



〔総説〕 世界最大級のレセプトデータ（NDB）を用いた 臨床研究への誘い

高 林 克日己¹⁾ 安 藤 文 彦²⁾ 土 井 俊 祐³⁾ 鈴 木 隆 弘²⁾

(2024年4月20日受付, 2025年2月10日受理, 2025年6月10日公表)

要 旨

The National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB) は日本国中の電子化レセプトの集積を厚労省がデータベースとしたもので国内の研究者であれば申請して利用することができる。NDBはわが国特有の国民皆保険制度と国内全域がほぼ均一な医療を行っている結果として、1億2千万人、2009年から13年間で約250億枚のレセプトデータを包含している世界最大規模の悉皆性をもった医療データベースである。NDBデータの内容は限定的であるが、これをうまく利用することで記述研究だけでなく臨床研究に匹敵する研究ができることを自験例をもとに本稿で紹介する。10年間にわが国で生物学的製剤およびJak阻害薬を用いた関節リウマチ患者をすべて抽出し、これら分子標的薬8薬剤の中で継続率を計算して有効性を比較するとともに、安全性については日和見感染症の発生頻度を比較した。また薬剤開始時期に季節変動を認めることを発見し、膨大な量を俯瞰的に検討することで通常は気が付かないような知見も発見できることを示した。今後多くのNDB研究が本学から発表されることを希望する。

Key words: NDB, claims data, real world data, rheumatoid arthritis

I. 緒 言

ビッグデータ, Real World Dataという言葉が一般化してきたが、実際にこれに関与したその恩恵を受けている医学研究者はまだ僅かである。しかし多くの医学研究者が臨床治験や基礎動物実験を行っているのと同様に、ごく近い将来

ビックデータの解析は医学研究の中で一つの重要な手法になるだろう。それは圧倒的な大きな数をもって俯瞰的に眺めるという点ではRandomized Controlled Trial (RCT) とは相反する世界である。RCTの数十あるいは数百名の精緻なデータの収集と比較解析から出てきた統計学的有意な結果に対して、中には信頼のできないデータも含ま

¹⁾ 三和病院, 千葉大学名誉教授

²⁾ 千葉大学医学部附属病院企画情報部

³⁾ 千葉大学医学部附属病院病院長企画室

Katsuhiko Takabayashi¹⁾, Fumihiko Ando²⁾, Shunsuke Doi³⁾, and Takahiro Suzuki²⁾. Invitation to clinical research using the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB).

¹⁾ Sanwa Hospital, Matsudo 270-2253. Professor Emeritus, Department of Management and Medical Informatics, Chiba University Hospital, Chiba 260-8677.

²⁾ Department of Management and Medical Informatics, Chiba University Hospital, Chiba 260-8677.

³⁾ Director Planning Office, Chiba University Hospital, Chiba 260-8677.

Phone: 047-712-0202. Fax: 047-712-0201. E-mail: takaba@kanai.or.jp

Received April 20, 2024, Accepted February 10, 2025, Published June 10, 2025.

れてはいるが何十万、何百万という数で研究する方法である。今回こうしたデータ、とくにthe National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB)の利活用を是非千葉大学の関係者にも知っていただきたいと考えて本稿を寄稿した。

ここでは特にわれわれが進めてきたNDBを用いた臨床研究を例にして述べる。厚労省が保有するわが国の電子レセプトデータの集積であるNDBの利用が従来の一部の研究施設から一般研究者にまで開放されるようになり、申請が認可されれば誰でも自由に活用できる時代になった(図1)(厚労省ホームページ; NDBの利用を検討している方へのマニュアル<https://www.mhlw.go.jp/content/12400000/001374751.pdf>; 2025年2月1日現在)。NDBは世界最大規模の電子レセプトデータベースであり、わが国特有の国民皆保険制度のお蔭で1億2千万余のほぼ全国民の匿名化された健康情報が13年間以上の長期間について悉皆性をもって活用できる点は特筆すべきものがある。その内容として2023年5月現在で約250億のレセプトデータを擁している(厚労省ホームページ; NDBデータの利活用の現状, 2023/11/13 医療等情報の二次利用に関するワーキンググループ<https://www.mhlw.go.jp/content/10808000/001166477.pdf>; 2025年2月1日現在)。この世界に誇るべき我が国の貴重な医

療ビッグデータベースはただ単に記述研究に利用するだけでなく、その特性をよく理解していれば様々な因果推論を行う臨床研究に応用することも不可能ではない。しかしこのことがよく理解されずに未だ十分には臨床研究に活用されているとは言えない状態にある。もしNDBで臨床研究が可能であるということが周知されれば、NDBを利用したいという研究者はあらゆる臨床分野に存在すると思われる。われわれはNDBに先立ち13国立大学病院のレセプトデータ解析(CISA研究)からレセプトデータを活用した臨床研究を進めてきた[1]。検査値もないレセプトデータで何ができるかと疑問をもたれるかもしれないが、ちょっと考えていただきたい。例えばある薬剤Aを投与中に特別な感染症Bに使う薬剤Cが併用されたとすれば、それはその薬剤Aにより合併症Bを起こしたことを意味する。整形外科で大腿骨置換術を行ったあとに深部静脈血栓症をおこした症例は、整形外科の術式、術後のヘパリンなどの抗凝固薬の投薬を組み合わせで症例を抽出し、さらに両者の発生日により手術から血栓症までの発症期間を計算することができる[2]。このように匿名化され分散したレセプトを個々の症例に名寄せしたうえで、複数の事象の存在をレセプトデータから時系列で探してくることで、単に記述のみでなく因果推論の研究もできる。本稿では臨床研究者にNDBに関心をもっていただくことを主眼として、自験例をも

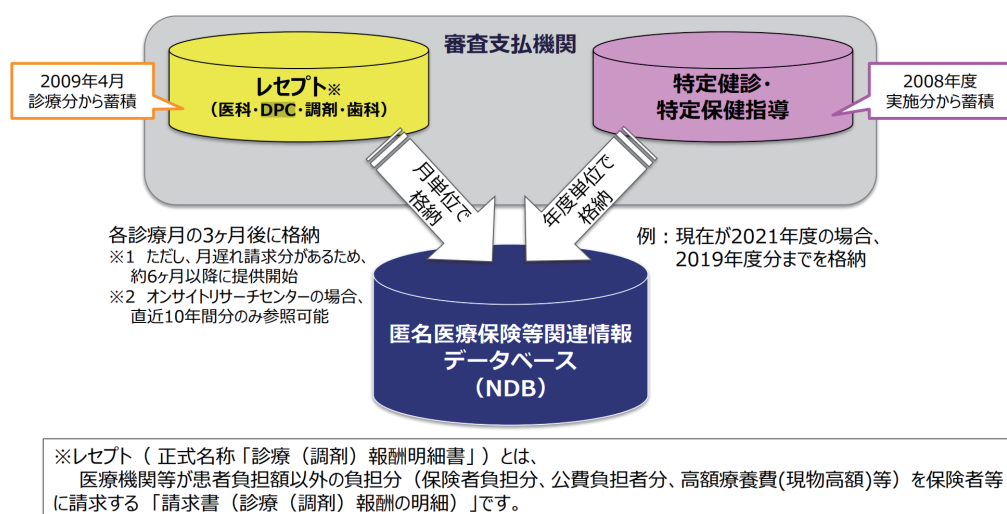


図1 NDBに含まれる情報 (厚労省ホームページより一部改変)

2009年度から2022年度まで248億1,200万枚、さらに特定検診データ、死亡データも取り込まれるようになっている。

とにどのような臨床研究が可能か、またどのような課題があるのかを紹介することとする。

II. NDBデータについて

1. NDBデータの特徴

NDBデータは厚労省に集まる電子レセプトデータの集合である。レセプトデータは単なる会計のための情報であって、当然ながら研究目的を意図して作られたものではない。そこにあるのは病名と検査項目、処方薬剤名、手術の術式などに留まる（最近ではDPC（Diagnosis and Procedure Complex）データの一部も包含されるようになった）。レセプトの病名はいわゆる保険病名としてその信頼度が高いとはいえないことはよく知られた事実である。また検査結果やその転帰とその事由についても記載はない。こうしたことから臨床研究以前に記述研究をするうえでも当然制約がある。さらに臨床研究となるとRCTとは異なり、検査結果など必要なデータが揃っているわけではなく、揃えるべきベースラインデータも不足し、転帰も不明瞭である。このように臨床研究に利用するには欠陥だらけのデータの集合にみえるが、RCT全盛の時代において実はRCTでは施行することのできない領域で補完的役割を果たすことが期待できる部分も存在する。RCTといっても万能ではない。例えばある薬剤の長期間（10年以上）投与期間における合併症の研究、多数の同効薬剤間のhead-to-headの比較、対象数が得られにくい希少疾患の解析、治験を行うことが難しい超高齢者を対象とした研究などはRCTの実施が困

難である。こうした分野においてはNDBによる解析に期待できることがある。またRCTで行ったことを、より大きな数で検証することにも用いることができる。

NDBを用いた研究の最大の特徴は言うまでもなくそのデータ数が十分に大きいことである。いままで動物実験や治験に慣れ親しんできた研究者にとっては愕然とするくらい研究手法において大きな差異がある。多くの医学研究者が医学統計学の基本として必要条件にしてきた χ 二乗テスト、p値なるものが、NDBの世界ではもはや意味を持たなくなる。NDBは抽出した集団ではなく、日本国民全てを対象とした母集団そのものであり、そこで見える差異がそのまま有意差をもつことになる。一方でこれは外れ値、脱落例、バイアス、交絡因子そうしたものが全て混在したデータの集積でもある。そのままではただのごみの山にしか見えないものの中から、真理を探し出すのはまさに臨床研究者の腕のみせどころでもある。

さらにこの国独特の国民皆保険制度のお蔭で、世界の中で1億人を超える巨大規模でしかも悉皆性のある、均一データを有する国は他に存在しないといってよく、まさに日本が独壇場になれる圧倒的に優位な状況にある。ここではNDBデータを用いて関節リウマチにおける分子標的薬間の比較についてのわれわれの研究を例に解説する。

2. NDB研究の種類

NDBを用いた研究としては、厚労省が既に公表しているオープンデータを用いるだけでなく、第三者提供として特定のデータを独自に依頼し、

表1 NDB第三者提供データの提供形式とその特徴

	特別抽出	サンプリングデータセット	集計表情報
内 容	研究者が必要なすべての対象患者の中の必要なデータ	外来の1%あるいは入院10%のレセプト1か月分	データを加工した集計表
提供データ	個票	個票	集計表
問 題 点	申請書の作成準備が膨大で、さらに申請から抽出までに1年前後の時間がかかる	連続したデータの解析が困難	自分でさらなる解析はできない
対象研究者	レセプト研究に一定の知見を有し、大量のデータに対して抽出条件を吟味して解析できる研究者	レセプト研究に関心を持っている初心者やある程度の概要を把握したいと考えている研究者	集計データの結果を利用するのみでデータ処理は考えていない研究者

厚労省のホームページ（厚労省のホームページを一部改変）

これを用いて行う研究がある。これはさらに特別抽出と呼ばれる独自の研究に必要なデータを申請してそれを受理して行う研究、一部のサンプリングデータを用いた研究、集計表の作成を依頼する方法に分かれる（表1）。サンプリングデータセットとは全体の1%（外来患者）あるいは10%（入院患者）のデータを1か月分のみ申請して受理する方法である[3]。したがって1か月以上連続する時系列データの解析はできない。第三者提供ではさらにデータを直接受理して自らの研究室を隔離されたスペースとしてそこで研究する方法と、オンサイトセンターに出向いてそこでサーバにアクセスして研究する方法に分かれる。オンサイトセンターは東大と京大および厚労省にあり、ここに出向いて自分に与えられた時間にサーバに直接アクセスする方法である[4]。サーバにアクセスできる部屋には自分のパソコンなどは持ち込むことはできず、情報の持ち出しはできない。これらに対して自施設で行う場合にはあらかじめ自分で設置した特別に隔離された安全な場所での研究が求められる。われわれは特別抽出でデータを受理し、千葉大学内に専用の解析室を設置してデータ解析を行った。

Ⅲ. 方 法

1. リウマチ分子標的薬に関する研究

われわれのNDBを用いた研究は、関節リウマチ（Rheumatoid Arthritis; 以下RA）の分子標的薬においてその有効性と安全性の比較を一般の臨床研究に近い手法で検討することにある[5,6]。ここではRAに用いられる7種類の生物学的製剤アバタセプト（ABT）、アダリムマブ（ADA）、セルトリズマブペゴル（CZP）、エタネルセプト（ETN）、ゴリムマブ（GLM）、インフリキシマブ（IFX）、トシリズマブ（TCZ）と1種類のJak阻害薬（Jaki）トファシチニブ（TOF）、計8種類の分子標的薬を対象として相互間を比較する研究を試みた。

RAの分子標的薬は2000年ごろからRAとその関連疾患に使用が開始され、RAの予後を大きく変えた薬剤であるが、すべての例に効果があるわけではなく、また薬物により効果も異なる。保険

適用されてはいても高価な薬剤であり、また長期間にわたっての投与が必要で患者に大きな経済的負担がかかるので、効果がなければ継続することはできない。このことから薬剤の継続性を有効性の指標とみなすことができると考えた。2017年当時RAにおける分子標的薬は先発薬品として生物学的製剤7種類またJak阻害薬1種類があったが、これらの相互の多剤間比較の解析はRCTでは行われていなかった。有効性については分子標的薬の継続性をみることで、安全性については治療後に発生する日和見感染症の発症率を比較することで検討を行うことにした。

2. 方法とプロセス

2. 1. 申請から受理まで

特別抽出での研究を行うためには厚労省に詳細な研究計画の申請し、また学内の倫理審査を受けなければならない。本研究は介入試験ではなく、また匿名化に関して個人情報はそもそも除去されているが、希少疾患などで症例数が10未満になる場合などは本人を同定することができる可能性を考えて発表ができない。また申請書には必要なコードの一覧（疾患名、薬品名、検査、処置など）を漏れなく記載しなければならない。NDBの申請ルールとして、すべてのデータを要求することは許されず、必要最低限のデータ項目に絞った申請が求められる。しかし主要解析に用いるデータについてはもちろんであるが、論文投稿時の査読等でさらに別項目が要求されることも想定し、関連するデータをあらかじめもれなく取得しておくことも必要である。このようにある程度拡張したデータでカバーしておかないと、わずかな項目の不足が研究の続行を不可能にすることがある。再申請はできるが、データの不足からの再申請はさらに長時間をかけることになる。また申請時には最終的に出来上がると想定される図表も提出する必要がある。こうした意味から考えると現時点では特別抽出において探索的研究は原則認められないことになる。

われわれは2010年4月から2017年3月までのNDBのデータから8種類の分子標的薬を用いたRA症例を抽出するのであるが、レセプト病名からのRA患者の抽出は不正確であるのに対し、分

表2 薬剤の組み合わせによる日和見感染疾患の定義

OI		DRUG					Patients
PCP	ST	PI					
	+						572
		+					103
TB	RFP	EB	INH	SM	PZA		
	+		+				602
	+	+	+		+		527
	+	+	+				518
	+	+					338
		+	+				59
	+		+		+		41
	+	+		+			41
	+	+	+	+			31
	+	+	+	+	+		29
	+		+	+			27
		+	+		+		24
	+		+	+	+		16
	+	+			+		14
			+		+		14
		+	+	+			13
			+	+			12
		+		+			10
NTM	RFP	EB	INH	SM	PZA	CAM	
	+	+				+	656
	+					+	245
		+				+	133
	+		+				39
	+	+	+				34
	+	+					31
	+	+	+			+	26
	+	+		+		+	23
	+		+			+	13
HZ	VACV	FCV	ACV	VIDA			
	+						15,160
		+					8,731
			+				1,401
				+			415

日和見感染症患者抽出のための疾患別投与薬剤の組み合わせおよびその該当患者数を示す（文献[6]）。

OI: opportunistic infection, ST: sulfamethoxazole/trimethoprim, PI: pentamidine isetionate, RFP: rifampicin, EB: ethambutol, INH: isoniazid, SM: streptomycin, PZA: pyrazinamide, CAM: clarithromycin, VACV: valaciclovir, FAV: famciclovir, ACV: acyclovir, VIDA: vidarabine

子標的薬を使用してレセプト病名にRA名がある患者はより確度が高いと考えられることから、まず分子標的薬を使用している患者の全抽出を依頼し、その中からRAのレセプト病名のある症例を選ぶ方法とした。このような選択の順序を過つと

膨大な作業量が発生する可能性があるため、症例の抽出順序は十分に検討する必要がある（表3）。

申請書の作成に当たっては数百人時を要するともいわれている[7]。これをできるだけ簡素化するためには先人たちの意見や参考例を利用

表3 NDB研究を成功させるためのポイント

1	まずサンプリングデータセットで試してみる
2	コンピュータサイエンティスト・レセプトの専門家とのチームをつくる
3	時間と予算には十分な余裕を持つ
4	レセプト病名と特定の薬剤や検査の組み合わせで疾患を同定する
5	第一段階から解析数をできるだけ小さく絞れる手順を考える
6	解析に必要なとする項目以外に周辺の薬剤、病名などもデータ請求しておく

自験例などからNDB研究でのポイントをまとめた。

することが不可欠である。また一旦申請書を提出しても、内容の吟味で厚労省側との質疑が繰り返されるため、提出後にも相当のやり取りがあることは覚悟しなければならない。2024年現在申請してからデータ受理するまでの期間は平均300日を超えており（厚労省ホームページNDB利活用促進に向けた取組について、https://www8.cao.go.jp/kisei-kaikaku/kisei/meeting/wg/2210_03medical/221107/medical02_020103.pdf: 2025年2月1日現在）、われわれが2022年から請求中の別研究も申請からデータ入手まで1年以上がかかっている。

2. 2. 実行環境の整備

得られた情報を扱う場所については周囲から隔絶され第三者が立ち入ることのない鍵のかかる専従のスペースで、かつ入室者の記録がとれる空間を用意する必要がある。さらにデータの受領準備では、物理的な作業部屋の準備と作業を行う分析用端末の準備が必要となる。特に分析用端末の準備では、事前に受領するデータがどの程度のデータ容量となるか見積りができないことが問題である。処理するファイルの容量が受理するまではわからないため、準備する分析用端末のスペックや必要となる処理ソフトウェアの選択の手がかりがなく、事前に準備した環境がオーバー・スペックであったり、反対に処理能力が不足し、環境の構築し直しが必要となることがある。自験例では、細かな加工作業で発生する数分から数十分の待ち時間の発生が作業を早く進める上での一番の課題となり、その原因としてファイルの読み書き等の入出力処理が大きなボトルネックとなった。デー

タ容量の予測がつかない事前の準備においては、できるだけスペックの高いマシンを用意しておくに越したことはないが、限られたリソースの中では、読み書きの速度に注目した環境構築も重要である。

2. 3. 前処理・加工

2. 3. 1. 名寄せと連結

申請したデータは個人ごとではなくそれぞれ月別施設別の分散されたレセプトデータとして受領することになる。これを臨床研究に用いるには分散している個人のデータを寄せ集めた（名寄せ）うえで、個々の症例の時系列データとして並び替える必要がある。匿名化されているが被保険者番号あるいは姓名と生年月日などからハッシュ関数を用いて元の値に復号できない値（ハッシュ値）に変換されたID1とID2が用意されているので、この二つのIDを用いることで同一患者として追跡し時系列に並べていく[8,9]。ID1は、保険者番号、被保険者証の記号・番号、生年月日、性別を元に作成されたハッシュ値であり、ID2は氏名、生年月日、性別を元に作成されたハッシュ値である。どのIDを用いるかは、個々の研究において対象とする患者にあわせて検討する必要がある。今回の研究ではID1とID2を組み合わせたID0を用いて研究している。しかしそれでも同時に氏名と保険者が変更となる患者や性別変更が行われた患者などは追跡できておらず、100%追跡できるわけではない。死亡例についても2024年以降は死亡フラグがつくことで死亡の確認ができるようになったが、それ以前の死亡確認については直接のデータを得られないので前後のレセプトの状況などから死亡の有無を推定しなければならない[10]。

2. 3. 2. 症例の選別

名寄せをして各個人のデータを時系列的に並べた後、RAのレセプト病名がついている症例を約23万人抽出し、さらにリウマチ分子標的薬8薬剤で保険適用が許されている疾患名（例えば潰瘍性大腸炎）がレセプト病名にある症例は除外し、RAと考えられる症例に絞った。さらに月に数種の分子標的薬を投与するなどの外れ値症例を除去

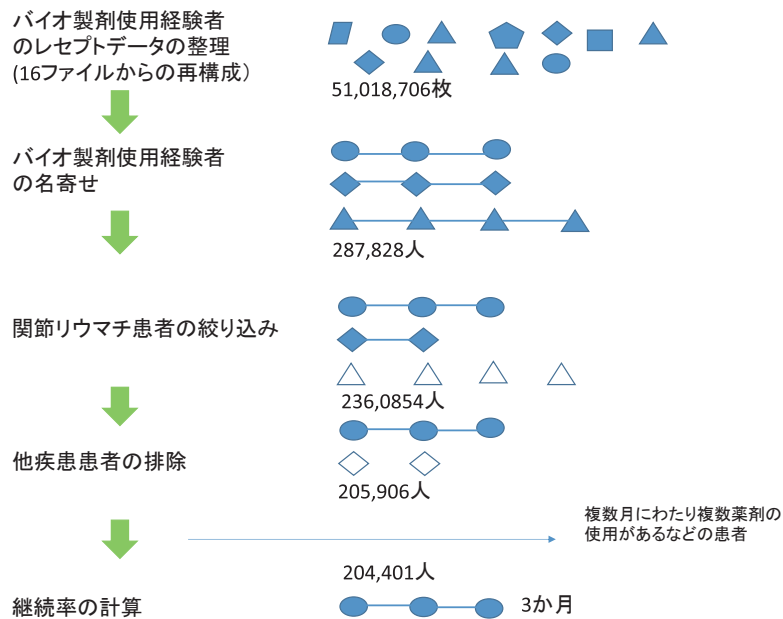


図2 分子標的薬使用関節リウマチ患者の抽出方法

厚労省から受領した51,018,706枚のレセプトをID0で追跡名寄せして月次の連続データに並び揃えた。さらにその中で関節リウマチのレセプト病名のあるものを選択し、また異常値のあるものを除外して最終的に204,401人とした。

した20万4千人の患者データを解析した(図2)。

継続率の観察期間としては2010年4月から2011年3月まで1年間を観察期間としてこの間にこれら薬剤が投与されていない症例をナীব症例として選んだ。継続の定義はその薬剤が連続する月のレセプト内に存在をみることで行うが、処方最大日数が90日であることから欠損月が出ることも考慮して、4か月以上処方のない例を中止症例と判断し、投与最期の月までを継続期間とした。またスイッチ例は前投薬剤が休止されてから4か月以内に別の分子標的薬が開始されたものとした。そして薬剤別の継続率を算出した。

日和見感染症の定義は、分子標的薬投与中にレセプト病名が出現し、(表2)にあるようなニューモシスチス肺炎(PCP)、結核(TB)、非結核性抗酸菌症(NTM)、带状疱疹(HZ)各疾患において使用される薬剤あるいは複数薬剤の組み合わせが投与されたものを日和見感染症と定義した[6]。

IV. 結 果

1. 分子標的薬使用者患者の記述的解析

月別使用患者数データを取ることで、2010年から2017年までの各薬剤の使用患者数の推移を見る

ことができる(図3)。各製薬会社はある月にどれだけ市販されたのかのデータは持っているが実際の患者数は把握できておらず、また他社のデータは持っていないためこうした全国の総使用患者の推移は今まで報告されていない。性別、年齢別使用患者数で各薬剤の特色を比較した。男女比はどの薬剤でも女性が80%前後であった。ABTとGLM、TOFが65-70歳と比較的高齢者に使われていることが示された(図4)。

またさらにこれを各月ごとの新規使用患者数をグラフにすると図5のようになり、分子標的薬の使用開始は春に増加し、秋から冬に減少する季節性の傾向が観察された[11]。これはまた各薬剤別、年齢別にみても同様な傾向があり、国内を地域別に分けても同様なことが示された。

2. 継続率

継続率はナীব群、スイッチ群どちらもTOF、TCZ、ABT、ETN、GLM、CZP、ADA、IFXの順になった(図6)。これらは従来報告されているものよりも除外症例がないことなどから低い数値となっている。関節リウマチにおいて1万を超える使用者数で相互比較した国内の論文はないが、関西地区で行ったレジストリー調査とは近い

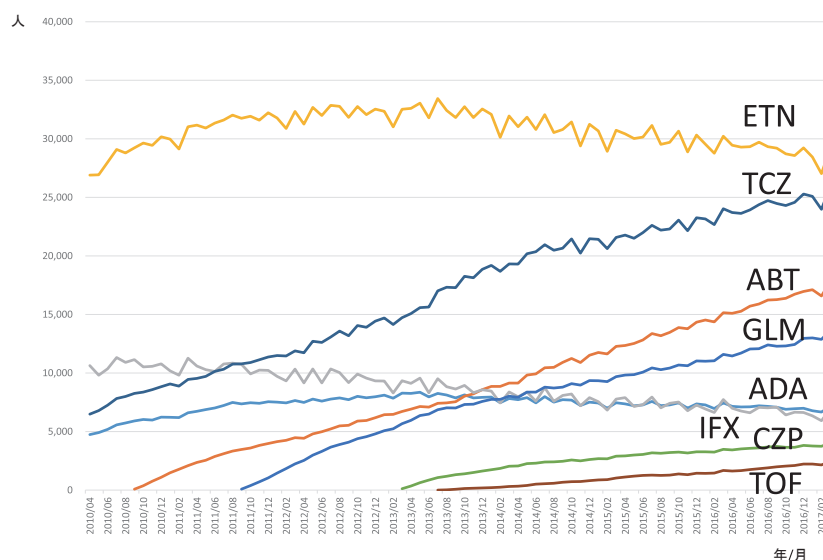


図3 各薬剤使用患者数の月別推移

ETN: etanercept, TCZ: tocilizumab, ABT: abatacept, GLM: golimumab, ADA: adalimumab, IFX: infliximab, CZP: certolizumab-pegol, TOF: tofacitinib

	ABT	ADA	CZP	ETN	GLM	INF	TCZ	TOF
総数	29,506	19,524	9,348	42,316	26,081	13,824	38,695	3,732
(naive)	17,649	14,249	4,797	36,891	15,198	12,001	19,321	1,182
(switch)	11,857	5,275	4,551	5,425	10,883	1,823	19,374	2,550
年齢（中央階級）	65～69	60～64	60～64	60～64	65～69	55～59	60～64	65～69

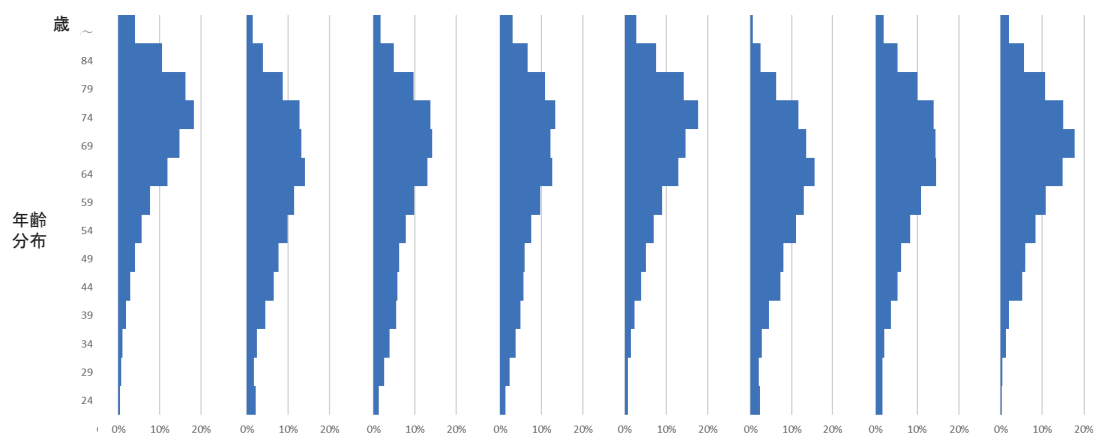


図4 各薬剤使用患者の背景

NDBでは年齢は5歳刻みでないと開示できないのでこのような表現になる。～24歳は24歳以下をまとめた。
 ABT: abatacept, ADA: adalimumab, CZP: certolizumab-pegol, ETN: etanercept, GLM: golimumab, IFX: infliximab, TCZ: tocilizumab, TOF: tofacitinib

データになっている[12]。

またナイーブから他の薬剤にスイッチしたときに継続率の高い薬剤を比較したが、どの薬剤においてもほぼ同様の傾向となった。さらにA、Bの薬剤をナイーブとして使い、それぞれを入れ替えたスイッチ例でのA、Bの継続率を比較すること

(図7)で、擬似クロスオーバー試験としてデータを取ることでhead-to headの比較ができる。ここではTCZと他剤の比較を図8に示す。

3. 安全性

安全性はここでは4つの日和見感染症の発症

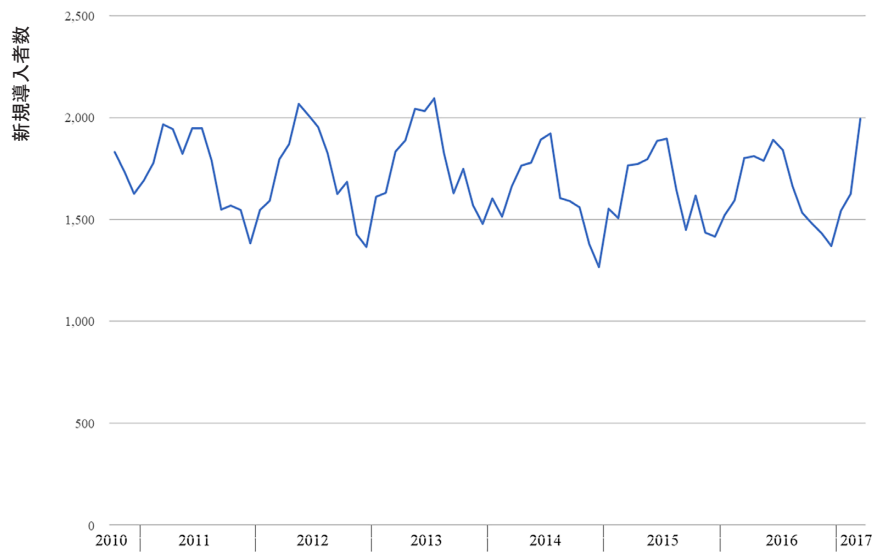


図5 新規分子標的薬導入者数の季節変動

分子標的薬の新規導入者の数を月別でプロットすると春夏に高く秋冬に低いパターンが見えてくる。これは関節リウマチが増悪する季節性変化と呼応するように見える。

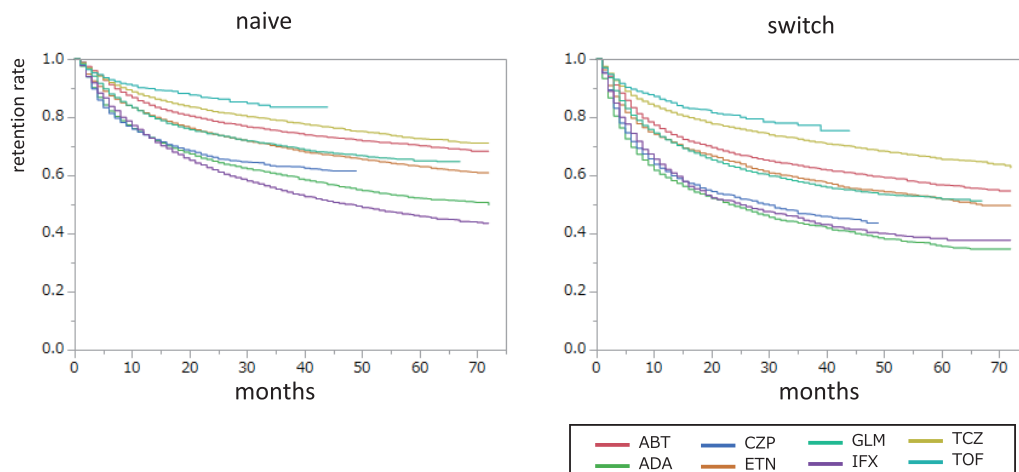


図6 薬剤別継続率 (naive/switch別)

ナイブ、スイッチどちらもトファシチニブ (TOF) が最も高く、トシリズマブ (TCZ)、アバタセプト (ABT) がこれに続いている。

ABT: abatacept, ADA: adalimumab, CZP: certolizumab-pegol, ETN: etanercept, GLM: golimumab, IFX: infliximab, TCZ: tocilizumab, TOF: tofacitinib

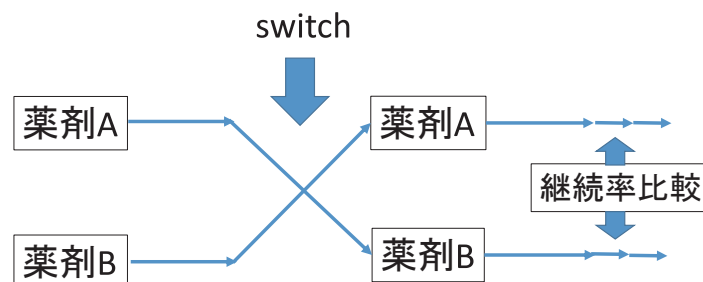


図7 Switch例を利用した疑似Cross Over Test

スイッチ例はそれ以前のナイブ例が無効なものを切り替えたものであり、その継続率を相互で比較することで擬似クロスオーバー試験として相互の有効性の比較を行った。

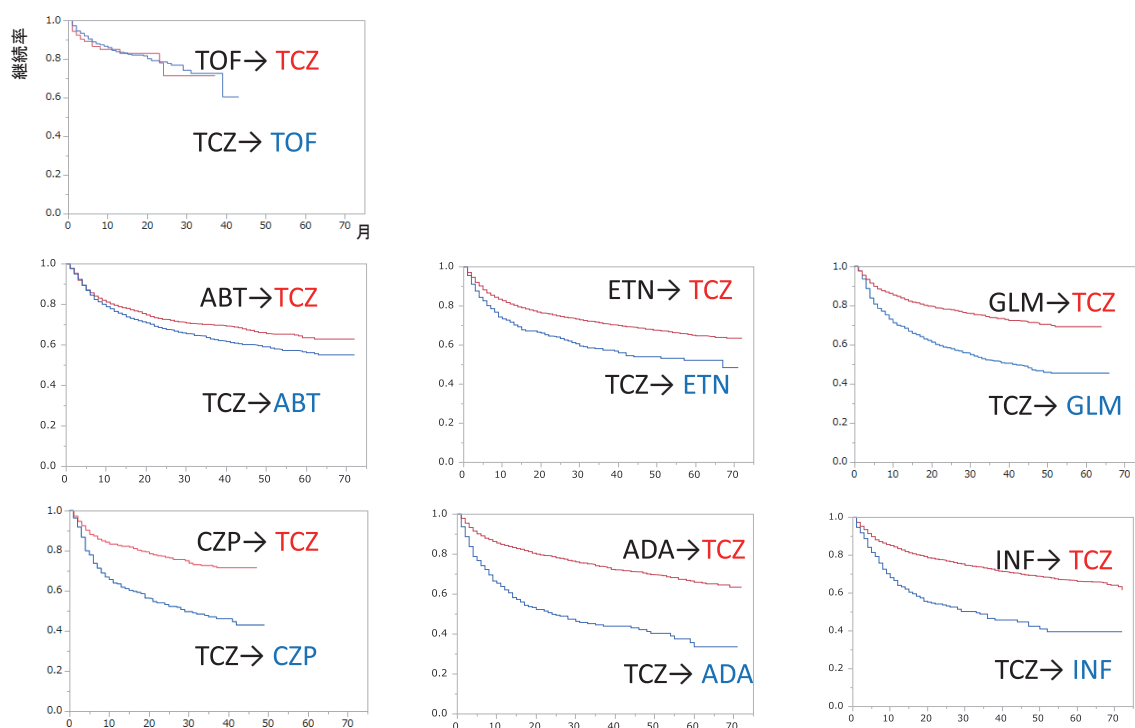


図8 Switch例を利用したCross Over継続率の比較

トシリズマブ（TCZ）を別の薬剤にスイッチした後の継続率と、その薬剤からTCZにスイッチした後の継続率を比較した。これによりhead to headの比較がクロスオーバー試験に近い形で示されると考えた。例えばTCZとTOFとのスイッチ例の継続率はどちらもほぼ同等だが、それ以外の薬剤ではいずれもTCZに変更した方が継続率が高いことが示されている。

ABT: abatacept, ADA: adalimumab, CZP: certolizumab-pegol, ETN: etanercept, GLM: golimumab, IFX: infliximab, TCZ: tocilizumab, TOF: tofacitinib

について、100人年のデータで薬剤間の比較をした。各薬剤間の差異は大きなものではないが、特にNTMにおいてABT使用者に高く、HZにおいてTOF使用者に高いことが示されている（図9）[11]。

V. 考 察

月別の各薬剤使用患者数などは本来基礎的記述データとして公的に提示されるべきもののように思われる。こうした取り組みが様々な領域で常時公表される時代が来ることを期待する。今回日和見感染症では、例えば結核患者の定義のためにどのような薬剤の組み合わせがあるか検討しそれぞれの患者数をみることができたが[6]、これはこの組み合わせで治療された国内の全ての患者数を示しているわけで、さらにすべての結核患者、あるいは他の疾患の記述研究にも応用できる手法である。

また今回示されたデータで、今まで薬剤間の比

較が示されていなかった継続率の差異が明らかになった。特に擬似クロスオーバーテストなどを行うには大量の症例数が必要となるために、NDBクラスの国内を包含するようなビッグデータでないと解析が困難である。一方でこの継続率が薬剤間の正当な有効性を示しているかについては多少の疑念がある。すなわちこれは医師の判断の集積であり、新たな薬剤ほど、あるいは他剤に変更することができない薬剤ほど長く使われることになる（代替品がない）とか、毎週よりも隔週、あるいは月1回の注射製剤が、さらに注射薬よりも経口薬の方が患者に好まれること、あるいは薬価も関係してくると考えられる。すなわちこの継続率は「有効性」というよりも「有用性」を反映していると考えた方が正しいように思われる。この継続率の研究手法は例えば抗がん剤の効果、生存率など多くの臨床研究に応用が可能である。われわれは非小細胞肺癌に対する免疫チェックポイント阻害薬の効果についてNDBで解析し、ニボルマ

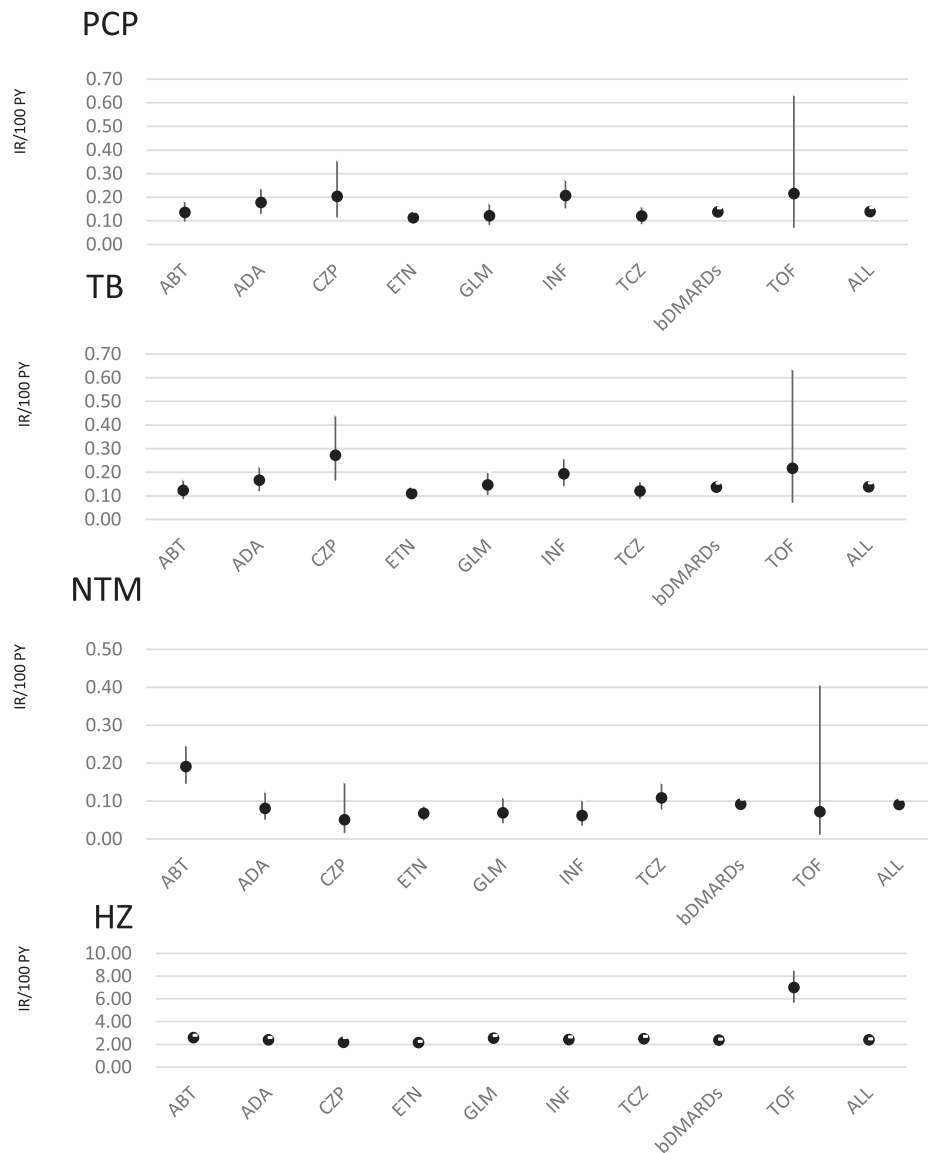


図9 薬剤別日和見感染症発症率

PCP: pneumocystis pneumonia, TB: tuberculosis, NTM: non-tuberculosis mycobacteria, HZ: herpes zoster
日和見感染の発症率に関してはほとんどの分子標的薬間には差がない。これは市販後の5,000名の調査であったデータとは異なるもので、発症率に対して十分に大きな数で長期間を観察しないと真実が見えてこないことを意味している。

ABT: abatacept, ADA: adalimumab, CZP: certolizumab-pegol, ETN: etanercept, GLM: golimumab, IFX: infliximab, TCZ: tocilizumab, bDMARDs: biological disease modifying anti-rheumatic drugs, TOF: tofacitinib

ブでは従来のRCTと比較し、実臨床データでもほぼ同等であることを示している[13]。抗リウマチ分子標的薬の安全性に関しては既に各社に義務付けられた市販後全例調査である5,000例のデータ解析が行われ論文化されている[14-16]。これらは日本からの発表として国際的にも認められているのだが、実は今回NDBで解析するとこれらの論文とは大きく異なり、一部薬剤を除き薬剤間の日和見感染症の発症率にはほぼ差がないことが

示された。5,000人の調査といえば十分のように思えるが、これは日和見感染症の発症率が小さいので5,000人程度では罹患患者が一人増加するだけで大きな差になってしまい、より大量の数で検討しなければ正確な推定ができないことによるのではないかと考えている。また帯状疱疹においてはその発生頻度が当初の市販後調査に比べその後の結果ではかなり大きくなっているが、これは帯状疱疹の発症の中央値が今回のNDB研究では13

か月であり、市販後調査であっても期間が数か月と短いと発症を捉えられない例が多々あるためではないかと考えている[16]。ここで示したように、NDBデータが膨大な人数と長期間のデータであることが薬剤間の安全性の正しい比較を可能にしている。

現時点でのNDB研究の大きな問題は申請からデータを受領するために平均330日もかかっているということである[7]。このため科研費などを利用して研究を計画しても期間内に研究が完成できないことがしばしばおこる。このことについては厚労省側も改善するとして[17]、2025年2月からはNDB- β として迅速に対応するとの発表がある（厚労省ホームページ: NDB匿名医療保険等関連情報データベースの利用に関するホームページhttps://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/iryuhoken/reseputo/index.html: 2025年2月1日現在）。また受領後の研究期間も定まっており、いつまでも研究できるわけではなく、成果の公表については逐一厚労省の認可を受けなければならない。さらに一般の研究と異なり、レセプトデータだけの解析であるから、あらゆる研究が可能というわけではなく、ある限られた疾患、特定の薬剤などに絞られてしまう。それではどのような研究が可能かどうかはその分野の専門の臨床医でないと判断ができない。

NDBでは最近より詳細な情報が包含されるDPCデータによる解析ができるようになった。DPCデータはDPC病院のみのデータだが、全国総病床数の2/3を占め、主要な病院のデータとして限定できるのであれば利用が可能である。今後厚労省の感染症データベース、予防接種のデータベース、難病データベース、小児慢性データベースなど、また次世代医療基盤法に基づき検査結果などの蓄積が行われている千年カルテ[18]などの他のデータベースと連結するような研究が期待される。

今回の研究では、当初から月別での患者数の集計することとしていたが、その結果から偶然見出したものにリウマチ新規治療患者数の周期性がある。新規リウマチ治療患者が春から夏に多くなるというもの（図5）で、その原因は春になると関

節リウマチが悪くなるという事実と相応することと考えられた。このような現象は患者のアンケート調査の結果から報告されていたが[19,20]、専門家においても今まであまり知られておらず、筆者らも今回の研究ではじめて認識した。最近森らも大規模データを解析することで春に悪くなるRAの周期性を示している[21]。本研究に限らず少数例ではわからないことが大規模データにして捉えられたといえるだろう。実際に大量のデータにすることで逆にノイズが小さくなり初めてみえる事実や、より明確になる事実は大規模データ解析に期待をもたせるものである。例えばLittleJohnsは5,000人の解析ではわからない年齢別の高血圧と虚血性心疾患死亡率の相関が、50万人では明確になることを示した[22]。

一方で交絡因子については常に十分な検討が必要である。retrospectiveな研究であり、数が庞大でもバイアスが排除できなければ解析結果の意味は失われてしまう。その意味でNDB研究において交絡因子の排除は最大の課題といえるかもしれない。自験例の研究でも新薬ほど継続率が高く出る傾向があることや、安全であると考えられている薬剤ほどリスクの高い症例に投与されていると考えられるような結果がみられた。できる限り層別化するなどしてバイアスの排除をすることが必要である。

もし本稿を読んでNDBに興味を持っていただいたら、まずはオープンデータを参照し、サンプリングデータセットなどでのトライアルを行うことをお勧めする。はじめから特別抽出となるとかなりハードルが高いため、小規模（といっても十分大規模であるが）のデータでトライして、そののち特別抽出へと移られた方が賢明であるとするのが、先人たちの多くの意見であることを添えておく。千葉大学には次世代医療構想センターなど厚労省でNDBデータの創設にかかわった先生方やNDBに関して研究されている先人が少なくないので、まずはこうした先輩にご相談されることを勧める[23]（佐藤大介（2020）第2回「医療ビッグデータを活用した研究セミナー」<https://www.youtube.com/watch?v=vSu9I1Xo8dQ>, 梅澤耕学（2020）NDBの今とこれから

seminar-1st-shiryo_2umezawa.pdf)。また大量のレセプトデータを扱うためにはレセプトをよく理解したコンピュータサイエンティストの能力が必要である。臨床研究を計画し、レセプトからの抽出まですべて医師だけで行うのは困難で、臨床研究者とコンピュータサイエンティストの合同チームで作業の方が効率がよい。その意味で適切なパートナー探し、チームを作った人たちには大きな恩恵が与えられるであろう。またNDBの解析のためには鍵のかかる管理された専用の個室が必要となる。一研究者が学内各教室でこのような部屋を準備することは容易ではない。スペースとしては駅のワークステーションのように小規模ですむものであるから、学内の空きスペースを希望者に活用できるよう医学部、病院として検討されることを期待する。本学からNDBを用いた研究が今後続々と出てくることを切に願うものである。

貢献者

本論文の作成にあたっては、高林は論文作成を、安藤はNDBデータの具体的解析を、土井、鈴木は論文校閲を担当した。

財源支援

文科省科学研究費基礎研究C(21K12119)を受けて研究を遂行した。

利益相反

今回の論文に関して申請すべき利益相反はない。

倫理的承認

今回の発表内容は厚労省申請時に際して千葉大学大学院医学研究院倫理審査委員会の承認(承認番号: 2842)と発表に際しては厚労省に届け出て認可を受けている。

謝辞

抗結核薬の組み合わせによる結核、非結核性抗酸菌症の定義についてはそれぞれ現代の本邦を代表する専門家である山岸文雄先生、Ho Namkoong先生にお教えいただいた。この場を借りて御礼申し上げます。

Abstract

The National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB) is the large and nationwide database which is composed of 25 billion of claims data for more than 120 million people in the entire country and the government opened it to Japanese researchers for 13 years since 2011. As the NDB includes nearly all claims data in this country, it is unprecedentedly big real-world data with completeness in the world. In this article we introduce how to utilize the NDB resources in clinical research as well as epidemiology. For example, we extracted all cases with rheumatoid arthritis who received eight molecular targeted drugs and compared the retention rates and occurrence of opportunistic infections between the drugs. We also discovered the seasonal change of the number of new cases for these therapy by using big data.

文献

- 1) Takabayashi K, Ando F, Suzuki T. (2019) Comparing the effectiveness of biological disease-modifying antirheumatic drugs using real-world data. *Mod Rheumatol* 29, 87-97.
- 2) 寺島健史. (2008) 下肢人工関節置換術実施患者の周術期における静脈血栓塞栓症の発症予防 第31回日本血栓止血学会学術集会.
- 3) Fujita M, Hashimoto H, Nagashima K, Suzuki K, Kasai T, Yamaguchi K, Onouchi Y, Sato D, Fujisawa T, Hata A. (2023) Impact of coronavirus disease 2019 pandemic on breast cancer surgery using the National Database of Japan. *Sci Rep* 13, 4977-86.
- 4) Miyashita K, Hozumi H, Furuhashi K, Nakatani E, Inoue Y, Suzuki Y, Karayama M, Enomoto N, Fujisawa T, Inui N, Ojima T, Suda T. (2024) Impact of preexisting interstitial lung disease on mortality in COVID-19 patients from the early pandemic to the delta variant epidemic: a nationwide population-based study. *Respir Res* 25, 95-103.
- 5) Takabayashi K, Ando F, Ikeda K, Fujita S, Nakajima H, Hanaoka H, Suzuki T. (2022) Trend in

- prescription and treatment retention of molecular-targeted drugs in 121,131 Japanese patients with rheumatoid arthritis: A population-based real-world study. *Mod Rheumatol* 32, 857-65.
- 6) Takabayashi K, Ando F, Ikeda K, Nakajima H, Hanaoka H, Suzuki T. (2023) Incidence of opportunistic infections in patients with rheumatoid arthritis treated with different molecular-targeted drugs: A population-based retrospective cohort study. *Mod Rheumatol* 33, 1078-86.
 - 7) 野田龍也. (2023) NDBを使いやすくするための研究・取り組み. 日本薬学疫学会シンポジウム 実務担当者のためのNDBの現状・課題・展望.
 - 8) Kubo S, Noda T, Myojin T, Nishioka Y, Higashino T, Matsui H, Kato G, Imamura T. (2018) National Database of Health Insurance. Claims and Specific Health Checkups of Japan (NDB): outline and patient-matching technique. *bioRxiv*. <https://www.biorxiv.org/content/early/2018/04/02/280008.full.pdf>.
 - 9) Sato J, Yamada H, Goda K, Kitsuregawa M, Mitsutake N. (2019) Enabling Patient Traceability Using Anonymized Personal Identifiers in Japanese Universal Health Insurance Claims Database. *AMIA Jt Summits Transl Sci Proc* 2019, 345-52.
 - 10) Nishioka Y, Kubo S, Okada S, Myojin T, Higashino T, Imai K, Sugiyama T, Noda T, Ishii H, Takahashi Y, Imamura T. (2022) The age of death in Japanese patients with type 2 and type 1 diabetes: A descriptive epidemiological study. *J Diabetes Investing* 13, 1316-20.
 - 11) Ando F, Takabayashi K, Fujita S, Nakajima H, Hanaoka H, Suzuki T. (2023) Seasonal exacerbation of rheumatoid arthritis detected by big claims data analysis: A retrospective population study. *Mod Rheumatol* 33, 46-53.
 - 12) Ebina K, Hashimoto M, Yamamoto W, Hirano T, Hara R, Katayama M, Onishi A, Nagai K, Son Y, Amuro H, Yamamoto K, Maeda Y, Murata K, Jinno S, Takeuchi T, Hirao M, Kumanogoh A, Yoshikawa H. (2019) Drug tolerability and reasons for discontinuation of seven biologics in 4466 treatment courses of rheumatoid arthritis-the ANSWER cohort study. *Arthritis Res Ther* 21, 91-100.
 - 13) 高林克日己. (2024) シンポジウム ナショナルデータベースの今後 第28回日本医療情報学会春季シンポジウム.
 - 14) Takeuchi T, Tatsuki Y, Nogami Y, Ishiguro N, Tanaka Y, Yamanaka H, Kamatani N, Harigai M, Ryu J, Inoue K, Kondo H, Inokuma S, Ochi T, Koike T. (2008) Postmarketing surveillance of the safety profile of infliximab in 5,000 Japanese patients with rheumatoid arthritis. *Ann Rheum Dis* 67, 189-94.
 - 15) Koike T, Harigai M, Inokuma S, Ishiguro N, Ryu J, Takeuchi T, Tanaka Y, Yamanaka H, Fujii K, Freundlich B, Suzukawa M. (2011) Postmarketing surveillance of safety and effectiveness of etanercept in Japanese patients with rheumatoid arthritis. *Mod Rheumatol* 21, 343-51.
 - 16) Kanbori M, Suzuka H, Yajima T, Kishino E, Morishige R, Momohara S, Kawakami A, Ota M. (2018) Postmarketing surveillance evaluating the safety and effectiveness of golimumab in Japanese patients with rheumatoid arthritis. *Mod Rheumatol* 28, 66-75.
 - 17) 吉村健佑. (2024) シンポジウム ナショナルデータベースの今後 第28回日本医療情報学会春季シンポジウム.
 - 18) 吉原博幸. (2022) 千年カルテ: 次世代医療基盤法に基づく診療データの研究開発への二次利用. *薬剤疫学* 27, 3-10.
 - 19) 滝沢健司, 島井哲志, 田中正敏. (1991) 慢性関節リウマチと環境因子の関係 日本各地における悪化の訴え率の季節変動について. *公衆衛生* 55, 887-89.
 - 20) Iikuni N, Nakajima A, Inoue E, Tanaka E, Okamoto H, Hara M, Tomatsu T, Kamatani N, Yamanaka H. (2009) What's in season for rheumatoid arthritis patients? Seasonal fluctuations in disease activity *Rheumatology* 46, 846-48.
 - 21) Mori H, Sawada T, Nishiyama S, Shimada K, Tahara K, Hayashi H, Kato E, Tago M, Matsui T, Tohma S. (2019) Influence of seasonal changes on disease activity and distribution of affected joints in rheumatoid arthritis. *BMC Musculoskeletal Disorder* 20, 30-7.
 - 22) Littlejohns TJ, Sudlow C, Allen NE, Collins R. (2019) UK Biobank: opportunities for cardiovascular research. *Eur Heart J* 40, 1158-66.
 - 23) Sako A, Gu Y, Masui Y, Yoshimura K, Yanai H, Ohmagari N. (2023) Prescription of anti-influenza drugs in Japan, 2014-2020: A retrospective study using open data from the national claims database. *PLoS One* 18, e0291673.
-