

**i. 台湾 DOH Office of Statistics (衛生署統計室) における National Health Informatics Project (NHIP)**

台湾は、韓国と同様に個人識別番号を第二次世界大戦後に導入し、個人識別番号は台湾における社会生活上に深く入り込んでいる。台湾の医療保険請求のデータは研究用データベース (National Health Insurance Research Database, NHIRD) として、広く台湾の研究者・企業に公開されているが、行政レベルでは、医療保険請求のデータのそれ以外のデータとのレコードリンケージを積極的に行い、保険行政に活用する取り組みが進行している。

Department of Health (衛生署) 内の collaboration center of health information application project が中心となって、2009-2011年の3年計画で進められている National Health Informatics Project (NHIP) がその取り組みである。NHIP では 14 カテゴリー、26種類の Database(全て個人単位でリンク可能)の処理・分析能力、レコードリンケージの能力の向上を課題としており、その結果を行政支援に結びつけることを目標としている。

このプロジェクトで連結することが可能な 14 カテゴリーのデータは以下の通りである。

- [1]. 外来受診の支出データ (National Health Insurance, NHI)
- [2]. 外来受診に関する詳細データ (薬、受診など) (NHI)
- [3]. 保険加入者 DB (NHI)
- [4]. 高額医療費の患者 (catastrophic illness patients) 登録 (NHI)
- [5]. 世帯データ登録 (台湾内政部 Ministry of the Interior)
- [6]. 死亡登録と死亡の複数原因データ

(Division of Statistics, DOH)

- [7]. 先住民のデータ (Council of Indigenous Peoples, CIP 原住民族委員会)
- [8]. 医療機関の医療資源 (Office of Statistics, DOH)
- [9]. 医療機関の医療サービスの利用 (Office of Statistics, DOH)
- [10]. 医療機関の認定プロファイル (Bureau of Medical Affairs, DOH)
- [11]. 出生登録 (Bureau of Health Promotion, BHP、国民健康局 DOH)
- [12]. 国民健康訪問調査(サンプル世帯に対するインタビュー) (BHP、DOH)
- [13]. がん登録 (BHP、DOH)
- [14]. 家庭/国の健康支出サーベイ Office of Statistics DOH

これら、14 のデータソースのいくつか、たとえば健康保険とがん登録や死亡に関するデータをサーバー上でリンクさせ、解析・報告のためのツールにつなげている。

これらの作業は現在、18 台の端末がおかれた Project Office 内の independent operation area において行われている。

2009年6月から11月の6ヶ月間に Office of Statistics (統計室) 22 件、Center for Disease Control (CDC)12 件、Bureau of Health Promotion (BPH) 12 件など 6 つの行政内の部署からの依頼で合計 53 件のレコードリンケージが実施された。

**2. 研究者・企業レベルにおけるレコードリンケージ**

**① 米国**

**i. Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital**

Dr. Setoguchi が所属する Division of Pharmacoepidemiology and Pharmaco

-economics, Brigham and Women's Hospital において利用しているデータベースについて、個人識別子の有無と個人識別子によるレコードリンケージについて記載する。

**a. Medicare**

Social Security Number (SSN) のデータが含まれている。ただし、近年 SSN は得にくくなっている。Pennsylvania 州と New Jersey 州のデータを利用しており、薬に関しては Pennsylvania 州と New Jersey 州の State Pharmacy Assistance Programs をその他のデータと SSN でリンケージして利用する。

**b. Medicaid**

New Jersey 州の Medicaid のデータを利用しており、このデータには SSN が含まれる。

**c. MCBS (Medicare Current Beneficiary Survey)**

通常の Medicare のデータにはないデータを含むが一回限りのサーベイのデータである。SSN のデータを含んでいる。

**d. Medicaid Analytic Extract from 50 states**

2006 年における全米 50 州の 5630 万人分のデータ。SSN は含まない。

**e. Pennsylvania 州と New Jersey 州のがん登録**

State Cancer Registry のデータであり SSN を含む。

**f. Massachusetts 州の HMO の Blue Cross Blue Shield のデータベース**

SSN を含まない。

**g. 疾患登録のデータ**

Coronary artery diseases, heart failure, and stroke registries に関するデータであり、Outcome Science Inc. から得たもの。SSN は含まないが、入・退院日、DOB、性、州のデータを含む。

**h. British Columbia の Provincial Claims Data**

カナダの British Columbia のデータベース。

**i. USRDS (US Renal Data System)**

Medicare の対象者は 65 歳以上、末期腎不全、その他特定疾患であるが、このうち腎疾患の患者データであり、Medicare の対象となることから Medicare のデータが含まれている。

**j. Dispensing data**

US の 50 州の大手 Pharmacy のチェーン"CVS"の Dispensing data。SSN は含まれないが、Division of Pharmacology and Pharmacoepidemiology and Pharmacoeconomics において氏名など SSN 以外の情報を使って Medicare のデータとリンクされている。

**k. Premier データベース**

米国の全入院の 6 分の 1 をカバーする入院患者のデータ：現在アメリカの保険会社の United Health のレコードとのリンケージが試みられつつある。

**1. リンクに関する詳細**

Medicare のデータとリンクされている USRDS のデータのように、入手前にすでになんらかのレコードリンケージが行われているデータベースも存在するが、研究者レベルでレコードリンケージが必要な場合も多い。

レコードリンケージのためには、SSN などの個人ごとに完全に unique な識別子によるリンクが最良である。ただし、SSN のレコードにエラーがある場合もあり、その他の情報も同時に使われることが多い。

SSN 以外の複数の完全に固有ではない識別子 (DOB、性、人種、州、保険者または病院の ID、入院日、退院日、DOD、入院の Primary Diagnosis、手技 (procedure)、患者の ZIP など) を使う probabilistic linkage も行われるが、一般に SSN を使う場合よりは一致率は落ちる<sup>8)</sup>。

また、Dr. Setoguchi は 2000-2002 年の

Medicaid のデータ内で母と子のリンクを試みたことがある。Medicaid case Number には家族単位を識別可能な単位をもち、ほかに母の出産日と子の DOB、居住郡 (county)、出生した病院、母の名前、子の名前を用いてリンクを実施し、92-93%をマッチしえたという。

#### m. 個人情報の取り扱い

個人情報を含むデータの扱いにあたっては、研究計画書の作成が必須であり、さらにデータを取得する際の契約に

- [1]. データをリンクして研究用データを構築するために使う個人識別子は研究に従事するプログラマーのみがアクセス可能である。
- [2]. リンクにより研究用データが作成されたのち、SSN などの個人識別子は研究用データから削除する。

などを含めることが要求されることが多い。

そのほか、以下のような項目が研究計画書や契約書に含まれることが多い。

- [3]. データを入れるコンピュータは他の部屋から隔離された鍵のかかるコンピュータ室におかれる。
- [4]. データを入れるコンピュータはサーバー、イントラネット、インターネットに接続しない。
- [5]. オフィス自体がセキュリティの高いビルディング内に存在する。
- [6]. コンピュータ室には施錠され、コンピュータプログラマのみが入退室が可能。
- [7]. もともとのデータは金属のクローゼットにおいて施錠されて保管される。
- [8]. コンピュータのデータへのアクセスにはパスワードが必要。

#### ii. i3 Drug Safety

##### a. i3 aperio

Dr. Arnold Chan が属する i3 Drug Safety 社

の i3 aperio は薬の安全性に関するシグナル検出のツールである。

このツールの基本は、米国で最大規模の医療保険会社である United Health 社による被保険者の各種の電子的データをレコードリンクさせて、リアルタイムで利用可能とするところにある。新薬使用者と”対照”薬使用者が傾向スコアマッチングであらかじめ抽出されており、診断、薬の（新規）使用などのイベント情報の発生率の比較によってシグナルを検出する<sup>9)</sup>。

##### b. in-depth study

シグナルが検出された場合、さらに詳細な調査が必要とされれば、カルテに遡及するなどして病名を検証 (validation) するなど、より質の高い研究に結びつけることも可能である。

#### ② 韓国 HIRA データベース

##### i. HIRA データベースの成立直後

HIRA が設立されたのは、2000 年 7 月 1 日であるが、HIRA のデータから釜山 (Busan) の病院/クリニックを 2000 年 11 月 1 日-2002 年 4 月 20 日に受診し、シサプリドを新規処方された 85 歳以下の 36865 人のデータを抽出し、Korea National Statistical Office から提供された National Mortality Data と個人レベルでリンクする研究が実施された<sup>10)</sup>。

これを計画した段階では、本研究は Seoul 大学の倫理委員会の審査対象ではなかったという。Cox 回帰モデルで検討した結果、併用禁忌薬と併用されていた患者における死亡率は併用禁忌薬と併用されていなかった患者の死亡率より高かった。

##### ii. 個人データの扱いに関する規制

韓国の National Statistical Office は 2005 年に National Mortality Data の研究への提供を停止した。これに伴い HIRA もそのデータの研

究者への提供に慎重な姿勢をとるようになった。ただし、KFDA など行政によって実施される調査においては、レコードリンケージは通常的手段として利用されている。

このような状況の変化の中で、韓国の研究者は KFDA の調査を受託する形でレコードリンケージを含む HIRA データベースを用いた研究を実施するようになった。たとえば 2007 年 3 月 21 日から同年 11 月 30 日にかけて KFDA の監督の下に実施された調査("drug utilization monitoring for frail population, specific impaired patients, contraindicated drug combinations, age-limited drugs and drugs with narrow therapeutic index") には以下の 7 つが含まれる。

- [1]. drug use in elderly
- [2]. pediatric drug use
- [3]. drug use in hepatic diseases
- [4]. drug use in renal diseases
- [5]. drug-drug interaction
- [6]. drug use in pregnancy
- [7]. drug use of narrow therapeutic index

このうち[1]については Seoul 大学が受託し、結果の一部については、医学雑誌への公表が許された<sup>11)</sup>。

## ii. 韓国内におけるプライバシー保護に関する論争

上記の規制強化は、韓国内の人権保護団体による、個人情報を用いた研究を全て禁止させようとする運動などと関連している。

Korean Society of Preventive Medicine は、学会内での 1 年間の検討後、2007 年の夏に HIRA データベースを研究者が提案した研究のために使用する際の基準に関する Recommendation を発表し、現在 HIRA 内でのこの問題の検討に利用されている。

【平成 22 年度報告書への補遺】

Seoul 大学の Professor BJ Park によれば、2010 年に入り、研究者の HIRA データベースの利用のあり方が相当程度前進した。たとえば、研究者による HIRA データベース利用に関する申請は最近ではほとんど全てが承認されている。ただし、データの HIRA 外部への持ち出しは実現しておらず、研究者は HIRA の内部で解析をすることが義務付けられている。結果の公表に関するルールについては 2010 年 12 月現在不明である。

さらに台湾と同様、韓国における 2009 年の入院患者と外来患者を代表するそれぞれ 50 万人のコホートを抽出し、研究への利用を促している。現時点では、これらのデータを利用可能なのは研究者のみで、企業については許可されていない。

## ③ 台湾 NHIRD

台湾において、研究者・企業は医療費請求のためのデータを他のデータソースとのリンケージの上利用することは許されていない。したがって、死亡に関するデータ、検査結果に関するデータを解析対象とすることは不可能である。ただし、この制約の元ではあれ、台湾においては、研究者・企業に医療費請求のデータを研究用にデータベース化した National Health Insurance Research Database (NHIRD) の利用を積極的に促す仕組みが存在する。ここでは、その仕組みについて記述する。

### i. 台湾のデータベースの研究者・企業による利用の現状

#### a. データの流れ

台湾における医療費請求のために医療機関、薬局が発行する claims のデータは NHIB(National Health Insurance Bureau) から保険局 (BNHI、Bureau of National Health Insurance) に送られ、ここで患者、医師、病院、

薬局の ID が scramble(暗号化)された上で、国家衛生研究院 (National Health Research Institute) NHRI に 1 年に 1 回送付される。NHRI でさらに標準データまたは、Special request に対するデータ抽出が行われる。データの発生から研究者などの手にデータが届くまでには、現在 12~24 ヶ月を要する。

#### b. NHRI に送付されるデータ

BNHI から NHRI に送付されるのは Original Claims Data、Summary Data、Registration File であり、前述の通り個人 ID は scramble (暗号化) され、NHRI にあるデータから個人を識別することはできない。

Registration Files に含まれるのは以下の 9 種類である。

- [R1]BED : registry for contracted beds
- [R2]DETA : registry for contracted specialty services
- [R3]HOSP : registry for contracted medical facilities
- [R4]HOSX : supplementary registry for contracted medical facilities
- [R5]DOC : registry for board-certified specialties
- [R6]PER : registry for medical-certifier
- [R7]HV : registry for catastrophic illness patients
- [R8]DRUG : registry for drug prescription
- [R9]ID : registry for beneficiaries

Claims data に含まれるのは以下の 9 種類である。

- [C1]DT : monthly claim summary for inpatients claims
- [C2]CT : monthly claim summary for ambulatory care claims
- [C3]DD : inpatient expenditure by admission :  
Diagnosis: 1 prime + 4 (max) sub <

1 record= 1 admission per month>

- [C4]DO : details of inpatients others
- [C5]CD : ambulatory care expenditure by visit : Diagnosis: 1 prime + 2 (max) sub <1 record= 1 visit>
- [C6]CO : details of ambulatory care others
- [C7]GD : expenditure for prescription dispenses at contracted pharmacies
- [C8]GO : details of prescription dispenses at contracted pharmacies
- [C9]OO : drug number.

#### c. NHRI において作成される標準データ

##### [1]. Longitudinal Health Insurance Database (LHID)

2000 cohort: 2000 に抽出した 100 万人のデータ : 1 年分で CD20 枚 (1 枚 50,000 人/年)

2005 cohort: 2005 に抽出した 100 万人のデータ : 1 年分で CD25 枚 (1 枚 40,000 人/年)

2000 cohort と 2005 cohort には overlap はない。2005 年以後に出生した者は 2000 cohort には含まれないが、2005 年以前に死亡した者は 2005 cohort には含まれない。

これら 100 万人 X2 分のコホートには CD, CO, DD, DO, GG, GO ファイルの内容が全て含まれる

##### [2]. Systematic Sampling (各年ごと)

DD の 5% : 1 record= 1 admission per month、一人の患者は最大(1 年間連続して入院している場合)12records をもちうる。この場合、そのうちのひとつまたはそれ以上がサンプルされうる。

CD の 0.2% : 1 record= 1 visit、一人の患者は多数の records をもちうる。そのうちの 1 つまたはそれ以上がサンプルされうる。

##### [3]. Specific Subject Datasets

特定の診断または薬の使用に関する記録

が抽出される。これらの診断には以下が含まれる。

- Diabetes
- Psychiatric Impatient
- Traditional Chinese Medicine
- Dental Care

#### [4]. Teaching demo dataset

教育用のデータセットであり、1000 人分の 3 年間分の longitudinal data が含まれる。

#### d. NHRI で抽出される特別のデータ

上記標準的のデータに加えて、研究者などからの special request に対して特別の抽出が行われる。

#### e. データの保護

NHRI がデータ保護のためにとっている原則は以下の通りである。

##### [1]. Anonymous :

患者、医師、病院、薬局の ID は BNHI から送付する段階で scrambling されているが、NHRI から研究者などに送付する段階でさらに（二重に）scrambling する方策がとられている。

##### [2]. Research protocol を要求 :

Review:2 週/件 10 人の reviewers が 1 件あたり 2 人でレビューする。研究を促進することが目的であり、最終的に拒絶することはほとんどなく、あるとすれば「希望するデータはない」ために、利用を断念するケースであるという。

##### [3]. Minimal amount of data :

それぞれの研究に必要な最小限のデータのみが提供/抽出される。毎年提供を受けることができるのは全データの 10%以内。

#### f. データの提供

データを申請できるのは台湾国民に限定されており、台湾の研究者または台湾に籍を

おく企業に対して提供されており、研究者には 2000 年から、企業には 2003 年からデータが提供されている。CD ROM で提供される場合、研究者には 500NT/ICD (約 1500 円/ICD) で提供される。Special request に対しては tape で研究者には、200NTD/1GB(約 600 円/1GB)で提供される。

2001 年から 2007 年に 707 件の請求があり、最近 3 年間に急増し、2-300 件/年に達している。このうち 1/3 が special request であり、2/3 が標準ファイルを購入した。

#### g. 学術的成果

NHRI が PubMed で 2010 年はじめに検索したところ、NHID (national health insurance database):200 件、NHIRD (national health insurance research database): 163 件がヒットした。

年度と件数は表 1 の通りである。

#### h. 人員

年間 2-300 件のデータ要求に対して 7~8 人で対応。この 7~8 人が以下に対応

- Data extraction
- Contract
- Communication

#### ii. 台湾のデータベースの研究者・企業による利用の将来

現状では、台湾において研究者・企業はレコードリンケージは許されてはいない。しかし、個人 ID を有する医療費償還のための請求書、がん登録、死亡データのレコードリンケージは行政レベルでは盛んに行われている。

DOH の Head Office の Deputy Minister Chen(衛生署 陳再晋副署長)は、将来的に、研究者（など）行政以外のユーザーが個人レベ

ルでリンクされたデータを使えるようになる、との個人的見通しを本分担研究者に明らかにした。

#### 【平成 22 年度報告書への補遺】

台湾の National Cheng Kung University (成功大学) の Professor Yea-Huei Kao Yang から 2010 年 12 月 27 日に得た情報によれば、NHRI は 2000 年と 2005 年の 100 万人のデータについて、死亡と死因の情報の提供を開始した。ただし、死亡に関する情報の利用は NHRI 内の指定された部屋においてのみ許され、適切であることの確認を得てのち解析結果のみを持ち帰ることができる。また死亡に関する原資料の閲覧は許されていない (2010 年 12 月 27 日現在においてアクセス可能な NHRI による死亡と死因データの提供に関する情報については NHRI の中国語ページ参照<sup>14)</sup>)。利用に若干の制約があるが、リンクされた死亡データの提供が開始されたことで台湾のデータベース研究の質のさらなる飛躍が期待される。

#### D. 考察

本稿では医療費償還請求のデータの個人単位でのレコードリンケージのあり方を、行政レベル、研究者・企業レベルに分け、米国、韓国、台湾における現状について記述した。

韓国、台湾に共通しているのは、国民それぞれが SSN など固有の番号を有していること、国民皆保険が成立していること、また保険者が単一(国)であることであり、保険者としての本来の業務を司る国は、国民の医療費償還請求のデータを個人識別子とともに管理する立場にある。韓国、台湾ともに、行政内における個人識別子によるレコードリンケージにより、情報価値の高い行政的資料として活用している。これに対し、米国においては、国民が SSN という固有の番号を有

している点では共通しているが、医療保険は複雑で分散しており、Medicare、Medicaid など、行政が直接保険者となるものを除いては、行政資料として、個人識別子で医療費償還請求のデータと他のデータを結合した新たな情報を用いることには十分成功しているとはいいがたい。ただし、米国の保険者の多くは、それぞれのデータを他のデータとリンケージする学術研究を含めて、自らの有するデータを多様に利用するノウハウをもっており、Sentinel Initiative で構想されている通り、行政はこれらのデータ管理者の協力を得て、サマリのデータとしてではあれ、必要な情報を得る道筋を描くことに成功しつつある。

研究者・企業レベルにおいては、韓国、台湾ともに、個人識別子を用いたレコードリンケージは、行政による調査の「請負」という形態を除いては実施不可能である。この点は、研究者自ら SSN などを用いたレコードリンケージが相当程度可能な米国と対比をなしている。

スカンディナヴィア諸国やカナダなどでは、研究者による個人識別子を用いたレコードリンケージがより完全に近い形で実施されている。アジア諸国に比べ、欧米ではプライバシー保護と公衆衛生上の必要性のバランスをいかにとるべきかについての社会的コンセンサスが成立する種々の条件が整っているといえるのかも知れない。

ただし、台湾における NHIRD の利用を研究者に奨励するシステムは、よく機能しており、定着しつつある。近年、台湾の HHIRD を使った国際雑誌に受理される論文も急増しており、個人識別子によるレコードリンケージの許可と、研究へのデータベースの利用の奨励とは必ずしも並行しない。さらに、台湾の行政内では、医療費償還請求のデータと疾患登録や死亡データとのリンケージが盛

んに行われており、比較的近い将来これらのレコードと結合済みのデータを台湾の研究者・企業が利用できるようになる可能性も相当高いと考えられる。

わが国におけるレセプトデータの利用については、台湾におけるような利用促進の体制の重要性が十分認識されているとはいいたがたい。

また、わが国のレセプトデータベースに関しては、研究者レベルのみならず、行政レベルでも他のデータとのレコードリンケージは当面困難と考えられる。SSNのような個人識別子の重要性に関する社会的な議論を重ね、保険者が単一でないとしても、個人識別子を中央のレセプトデータベースにもたせることを可能とする法的措置を含めたなんらかの可能性の検討などが将来的には必要になると考えられる。

また、これらの方策が十分進まない間は、レセプトデータの有用性に大きな限界があることを認識し、使用成績調査における科学性の向上など一次データを用いた調査研究の改良、効率化などにも取り組むべきと考えられる。

#### 【平成 22 年度報告書への補遺】

わが国のレセプトデータベースの研究者などへの提供に関しては、現在厚生労働省保険局によって組織された「レセプト情報等の提供に関する有識者会議」（座長：開原成允国際医療福祉大学副学長）で議論されており、会議での審査基準の策定と平成 23 年 4 月からのデータ提供開始が計画されている<sup>13)</sup>。レセプトデータベース利用への期待が高まる一方、匿名化が前提であるため、レコードリンケージは困難であり、「専任の職員は 1 名」（平成 22 年 11 月 25 日第三回会議の資料 2 の 21/22）<sup>15)</sup>などデータ提供のあり方に関する

不安要素も多い。平成 22 年 11 月の医療情報学会においては医薬品の安全性確保を目的としたレセプトデータベースの利用には厚生労働省の安全対策課の職員から「利用は難しい」との悲観的観測が述べられており<sup>16)</sup>、今後も引き続きデータベース利用の環境のあり方に注視が必要と考えられる。

#### E. 結論

医療費償還請求のデータの価値を高めるために必要な、個人単位での他のデータとのレコードリンケージの米国、韓国、台湾における現状を行政レベルと研究者・企業レベルに分けて考察した。

国が唯一の保険者である韓国、台湾では、行政レベルでは、レコードリンケージをすでに実現し、価値の高い行政的資料とすることに成功している。個人単位でリンケージされたデータを研究者・企業レベルでも利用可能になるためのインフラの整備はすでに整っており、今後、国民的コンセンサスが成立すれば、研究者・企業が他のデータのリンケージされた医療費償還請求データを質の高い調査・研究に利用できるようになるのには、時間の問題とも考えられる。保険者が分散している米国においては、データ所有者の協力を得て、行政が単一の問いに対する回答を得るシステムを Sentinel Initiative として構築しつつあり、また研究者には、完全ではなくとも個人単位のレコードリンケージが可能な道が存在する。

わが国のレセプトデータベースに関しては、多数の保険者が個人データを管理し、当面中央レベルのデータベースを直接他のデータとリンクさせることができる可能性は研究者・企業レベルのみならず、行政レベルでも高いとはいいたがたい。個人識別子の必要性や、個人識別子を中央のレセプトデータベ



ースにもたせる法的措置を含めたなんらかの可能性の検討がいずれは必要となろう。さらに、データベースの利用を促すためには、台湾の NHRI のような組織を構築するなど、相当程度の投資を含めた環境作りが必要であることを認識し、その実現に向けた努力が必要と考えられる。

#### F. 健康危険情報

なし

#### G. 研究発表

##### 1. 論文発表

なし

##### 2. 学会発表

なし

#### H. 知的財産権の出願・登録状況

##### 1. 特許取得

なし

##### 2. 実用新案登録

なし

##### 3. その他

なし

#### I. 参考文献

1. Olsen J. Using secondary data. In: Rothman KJ, Greenland S, Lash TL. Eds. Modern epidemiology, 3rd ed. Lippincott Williams & Wilkins, Philadelphia, 2008, p481-91.
2. Edgren G, Hjalgrim H, Reilly M et al. Risk of cancer after blood transfusion from donors with subclinical cancer: a retrospective cohort study. *Lancet* 2007; 369: 1724-30.
3. Haukka J, Arffman M, et al. Antidepressant use and mortality in Finland: a register-linkage study from a nationwide cohort. *Eur J Clin Pharmacol* 2009; 65: 715-720.
4. Sørensen R, Hansen ML, Abildstrom SZ. Risk of bleeding in patients with acute myocardial infarction treated with different combinations of aspirin, clopidogrel, and vitamin K antagonists in Denmark: a retrospective analysis of nationwide registry data. *Lancet* 2009; 374: 1967-74.
5. Karp I, Behloul H, LeLorier J, Pilote L. Statins and Cancer Risk. *Amer J Med* 2008; 121: 302-309.
6. FDA's Sentinel Initiative. <http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm>. 2010年4月アクセス。
7. Velentgas P, Bohn RL, Brown JS. A distributed research network model for post-marketing safety studies: the Meningococcal Vaccine Study. *Pharmacoepidemiol Drug Saf* 2008; 17: 1226-1234.
8. Hammill BG, Hernandez AF, Peterson ED. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J* 2009; 157:995-1000.
9. Dorea DD, Seegera JD, Chan KA. Use of a claims-based active drug safety surveillance system to assess the risk of acute pancreatitis with exenatide or sitagliptin compared to metformin or glyburide. *Curr Med Res Opinion* 2009; 25:1019-1027.
10. Choi NK, Hahn S, Park BJ. Increase in mortality rate following coprescription of cisapride and contraindicated drugs. *Ann Pharmacother* 2007; 41: 667-673.
11. J Lee et al. The risk of fracture with taking alpha blockers for treating benign prostatic hyperplasia. *J Prev Med Public Health* 2009; 42: 165-70.
12. Stang PE, Ryan PB, Racoosin JA et al., Advancing the science for active surveillance: rational and design for the

observational medical outcomes partnership.

Ann Inter Med 2010; 153: 600-606.

13. 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.

<http://www.mhlw.go.jp/stf/shingi/2r9852000000amvy.html> 2010年12月アクセス。

14. 国家衛生研究院 全民健康保険研究資料  
庫資料加値現場分析(On-site)サービス作業須知.

[http://w3.nhri.org.tw/nhird/onsite\\_01.htm#A2](http://w3.nhri.org.tw/nhird/onsite_01.htm#A2)

2010年12月アクセス

15. 第3回レセプト情報等の提供に関する有識者会議資料2: データ提供にあたってのポイント.

<http://www.mhlw.go.jp/stf/shingi/2r9852000000x6hp-att/2r9852000000x6m9.pdf> 2010年12

月アクセス

16. 日本薬剤疫学会・日本医療情報学会・日本製薬工業協会 共同企画シンポジウム  
「医療データベースの二次利用による薬剤疫学研究の今後の課題」. アクトシティ  
浜松、2010年11月20日.

表1 台湾 NHIRD を利用した学術論文の発表数

	<b>NHID</b>	<b>NHIRD</b>
<b>96</b>	1	0
<b>02</b>	4	4
<b>03</b>	9	7
<b>04</b>	15	11
<b>05</b>	8	3
<b>06</b>	28	22
<b>07</b>	27	18
<b>08</b>	53	47
<b>09</b>	55	51
<b>Total</b>	<b>200</b>	<b>163</b>

平成 22 年度厚生労働科学研究費補助金  
(医薬品・医療機器等レギュラトリーサイエンス総合研究事業)  
分担研究報告書

レセプトデータと医療情報データベースとの匿名化連結可能性と  
レセプト病名の妥当性に関する研究

研究分担者 折井 孝男 NTT 東日本関東病院・薬剤部長

研究要旨：

初年度の研究として、医療機関（病院）におけるレセプト（診療報酬明細書）データをハッシュ関数による匿名化識別子を用い、電子カルテ等の医療情報データベースとの連結可能性について検討した。その結果、病院で発生するレセプトのデータには患者個々の診療に関する診断、処方、検査、放射線、あるいは看護等の情報が記載されており、電子カルテの保有するデータとともに匿名化することにより各々のデータのマッチングを行うことができた。この研究結果の詳細な内容としては、医科、傷病件数、医薬品、診療行為、材料について、匿名化したデータの連結について実証できることを明らかとした。

そこで、今年度以降の研究として、診療に関する診療年月、患者の年齢、性別、医療機関のベッド数、経営体、診療科、診断された全ての疾病名（確定/疑い）、診療実日数、指導管理、在宅、投薬、注射、処置、手術麻酔、検査、画像診断、さらに請求点数数などさまざまな情報が記載されているレセプトデータと電子カルテデータとを匿名化することにより、より多くの診療情報との連結について検討できることから、病院における様々な診療情報を保有する電子カルテデータについて調査した。

初年度の研究結果から、レセプトデータをハッシュ関数による匿名化識別子を用い、電子カルテデータとの連結は可能であることから、今後医療現場における匿名化連結可能性について実証実験を重ねることは必須といえる。さらにレセプト病名についてもその妥当性に関する検証が可能となることから、匿名化された電子カルテデータとの詳細な連結可能性について実証する必要がある。

A. 研究目的

NTT 東日本関東病院（以下、関東病院）で発生するレセプトデータ、電子カルテデータを匿名化し、電子レセプトファイルより個人情報情報を匿名化したデータを生成し、データセンター等が保有するデータと比較しデータの詳細な部分までの正当性を検証する体制を確立する。

B. 研究方法

本研究では、レセプトデータの収集をデータセンター等からの利用を考え、電子カルテ側の

データとのマッチングについて検証する。電子カルテには、病院内各部門における課題解決等の機能を有することが必須である。それらの問題を解決するためマルチベンダーによるシステムが構築されていることが多くある。このようにマルチベンダーにより構成されている電子カルテの医療情報データとハッシュ関数を用いた匿名化識別子によるレセプト情報との連結可能性を検証する必要がある。

（倫理面への配慮）

本分担研究内容については、関東病院 倫理・医療監査委員会の承認を得て電子カルテに関わる調査を実施した。(本分担研究班におけるレセプトデータと医療情報データベースとの匿名化の連結可能性とレセプト病名の妥当性に関する研究の検討において、倫理面として配慮すべき問題はデータのセキュリティの管理である。しかし、今回の研究の範囲内においては、個人の情報は取り扱わない。)

### C. 研究結果

本研究では方法論の検討として、ハッシュ関数を用いた匿名化識別子によるレセプト情報と電子カルテ等の医療情報との連結可能性を検証することになる。連結可能性について、電子カルテには、医科、傷病件数、医薬品、診療行為、材料等について突合を行うことのできるデータを有している。

関東病院における電子カルテを中心とした総合医療情報システムを示した(図1)。

### D. 考察

薬剤部門について、従来、処方オーダーシステムでは、情報のながれは医師から薬剤部に対し一方方向のながれであったといえる。患者に対しどのような薬剤が医師から処方されたかという情報は伝わるものの、どのような診療プロトコルに基づき医師が薬剤を選択したかという情報、薬剤の選択が適正であったかどうかという情報、その薬物療法の効果はどうであったかという情報は薬剤師には伝わらなかった。そのため、これらの情報を得て、薬剤師が医師らと検討する機会をもつことはほとんどみられなかったといえる。そのため薬剤師が診療プロトコルを閲覧し、診療記録情報の詳細をリアルタイムで入手できない場合には、薬剤師が有する情報を医師に提供する機会を逃す結果となることが多くみられた。電子カルテシステムの導入・稼働は、全ての医療スタッフが患者の診療記録情報を共有できる点で有用であると同時に保有するデータの信頼性、正確性は高まるものとなる。診療スタッフ間では患者情報が双方向に移動することなどから、情報の共有化が可能となる。

このように電子カルテに正確な情報が記載されることにより、レセプトデータとの詳細な突合が可能となる。また、平成22年度から始まった医療機関コードを提供する情報に含めた場合の問題については、病院等では個人情報とはならないものの、個人別では他の情報との

照合により、特定個人を識別できることから行政個人情報保護法との関係も整理されている。

レセプトデータの提供については、米国等で行われている行政機関、大学・大学院、医療の質を向上させることを目的とする公益法人、研究成果を一般に公表する学術研究等とする方向性の検討も必須となる。

従来、総理大臣の諮問機関である規制改革会議から「規制改革推進のための3か年計画(再改定)」が公表された。レセプトのオンライン請求義務化に関しては、レセプトのデータベースの構築と情報活用体制の整備が挙げられていた。

その内容は、

- レセプトオンライン化請求に合わせ、平成20年度末までにレセプトデータの収集・分析体制を構築し、平成23年度以降は全国のレセプトデータを収集・蓄積・活用できる体制を構築、運用する。
- レセプトオンライン化が効果を上げることができるよう、オンライン化を通じた医療情報・健康情報の収集・蓄積体制を早急に整備するとともに、例えば統計法(平成19年法律第53号)などの取り扱いを参考にしつつ、その恣意的な利用を防ぐとともに、国の独占利用を排除し、民間への開放を前提とした利用ルールを確立し、国民の健康の増進に資する幅広い分野での利用を図る。また、医療機関において蓄積された医療情報を共有化し、効率的かつ質の高い医療提供体制を構築するため、診療記録等の医療分野における電子化された情報については、外部保存を行うことができる環境整備を進める。等としたものであった。

レセプトデータを利活用することは、医薬品の使用についての実態を正確に把握するとともに、エビデンスを構築することができる。本研究において、病院のレセプトデータの連結を確認できたことは、薬害問題等の医薬品使用実態だけでなく、その患者の診療記録にまで遡ることが可能となる。従来は、このような体制が望まれながらも構築できていなかったことから、レセプトデータの紐付けができる体制の構築から、さらに詳細なデータとの連結を図ることの試みは非常に意義がある。

### E. 結論

本研究では、前年度の研究結果(平成18年

2月から平成20年11月までの34ヵ月分のレセプトを対象として調査を実施)より、医薬品については完全一致したものの、医薬品点数、回数については使用したレセプトデータの誤入力等による不一致がみられた。処方オーダーリングシステム等の正確なデータを有することが可能な電子カルテデータの利用は、匿名化したデータについて連結できる可能性について実証できることを明らかとした。

レセプトデータを利活用することは、医薬品の使用についての実態を正確に把握するとともに、エビデンスを構築することができる。本研究において、病院のレセプトデータの連結を確認できたことは、薬害問題等の医薬品使用実態だけでなく、その患者の診療記録にまで遡ることが可能となる。

電子カルテは、患者情報の一元管理による重複入力を排除し、情報を共有してチーム医療に貢献することができる。薬剤の重複投与・誤投与を防止し、抗がん剤インターバルのチェック等、各種チェック機能を実現する。さらに、情報のコード化等により、臨床データの収集・分析と診断へのフィードバックを容易にする。整理された経過記録を実現し、POS (Problem Oriented System) に基づいた SOAP (Subject, Object, Assessment, Plan) の記述方式を採用し、標準治療計画 (クリティカルパス) の導入により、患者中心の医療を推進する。最終的には、医療の質の維持向上等を実現する。

本研究結果を踏まえ、今後は、匿名化されたレセプトデータと電子カルテデータとのより詳細なデータの連結可能性までを視野に入れた研究を継続し実証することが必須である。

#### F. 健康危険情報

なし

#### G. 研究発表

##### 1. 論文発表

なし

##### 2. 学会発表

- 折井孝男. 医療データベースの二次利用による薬剤疫学研究の今後の課題: 第30回医療情報学連合大会(第11回日本医療情報学会学術大会) 浜松 2010 抄録集.

#### H. 知的財産権の出願・登録状況

##### 1. 特許取得

なし

##### 2. 実用新案登録

なし  
3. その他  
なし

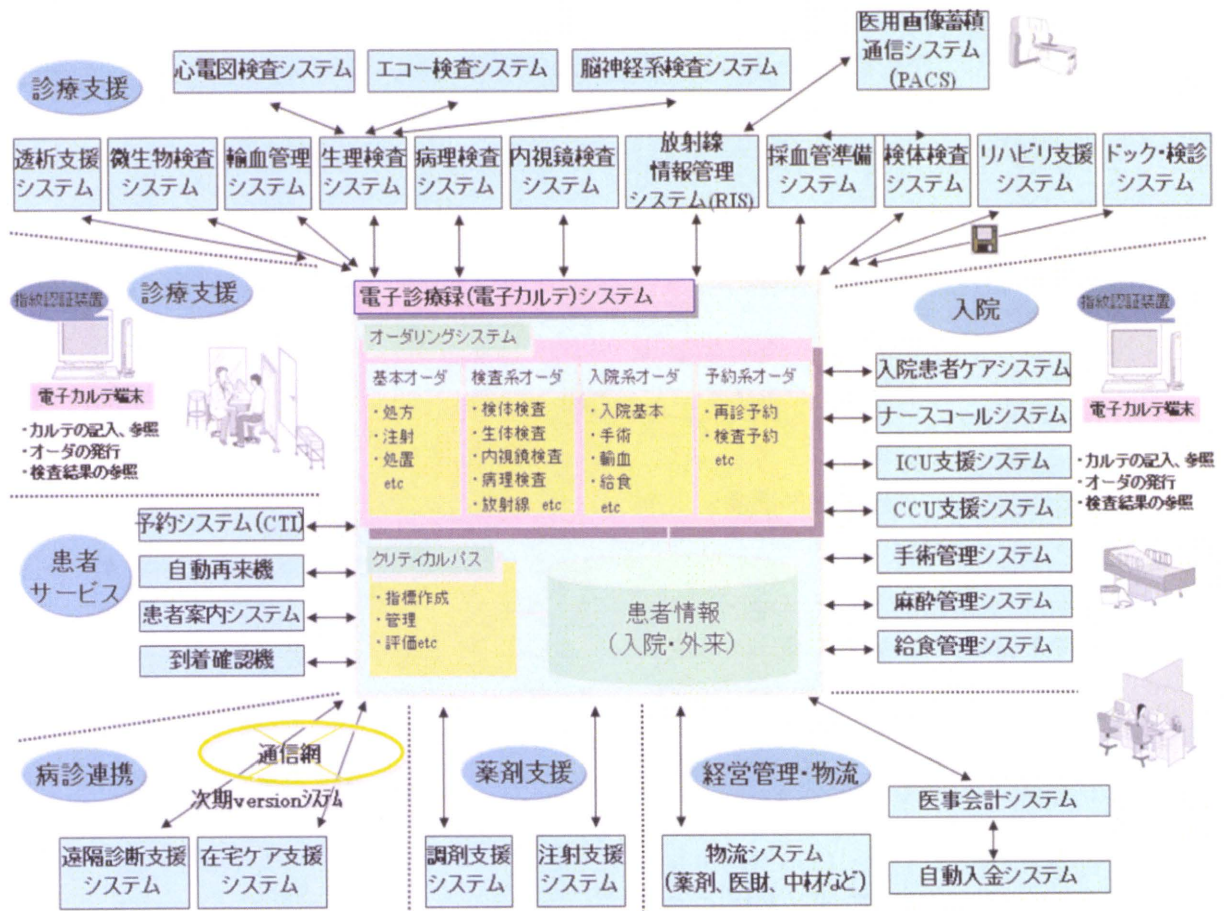


図1 電子カルテを中心とした病院情報システム

平成 22 年度厚生労働科学研究費補助金  
(医薬品・医療機器等レギュラトリーサイエンス総合研究事業)  
分担研究報告書

国外におけるレセプトデータベースの薬剤疫学的活用に必要な  
分析手法の状況とわが国への適用可能性

研究分担者 岡本 悦司 国立保健医療科学院経営科学部・経営管理室長

研究要旨：

わが国においてもレセプト情報・特定健診等情報データベース(以下、NDB)の研究利用が 2011 年度より可能となることより、今後はどのようなデータが使えるか、から、「どのように」活用するか、という分析手法に関心に移る。EBM では無作為割付による比較対照試験(RCT)が重視されその分析手法はかなり発展しているが、NDB のような既存データを用いた分析では、その分析手法は発展段階であり、中には RCT の手法をそのまま適用する誤用も見受けられる。諸外国や薬剤疫学分野において近年重視されている、とくにプロペンシティスコア(傾向得点)によるマッチング法を中心に、その研究状況を調査し、わが国 NDB の分析プログラムに MatchIT 等の適切なマッチングソフトを導入することが必要と結論した。

A. 研究目的

レセプトや電子カルテ(EHR)といった既存データを大規模に収集し、医療の質向上や薬剤安全対策に用いることは国際的にも主流になりつつある。振り返るに 1991 年頃より流行となった EBM(根拠に基づく医療)運動は、無作為割付比較試験(RCT)を最高のエビデンスと位置づけ「RCT でなければエビデンスにあらざ」という風潮さえ生れた。

しかし RCT の倫理性や限界についても批判があり、また 1983 年には Rosenbaum & Rubin がプロペンシティスコア(傾向得点、PS)法という既存データを用いた疑似的 RCT とも呼ぶべき手法を提案した。1994 年米国会計検査院(GAO)は乳癌に対する乳房全摘術と温存手術の優劣比較ががん登録データに PS を適用することにより RCT と同一の結果がえられることを示し、RCT に頼った結果、乳

房温存療法でも生存率に差が無いことを証明するために、介入群に割付けられた 2000 人以上の女性が失わなくてもよい乳房を失った、点を問題視した。

とりわけ医薬品安全性監視(pharmacovigilance、PV)のような RCT になじまない分野においては PS の有効性が次第に認識されるようになり、薬剤疫学の学会では大規模データベースと、それにいかにプロペンシティスコア法を適用するかという分析手法に関心が高まっている。と同時にこれまでに実施されたいくつかの分析には方法論の誤りもあることが明らかにされた。

本論ではわが国の NDB に PS を適用することを念頭に、PS 法の最新研究状況を国内外より収集し要約するとともに、具体的な適用方法を提示する。



## B. 研究方法

データベースへの PS 法の適用の手法に関して、薬剤疫学の国際学会、米国公衆衛生学会、等での発表内容を通じて PS 法の分析手法に関する文献を収集分析し、ミネソタ大学研究データ支援センター(ResDAC)の訪問調査により処方薬データの収集と研究者の利用を支援する体制について調査した。

## C. 研究結果

### 1. PS(傾向得点)とは

RCT において無作為割付が行なわれるのは介入を受けるか受けないかの確率を揃えるためである。観察研究では、介入(たとえば特定の薬剤の投与)を受けた群とそうでない群は受ける確率が同一ではないのでそのままでは比較できない。もし介入群と、介入を受ける確率が同一の対照群を選択できればそれはあたかも疑似 RCT とみなすことができる。Rosenbaum & Rubin は両群の様々な変数(たとえば生活習慣や健診データ)を説明変数として介入(たとえばある薬剤)を受けるか受けないかを目的変数としてロジスティック回帰分析を行い、介入を受ける確率を算出し、その確率をプロペンシティブスコア(傾向得点)と呼び、同一の PS の者同士で比較することで観察研究でも疑似 RCT が可能と提案した。

### 2. 観察研究と実験研究の相違

データベースは観察研究であって RCT のような実験研究ではない。それゆえ、因果関係を分析する手法も自ずと異なる。ところがこうした認識が欠けており、また EBM と RCT 万能の風潮もあって観察研究に実験研究の手法を誤用する危険が指摘されている。

たとえば RCT で介入群と対照群に割り付けると両群が類似(balance)していることを両群間の差の T 検定を行い有意差が無い、こと

で示すのが常道であり、既存データを用いた研究論文でも踏襲(stopping rule)されてきた。

しかし Imai らによると4つの理由からこの方法は正しくない、という。1)標本数を小さくすれば両群の差がどんなに大きくても p 値は大きくなり有意差無となる、2)標本数だけでなく標本分散によっても左右される、3)有意差検定は標本ではなく母集団に差があるかないかの検定だが知りたいのは標本そのものに差があるかないかである、4)両群が同じかどうかは絶対差の大小だけで決まり有意差検定など無意味・・・。

それゆえ、両群が同じかどうかの評価には、1)母集団ではなく標本そのものの差を評価すること、2)判断は標本数によって左右されないこと、の2点が重要という。Cochran は単純に両群の平均の差は標準偏差の4分の1を超えないこと、と提案した(1968)。別の方法は分位数(Quantile-quantile、QQ)プロットがある。Flury らは標準差(standardized difference)を提唱した。標準差とは両群の算術平均の絶対差を両群の分散を足して2で割ったものの平方根で除したもので Imai らの提案する2条件を満たしている。

### 3. PS マッチングの手法(MatchIT)

ある病気である薬剤の投与を受けた介入群を特定したら、次に同じ病気だがその薬剤を投与されていない対照群を抽出(マッチ)しなければならない。

介入群と対照群が一对一になるようにするためには「置き換え無し(without replacement)」とするのが原則である(図1)。

各患者について算出される PS 値がピッタリ一致する対照群を選ぶのが厳密(exact)マッチングであり理想であるが小数点以下までピッタリ合う例は少ないという問題がある。

そのため粗雑化された厳密マッチング (coarsed exact matching) という手法もある (caliper matching と呼ばれる)。

最近隣(next neighbor)法とは両群の PS 値の総当たり組を作り差の小さいものから順に縁組してゆくという方法(貪欲 greedy マッチングとも呼ばれる)。

最適法とは、最近隣法とは異なり、全体としての両群の差を最小化させる方法である。最適法も最近隣法も大体似た組合せとなるが最適法の方がやや優れている、とされる。

この他、1 対多の組合せとする完全(full)マッチングや遺伝子アルゴリズムを応用した遺伝マッチングというものもある。

Imai らはフリーウェアとして提供される R と呼ばれる統計ソフト上で稼働する MatchIT と呼ばれるソフトをフリーウェアとして提供している。こうした高度なマッチングを容易にできるプログラムとして NDB で薬剤疫学研究を行うにも有効と考えられる (図 2)。

#### 4. マッチトペアの評価手法

一対一マッチングされた介入群と対照群は「対(ペア)」をなす。それゆえ対のある分析を行わなければならない、介入群と対照群を群として群間比較をしてはならない。具体的には T 検定なら対のある T 検定、あるいはウィルコクソンのサイン付きランク検定等を用いる。また対ごとに層別化されたコックス比例ハザードモデルも使える。

Austin が 1996 年から 2003 年に公表された PS 法を用いた 47 論文を精査したところ適切な手法を用いていたのは 13 論文(28%)しかなかった。よくある誤った評価手法は、対のある標本なのに Kaplan-Meier 生存カーブにログランク検定を実施したり、Cox 回帰、ロジスティック回帰、カイ二乗検定、Wilcoxon ランクサム検定そして対の無い T 検定を実施する

こと、であった。

#### 5. 研究データ支援センター(ResDAC)

既に見たように、仮説検証型の「研究」を実施することは、単に特定の薬剤を処方された患者を抽出するだけでなく、PS 法等適切なマッチングによる対照群の選定が不可欠であり、こうした作業は単純なデータ抽出ではなく申請する研究者と共同して作業を行う必要がでてくる。

米国のメディケアを運営する CMS は、研究者の便宜を図るためミネソタ大学に委託して研究データ支援センター(ResDAC)を設置し、ワークショップやアウトリーチを通じて研究利用を促進している。高齢者対象のメディケアは処方薬は給付対象でなかったが 2007 年より対象となり(パート D)、薬剤疫学研究への利用可能性が一気に拡大した。それでも高齢者のみであり、全人口をカバーするわが国 NDBの方がはるかに有利である。

ResDAC は McBean 所長以下、7 人の技術職員、4 人の研究員を有し、アメリカ公衆衛生学会等で展示ブースを出したりワークショップを開催したりして、研究者に利用できるデータ内容、手続きそして統計的分析手法について助言を実施している。

#### D. 考察

近年その方法論が急速に発展した PS マッチングは NDB のような既存データを薬剤疫学的研究に活用する上できわめて有効な手法になると期待された。

わが国の NDB は、レセプトデータだけでなく 40 歳以上の相当数には特定健康診査・保健指導のデータ(検査値、体重そして質問票)が含まれており、これらは PS を算出するためのロジスティック回帰分析の説明変数として適切である。レセプトのナショナルデータベースを有する国は韓国、台湾等多数あ

るが、これはわが国 NDB の極めて有利な条件といえる。

## E. 結論

NDB の研究利用が 2011 年度早々に開始されるが、オーダーメイド集計等単なる特別集計に止まらない仮説検証的な「研究」目的に活用するためには、単にある薬剤の投与者を抽出するだけでなく有害事象の発生頻度を比較する対照群を選択する必要がでてくる。

NDB はレセプトだけでなく特定健診のデータや質問票等データを含んでおり、それらを説明変数として PS マッチングを行うことにより国際的にも貴重な薬剤疫学研究が可能となる。

NDB の研究利用申請を審査するための有識者会議が既に立ち上げられ、来年度からの申請受付に向けて準備が進められているが、今回明らかにしたような高度な分析手法も行える体制を整える必要がある。

具体的には、研究者からの「このようなデータが欲しい」という申請によりデータを抽出するだけでなく研究者と共同して適切なマッチング法を行うことを助ける米国の研究データ支援センター(ResDAC)のようなサポート機能が必要である。新たな組織や人員増が困難な中、国立保健医療科学院のような厚生労働省の研修機関を活用することも検討に値しよう。

## F. 健康危険情報

なし

## G. 研究発表

### 1. 論文発表

- 岡本悦司. 医薬品安全対策とナショナルデータベース. JAPIC news319 号 2~3(2010年11月)

### 2. 学会発表

- Okamoto E. Application of propensity score matching on health insurance claims using health-screening data for metabolic syndrome as covariates. The 5<sup>th</sup> Asian Conference on Pharmacoepidemiology; 2010 October, Tokyo.

## H. 知的財産権の出願・登録状況

### 1. 特許取得

なし

### 2. 実用新案登録

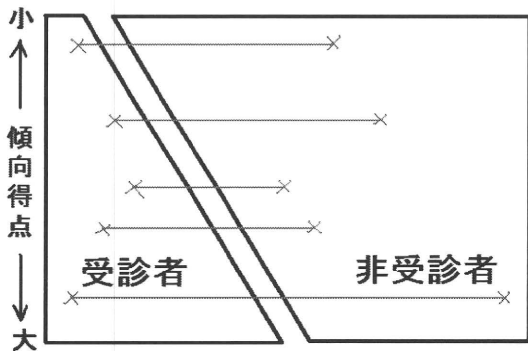
なし

### 3. その他

なし

## I. 参考文献

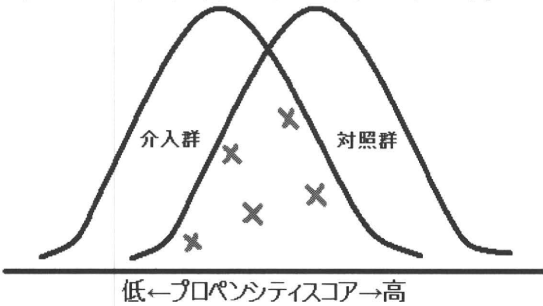
- Rosenbaum PR. Optimal matching for observational studies. *J Am Stat Assoc* 1989;84(408):1024-32.
- Imai K, et al. Misunderstandings between experimentalists and observationalists about causal inference. *J.R.Statist.Soc.A*(2008)171, Part2, pp.481-502.
- Ho DE, et al. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* (2007)15:199-236.
- Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine* 2008;27:2037-49.
- Ho DE, et al. MatchIT: Nonparametric preprocessing for parametric causal inference. August 12, 2010.(MatchIT マニュアル) [<http://gking.harvard.edu/matchit/>]



方法2 マッチング法

傾向得点が同一(又は極めて近い)受診者と非受診者を一対一で対応させる。欠点・・・相手を見つけられず、分析に使えない受診者や非受診者がでてる

### プロペンシテスコアマッチングの原理



プロペンシテスコアマッチングができない例  
予測があまり良過ぎて重なりがない(相手がいない)

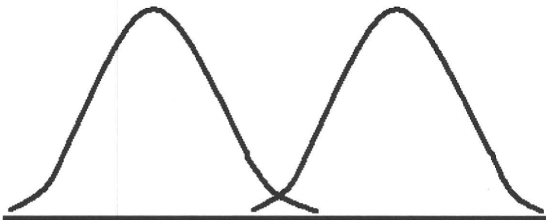
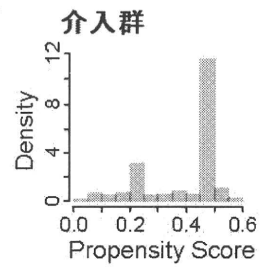
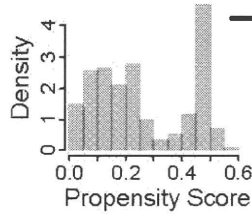


図1 プロペンシテスコアマッチングの原理

MatchITによるマッチングの結果  
対照群よりPS値が一致するものだけが抽出されている。



元の対照群



マッチされた対照群

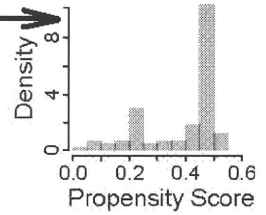


図2 MatchIT によるマッチングの原理