開発し、本テーブル用に新たなテーブル名称（略称）を以下のように付与した：患者（KAN）、レセプト（REC）、傷病名（SYO）、診療行為（SIN）、薬剤（IYA）、特定器材（TOK）、調剤（TYO）、調剤加算料等（TKA）、DPC診断群分類（BUD）、医療機関（IRK）、入院（ADM）。

IRKテーブルを除く各テーブルは解析用患者IDによって突合できる。また、RECテーブル、SYOテーブル、SINテーブル、IYAテーブル、TOKテーブル、BUDテーブルにおける入院IDを用いることで、ADMテーブルに突合できる。また、IRKテーブルはRECテーブルおよびBUDテーブルにおける医療機関IDによって突合できる。各テーブルの実体関連図を図1に示す。

11テーブルにおける各変数の定義方法をSupplement Table 1 ~ 11 (<https://www.niph.go.jp/journal/data/68-2/appendix.pdf>) に示している。

2) 補完退院年月日の設定

表3に、REレコードの「入院年月日」データとHOレコードの「診療実日数」データから作成した「補完退院年月日1」とSI、IY、TO、CDの各レコードにおける診療行為年月日から作成した「補完退院年月日2」のそれぞれについて、BUレコードから参照した「今回退院年月日」との一致状況を示している。

表1 エピソード別入院IDの付与方法

解析用患者ID	診療年月	レセプト区分	レセプト総括区分	診療行為最終日	DPC今回入院年月日(BU)	DPC今回退院年月日(BU)	入院年月日(RE)	医療機関ID	入院ID	エピソード別入院ID
A	201707	医科		20170731			20170727	a	A-20170727	A-20170727
A	201708	DPC	2	20170827	20170801	20170827		a	A-20170801	A-20170727
A	201708	DPC	3	20170831			20170727	a	A-20170727	A-20170727
A	201709	医科		20170930			20170727	a	A-20170727	A-20170727

表2 解析用データセットテーブルの概要

テーブル名	変数
患者 (KAN)	解析用患者ID, 男女区分, 初出診療年月, 初出診療時年齢区分, 最終診療年月, 最終診療時年齢区分
レセプト (REC)	レセプトID, 解析用患者ID, 患者SID1, 患者SID2, 男女区分, 診療(調剤)時年齢区分, 診療(調剤)年月, レセプト区分, NDBレコード順序, NDBレセプト通番, レセプト総括区分, 入院フラグ, 後期高齢フラグ, 保険者番号(匿名化後), レセプト種別, 診療実日数(処方せん受付回数), 合計点数, 給付割合, 支払機関, 支払額区分, RE入院年月日, BU入院年月日, レセプト連結入院年月日, エピソード別入院年月日, 死亡フラグ, 医療機関ID, 都道府県コード, 病床数区分, エピソード別入院ID
傷病名 (SYO)	傷病名ID, レセプトID, 解析用患者ID, 患者SID1, 患者SID2, 男女区分, 診療時年齢区分, 診療年月, レセプト区分, NDBレコード順序, NDBレセプト通番, レセプト総括区分, 診療開始日, 傷病名コード, 傷病名基本名称, 病名管理番号, 採択区分, 病名交換用コード, ICD-10-1, ICD-10-2, 転帰区分, 傷病名区分, 死因, 疑いフラグ, 主傷病フラグ, エピソード別入院ID
診療行為 (SIN)	診療行為ID, レセプトID, 解析用患者ID, 患者SID1, 患者SID2, 男女区分, 診療時年齢区分, 診療年月, レセプト区分, NDBレコード順序, NDBレセプト通番, レセプト総括区分, 診療識別, 負担区分, 診療行為コード, 数量データ, レセプト記載点数, レセプト記載回数, 一連番号, 一連順序, 補完後点数, 補完後回数, 実施年月日, 回数, 省略漢字名称, 医科診療行為(大分類), コード表用区分番号, コード表用枝番, コード表用項番, CD由来フラグ, CD重複フラグ, エピソード別入院ID
医薬品 (IYA)	薬剤ID, レセプトID, 解析用患者ID, 患者SID1, 患者SID2, 男女区分, 診療(調剤)時年齢区分, 診療(調剤)年月, レセプト区分, NDBレコード順序, NDBレセプト通番, レセプト総括区分, 診療識別, 負担区分, 医薬品コード, 使用量, レセプト記載点数, レセプト記載回数, 一連番号/処方番号, 一連順序/処方内番号, 補完後点数, 補完後回数, 実施年月日, 回数, 医薬品名・規格名, 单位名称, 薬効分類, CD由来フラグ, CD重複フラグ, エピソード別入院ID
特材 (TOK)	特定器材ID, レセプトID, 解析用患者ID, 患者SID1, 患者SID2, 男女区分, 診療(調剤)時年齢区分, 診療(調剤)年月, レセプト区分, NDBレコード順序, NDBレセプト通番, レセプト総括区分, 診療識別, 負担区分, 特定器材コード, 使用量, レセプト記載点数, レセプト記載回数, 一連番号/処方番号, 一連順序/処方内番号, 補完後点数, 補完後回数, 実施年月日, 回数, 名称, 单位名称, 単価, CD由来フラグ, CD重複フラグ, エピソード別入院ID
調剤 (TYO)	調剤ID, レセプトID, 解析用患者ID, 患者SID1, 患者SID2, 男女区分, 調剤時年齢区分, 調剤年月, レセプト区分, NDBレコード順序, NDBレセプト通番, 処方年月日, 調剤年月日, 処方せん受付回, 負担区分, 調剤料算定区分, 調剤料算定先No, 調剤行為コード, 調剤行為名称, 調剤料点数, 調剤数量, 調剤行為種類, 薬剤料点数, 一包化日数, 分割調剤種類, 前回までの一包化日数
調剤加算料等 (TKA)	調剤加算料等ID, レセプトID, 解析用患者ID, 患者SID1, 患者SID2, 男女区分, 調剤時年齢区分, 調剤年月, レセプト区分, NDBレコード順序, NDBレセプト通番, 処方せん受付回, 負担区分, 加算料等コード, 加算料等名称, 加算料等点数, 調剤加算料等種別, 調剤数量, 回数, 前回調剤年月日, 前回調剤数量
DPC診断群分類 (BUD)	DPC診断群分類ID, レセプトID, 解析用患者ID, 患者SID1, 患者SID2, 男女区分, 診療時年齢区分, 診療年月, レセプト区分, NDBレコード順序, NDBレセプト通番, レセプト総括区分, DPC診断群分類番号, 傷病名, 手術名, DPC今回入院年月日, DPC今回退院年月日, DPC転帰区分, エピソード別入院ID
医療機関 (IRK)	医療機関ID, 都道府県コード, 点数表, 医療機関コード, 請求年月, 病床数区分
入院 (ADM)	エピソード別入院ID, 解析用患者ID, 男女区分, 入院開始時年齢区分, RE入院年月日, BU入院年月日, エピソード別入院年月日, BU退院年月日, 補完退院年月日, 入院合計点数, 入院合計診療実日数, 死亡フラグ, 医療機関ID, 都道府県コード, 病床数区分, DPC診断群分類番号, 傷病名, 手術名, DPC転帰区分

※下線で示したものは、テーブル間の連結用変数である。

表3 DPCレセプトにおける退院年月日の推定方法の検討

κ統計量	「今回退院入月日」と一致患者		「今回退院入月日」と不一致患者における誤差の分布				
	%		最小値	25パーセント タイル値	中央値	75パーセント タイル値	最大値
方法1	0.982	98.18%	-305	-1	6	14	369
方法2	0.998	99.83%	-26	1	1	1	26

方法1: レセプト共通 (RE) レコードにおける「入院年月日」データに保険者 (HO) レコードにおける「診療実日数」を加算する方法

方法2: 診療行為 (SI) レコード, 医薬品 (IY) レコード, 特定器材 (TO) レコード, コーディングデータ (CD) レコードにおける最終行為発生日を用いる方法

NDB 解析用データセットテーブルの開発

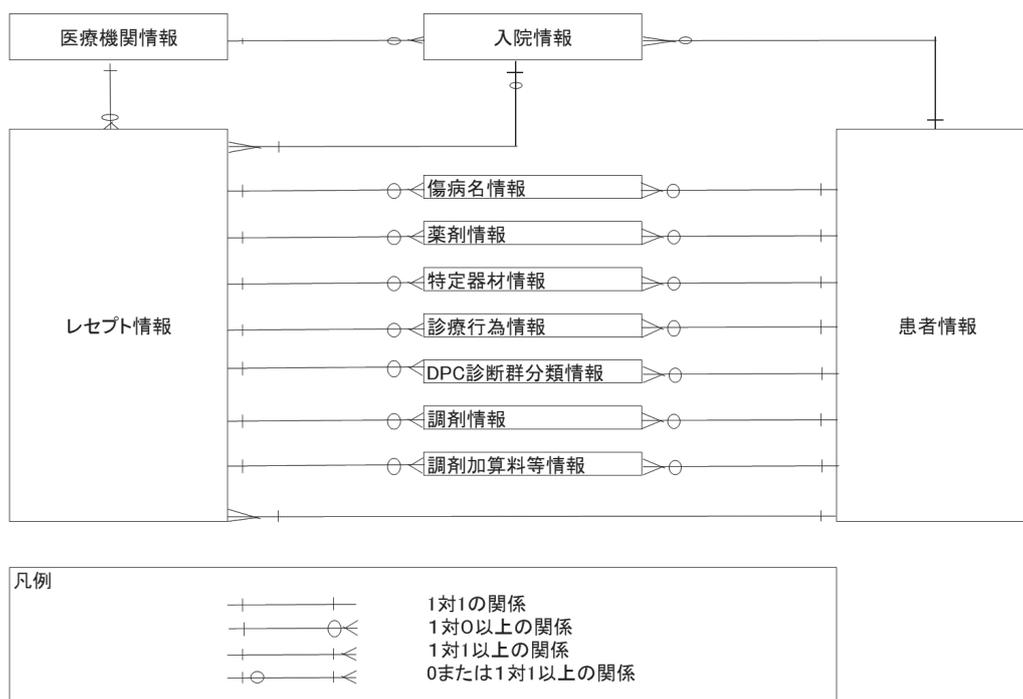


図1 解析用データベースのER図

6,585,250入院レコードを分析した結果、「補完退院年月日1」と「今回退院年月日」の一致率は98.18%であり、「補完退院年月日2」と「今回退院年月日」の一致率は99.83%であった。また、不一致した場合の入院レコードを対象に、両退院年月日との誤差は、「補完退院年月日1」を用いた場合は-305日から369日もの範囲があり、中央値は6日であった。一方、「補完退院年月日2」を用いた場合は-26日から26日の範囲であり、中央値は1日であった。 κ 統計量は、「補完退院年月日1」および「補完退院年月日2」でそれぞれ、0.982および0.998であった。

以上の検証の結果から、「補完退院年月日」はSI, IY, TO, CDの各レコードにおける最終行為発生日を用いる方法を用いることが妥当であると考えられた。我々はこの推定退院年月日データを入院情報（ADM）テーブルに付加した。

III. 考察

本研究では、NDBを利用している研究者等が、各々のリサーチ・クエスチョンを解明するための研究に即座に着手できるよう、データ解析用データセットテーブルを開発するものである。各テーブルに患者基本情報を付加することで、診療行為、医薬品、医療器材、傷病名、医療費などのレセプトの主要情報を各テーブル内で完結的に解析することを可能にした。また、研究者の解析目的に応じて適切なテーブルを新たに作成できるよう、

テーブル間に解析用患者IDや入院IDを付与した。さらに、入院エピソードは複数月にまたがることやDPCレセプトと医科レセプトにまたがることから、入院テーブルを作成し、分析の利用可能性を高めた。

本研究において開発したデータ解析用データセットテーブルは、様々なリサーチ・クエスチョンを解決するために活用可能である。例えば、外科的治療や感染症などの急性期を軸とした治療に関する分析を行う場合には、ADMテーブルを基軸とし、解析に必要な変数を他テーブルから突合し、必要な解析を実施可能なテーブルへと拡張することで、解析の利便性を高めることができる[8]。市中肺炎に関する研究であれば、IYAテーブルから抗菌薬に関するデータを抽出し、SYOテーブルから感染症病名に関するデータや併存疾患状況（例えば、Charlson comorbidity index [9]やElixhauser comorbidity index [10]）に関するデータを抽出し、これらのデータを入院IDを用いてADMテーブルに突合することができる。あるいは、糖尿病などの内科的治療が中心となる疾患に関する分析を行う場合には、KANテーブルから解析期間中の年月数分（2010年4月から2017年3月までなら84ヶ月分）の1患者ID・1月が1レコードとなるパネルデータを作成する。その後、IYAテーブルから糖尿病治療薬に関する医薬品データを抽出し、SYOテーブルから糖尿病性合併症に関する傷病名データや併存疾患状況に関するデータを抽出し、RECテーブルから各月の医療費に関するデータを抽出し、解析用患者IDと診療年月を用いて、パネルデータに突合することができる。それにより、

患者別に治療薬の使用状況, 合併症の発生状況および月別医療費を時系列的に観察可能な分析用データを構築することができる[11].

本研究において開発したデータセットテーブルで使用した解析用患者IDは, ID1あるいはID2によって設定するIDを用いた. これまでのNDBを用いた研究ではID1を使用する報告も[12-17], ID2を使用する報告[17-20]も認められている. しかしながら, ID1およびID2はそれぞれ個人追跡を妨げる要素が含まれていることから, 松居が提案するようにID1あるいはID2を用いて連結すべきであると考えられる[21]. ただし, ID2を使用した場合, 同姓, 同名, 同一生年月日の患者が同一IDが付与されてしまう問題は残されたままである. また, ID1とID2が同時に変更になった場合, 変更前後で解析用患者IDが変わってしまい連結不能になってしまう問題も残されたままである.

また, 本研究では, 医科入院レセプトに対して退院年月日を補完する方法を検討した. 最も簡便である「入院年月日」データに「診療実日数」を加算する方法では, 入院期間中に外泊などが発生する場合に, 退院年月日を正しく把握することができないためである. これまでのNDBを用いた報告では, 3研究において退院年月日が使用されていた[13,22,23]が, 医科レセプトに対する退院年月日の設定方法は述べられていなかった. しかし, 奥村は, 入退院日の不確実性について警鐘を鳴らしており, 入退院日を判定するために入院料と入院時食事療養費・入院時生活療養費の算定開始日と算定終了日を使用することを推奨している[24]. 本研究では, DPCレセプトを用いてBUに含まれている退院年月日を正解データとして上で, 入院年月日情報と診療実日数を用いて推測した退院年月日とあらゆる診療行為の最終発生日を用いて推測した退院年月日について, その一致状況を検証した. その結果, 診療行為の最終発生日を用いて推測することが高い精度で退院年月日を推測できることが明らかになった. しかし一方で, 本研究で使用したNDBは全ての診療行為情報を使用することが可能であったが, 通常の出発者は利用申請した診療行為情報のみしか使用することができない. したがって, 現実的には, 奥村が推奨するように, 入院料と入院時食事療養費・入院時生活療養費を申請した上で, 当該報酬項目の発生日情報を用いることが必要である. なお, DPCレセプトには入院料が包括化されており, 算定状況を把握することができないことから, 本研究において入院料と入院時食事療養費・入院時生活療養費の発生日から退院年月日をどの程度の精度で推測することができるかの検証は実施できなかった. なお, 「入院年月日」データに「診療実日数」を加算する方法において, 一部の症例において300日以上の誤差が発生する原因については, 本検証では明らかにすることができなかった. 本研究では利用申しなかった外泊(GA)レコードを用いたさらなる検討が必要である.

本解析用データセットの開発の過程において, NDB

を用いた診療分析を実施する際に注意すべき点があることも明らかになった. 第1にDPCレセプトにおいて, 診療行為, 医薬品, 特定器材の使用回数に関する分析を行う際の問題である. DPCレセプトには, これらの実施状況に関するデータは, それぞれSI, IY, TOに格納されることに加えて, CDにおいても格納されていることがある. そのため, SI, IY, TOのみを用いて分析した場合およびCDのみを用いて分析した場合は使用実態が過小評価され, SI, IY, TO, CDの全てを用いて分析した場合は, 一部のレコードが重複されることから使用実態が過大評価される. したがって, DPCレセプトを用いて使用量に関する分析を行う際には, SIN, IYA, TOKに付加した「CD重複フラグ」のない情報を用いて分析する必要がある. 第2にDPCレセプトにおいて, 医薬品等の使用量に関する分析を行う際の問題である. DPCレセプトにおけるCDレコードに記載のある医薬品情報の使用量に関する項目において, 医療機関におけるレセプトの入力ルールに2つの考え方が存在することが明らかになった. 例えば, 1缶あたり250mL含まれている栄養剤のレコードにおいて, 使用量として“1”が入力される場合もあれば, ごく一部ではあるが, “250”が入力される場合もあった. このようなレコードが支払基金によって査定されないのは, SIレコード, IYレコード, TOレコードでは, 「点数」項目の確認の際に点数計算で「使用量」項目が使用されるため誤った値が入力された場合は同定されるものの, CDレコードでは「使用量」項目の単独チェックで「0以外である」かどうかのみ査定される[25]. そのため, CDレコードにおいてこのような誤った値がレセプトデータに残ると考えられる. したがって, CD由来でIYAに格納されたレコードにおける使用量には注意が必要である(解析用データセットテーブルにおいて, SIN, IYA, TOKにCD由来フラグを付加している). この問題は, DPCレセプトに対して出来高換算医療費を推計する際にも障壁となる. 第3に, 診療行為をほぼ網羅的に把握可能なレセプトデータとはいえ, 医薬品の使用量を正確に把握することが困難な事例もある. 例えば, 実際には患者あたり1/2瓶使用しているにもかかわらず, レセプト上は1瓶を請求する実態がある. このとき, レセプトデータにおける使用量は実際の2倍になってしまう. このような場合, 各医薬品の添付文書情報から使用方法を確認した上で標準的な投与量を仮定するような処置が必要になることがある.

NDBは関係者の不断の努力により極めて有用なデータベースであることは間違いない. しかしながら全てのデータが完全であるとは言い難く, 一部のデータの扱いについては慎重な対応が求められる. 例えば, 医科レセプトに対する退院年月日の付与にあたっては, より妥当性が高いと考えられる最終行為発生日を用いる方法においても, 実際の退院年月日とは異なるレコードが0.17%において発生し, その大多数は1日の誤差であるものの, 最大で30日弱の誤差が生じる可能性があった. また,

- [3] 厚生労働省. オンライン又は光ディスク等による請求に係る記録条件仕様 (医科用). http://www.iryohoken.go.jp/shinryohoshu/file/spec/30bt1_1_kiroku.pdf (accessed 2018-11-22)
Ministry of Health, Labour and Welfare. [Online mataha hikari disc to ni yoru seikyu ni kakaru kiroku joken shiyo (ikayo).] (in Japanese)
http://www.iryohoken.go.jp/shinryohoshu/file/spec/30bt1_1_kiroku.pdf (accessed 2018-11-22)
- [4] 厚生労働省. オンライン又は光ディスク等による請求に係る記録条件仕様 (DPC用). http://www.iryohoken.go.jp/shinryohoshu/file/spec/30bt1_2_kiroku_dpc.pdf (accessed 2018-11-22)
Ministry of Health, Labour and Welfare. [Online mataha hikari disc to ni yoru seikyu ni kakaru kiroku joken shiyo (DPCyo).] (in Japanese)
http://www.iryohoken.go.jp/shinryohoshu/file/spec/30bt1_2_kiroku_dpc.pdf (accessed 2018-11-22)
- [5] 厚生労働省. オンライン又は光ディスク等による請求に係る記録条件仕様 (調剤用). http://www.iryohoken.go.jp/shinryohoshu/file/spec/30bt1_4_kiroku_chozai.pdf (accessed 2018.11.22)
Ministry of Health, Labour and Welfare. [Online mataha hikari disc to ni yoru seikyu ni kakaru kiroku joken shiyo (Chozaiyo).] (in Japanese)
http://www.iryohoken.go.jp/shinryohoshu/file/spec/30bt1_4_kiroku_chozai.pdf (accessed 2018.11.22)
- [6] 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. レセプト情報・特定健診等情報データベースの第三者提供. <https://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/0000117728.pdf> (accessed 2018-11-22)
Hoken system kodoka suishinsitsu, Iryo Kaigo Renkei Seisakuka, Health Insurance Bureau, Ministry of Health, Labour and Welfare. [Reseputo joho / tokutei kenshin to joho database no daisansha teikyo.] (in Japanese)
<https://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/0000117728.pdf> (accessed 2018-11-22)
- [7] 厚生労働省保険局. 診療報酬情報提供サービス. <http://www.iryohoken.go.jp/shinryohoshu/> (accessed 2018-11-22)
Health Insurance Bureau, Ministry of Health, Labour and Welfare. [shinryo hoshu joho teikyo service.] (in Japanese)
<http://www.iryohoken.go.jp/shinryohoshu/> (accessed 2018-11-22)
- [8] Fukuda H, Yano T, Shimono N. Inpatient expenditures attributable to hospital-onset *Clostridium difficile* infection: a nationwide case-control study in Japan. *Pharmacoeconomics*. 2018. in press.
- [9] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.
- [10] Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27.
- [11] Fukuda H, Ishida H, Sato D, Moriwaki K. Healthcare expenditures for the treatment of patients with hepatocellular carcinoma in Japan. *ISPOR Europe 2018*; November 14, 2018; Barcelona.
- [12] Toyokawa S, Maeda E, Kobayashi Y. Estimation of the number of children with cerebral palsy using nationwide health insurance claims data in Japan. *Dev Med Child Neurol*. 2017;59:317-321.
- [13] Okumura Y, Sakata N, Takahashi K, Nishi D, Tachimori H. Epidemiology of overdose episodes from the period prior to hospitalization for drug poisoning until discharge in Japan: An exploratory descriptive study using a nationwide claims database. *J Epidemiol*. 2017;27:373-380.
- [14] Hagiwara H, Nakano S, Ogawa Y, Tohkin M. The effectiveness of risk communication regarding drug safety information: a nationwide survey by the Japanese public health insurance claims data. *J Clin Pharm Ther*. 2015;40:273-278.
- [15] Hagiwara H, Nishikawa R, Fukuzawa K, Tohkin M. The survey of the compliance situation to the antihypertensive therapy guideline by analyzing Japanese national claims data. *Yakugaku Zasshi*. 2017;137(7):893-901.
- [16] 關真美, 椿広計. サンプルングデータセットを用いた併用禁止医薬品等の処方実態研究. *医療情報学*. 2015;34:293-304.
Seki M, Tsubaki H. [Drug utilization study of potential drug-drug interactions using the sampling dataset of the National Database of Health Insurance Claim Information.] *Japan Journal of Medical Informatics*. 2015;34:293-304. (in Japanese)
- [17] Kubota K, Kamijima Y, Sato T, Ooba N, Koide D, Iizuka H. Epidemiology of psoriasis and palmoplantar pustulosis: a nationwide study using the Japanese national claims database. *BMJ Open*. 2015;5:e006450.
- [18] 高田充隆. ナショナルデータベースを用いた低用量アスピリン療法における消化管傷害リスクに関する研究. *医療薬学*. 2013;39:471-481.
Takada M. [Study on risk of gastrointestinal complications in low-dose aspirin therapy using the National Receipt Database.] *Iryo Yakugaku*. 2013;39:471-481. (in Japanese)
- [19] 飯原なおみ, 吉田知司, 岡田岳人, 他. わが国のナショナルレセプトデータベースが示した運転等禁止・注

- 意医薬品の使用実態. 医療薬学. 2014;40:67-77.
- Iihara N, Yoshida T, Okada T, et al. [Survey of usage of medication with driving with prohibition or caution by the National Health Insurance Claims Database in Japan.] *Iryo Yakugaku*. 2014;40:67-77. (in Japanese)
- [20] 細見光一. ナショナルデータベースを用いた抗精神病薬による錐体外路系症状のリスクに関する解析. 医療薬学. 2016;42:87-97.
- Hosomi K. [Association of antipsychotic use with extrapyramidal symptoms: data mining of the Japanese National Insurance Claims Database.] *Iryo Yakugaku*. 2016;42:87-97. (in Japanese)
- [21] 松居宏樹, 佐藤大介. レセプト情報等オンラインリサーチセンターの試行的利用に関する中間報告. <http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000128901.pdf> (accessed 2018-05-21)
- Matsui H, Sato D. [Reseputo joho to on site research center no shigoteki riyo ni kansuru chukan hokoku.] (in Japanese)
- <http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000128901.pdf> (accessed 2018-05-21)
- [22] Okumura Y, Nishi D. Risk of recurrent overdose associated with prescribing patterns of psychotropic medications after nonfatal overdose. *Neuropsychiatr Dis Treat*. 2017;13:653-665.
- [23] Otsubo T, Goto E, Morishima T, Ikai H, Yokota C, Minematsu K. Regional variations in in-hospital mortality, care processes, and spending in acute ischemic stroke patients in Japan. *J Stroke Cerebrovasc Dis*. 2015;24:239-251.
- [24] 奥村泰之, 佐方信夫, 清水沙友里, 松居宏樹. ナショナルデータベースの学術利用促進に向けて: レセプトの落とし穴. *Monthly IHEP*. 2017年10月号. p.16-25.
- Okumura Y, Sakata N, Shimizu S, Matsui H. [National Database no gakujutsu riyo sokushin ni mukete: Reseputo no otoshiana.] *Monthly IHEP*. Oct. 2017. p.16-25. (in Japanese)
- [25] 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(DPC). http://www.ssk.or.jp/seikyushiharai/rezept/iryokikan/download/index.files/checklogic_dpc.pdf (accessed 2018-11-22)
- Health Insurance Claims Review & Reimbursement Services. [Reseputo online seikyu uketsuke / jimu tenken ASP ni kakaru check logic (DPC).] (in Japanese)
- http://www.ssk.or.jp/seikyushiharai/rezept/iryokikan/download/index.files/checklogic_dpc.pdf (accessed 2018-11-22)