

データベース研究入門

日本製薬工業協会

医薬品評価委員会 データサイエンス部会

2015 年度タスクフォース 3

「Real World Data の活用」

Ver 1.1

2015 年 11 月

目次

はじめに.....	1
1. 医療情報データベースとは？.....	2
1.1 本報告書での医療情報データベースの定義.....	2
1.2 医療情報DBに含まれるデータの臨床的妥当性の限界.....	3
1.3 医療情報DBの種類と特徴.....	5
1.3.1 <i>Clinical DB</i>	5
1.3.2 <i>Administrative DB</i>	7
2. 医療情報DBが役に立つ可能性のある場面とは？.....	11
2.1 臨床開発.....	11
2.2 安全性評価.....	11
2.3 医薬品の価値の相対的評価.....	13
3. DB研究の進め方.....	15
3.1 社内体制.....	15
3.2 研究フローと事前準備.....	16
3.3 REQUEST FOR PROPOSAL.....	17
3.4 DBの選択.....	18
3.5 研究実施計画書.....	22
3.6 解析と結果の解釈.....	24
おわりに.....	25
REFERENCES.....	26

はじめに

近年、Real World Data (RWD) が注目されている。「実験的」な環境で得られる臨床試験のデータに対して、「Real World」、すなわち臨床現場で得られるデータであり、より一般的な患者集団における医薬品の有用性や安全性、使用実態や経済効果を検討するデータソースとして期待される。ただし、目的に沿って前向きに収集される治験のデータとは異なり、他の目的で収集される RWD の二次利用においては、データの質や臨床的妥当性におのずから限界があることも、十分に認識されるべきである。

2015 年度データサイエンス部会タスクフォース 3 は、医療データベース協会の協力を得て、「Real World Data の活用」というテーマのもと、活動を開始した。具体的には、RWD の中でも医療情報データベース (DB) に含まれるデータを対象とし、日本の製薬企業における使用の現状を探るとともに、データの質やサイズ等の特質に関する考察を深め、その結果を広く発信する予定である。

本報告書は、医療情報 DB を用いた研究 (DB 研究) の経験がないかまたはごく少ない方を対象としている。第 1 章ではそもそも医療情報 DB とはどのようなものか、その由来や種類、特徴、限界などについて紹介する。第 2 章では、製薬企業の業務を考えたときに、医療情報 DB が役に立つ可能性のある場面や医療情報 DB の強みを活かせる研究の条件について概観する。第 3 章は実践編として製薬企業が DB 研究を企画・実施する場合に必要な準備や注意点などについて仮想事例とともに記載する。

海外にも多くの医療情報 DB があり、DB ベンダの研究経験も豊富であるが、本報告書では日本で研究に応用可能な DB を主に対象としている。なお、薬剤疫学研究やリサーチ・クエスチョンの設定など、DB 利用に限らない観察研究の基本的な知識や考え方の解説は、本報告書のスコープ外であり、成書をはじめ、これまでにデータサイエンス部会から公表されている報告書などを参照されたい。

医療情報 DB には観察研究を行うに当たって魅力的な利点が多くある一方、理解しておくべき限界もある。本報告書が医療情報 DB の適切な利活用入門の一助となれば幸いである。

1. 医療情報データベースとは？

データサイエンス（以下 DS）担当者の A さんは、近年、ベネフィット・リスク評価や費用対効果分析など、通常業務である治験の計画や解析とは異なる業務にまで駆り出され、忙しい日々を送っています。

「おー、A、なんか最近リアル・ワールド・データって流行ってるらしいじゃないか」

ある日、上司に軽い感じで声をかけられました。

「はあ」

A さんは RWD という言葉をネットで見かけたことがある気がしましたが、正直なところあまり現実味はありません。

「わが社も波に乗り遅れないように、新しいものをどんどん利用していかないとな！」

こ、これは・・・つまり調べておけてことか???

社会人 10 年目の A さんは勘を働かせます。とはいえ、どこから手を付けたものか。



近年、Real World Data (RWD) が注目されている。"Data"の部分は"Evidence"や"Informatics"などのバリエーションもあり、それぞれ RWE, RWI などと略されるが、似たような用語である Big Data とは異なり、RWD といえば主に医療情報に関するものを指しているようである。製薬業界で考えるなら、「実験的」な環境で得られる臨床試験のデータに対して、「Real World」、すなわち臨床現場で得られるデータとして位置づけられるだろう。RWD は営業管理やマーケティング戦略などでは以前から活用されてきたが、本報告書では医学的・科学的妥当性がより求められる観察研究目的での利用を対象とする。観察研究の症例収集法には、目的に沿って一例ずつ収集する方法と、別の目的ですでに収集されていたデータを二次利用する方法があるが、RWD を利用する研究は後者に分類される。RWD は広義には個人のブログや SNS など含まれるかもしれないが、ここでは一般的によく研究に利用されるデータソースである「医療情報データベース (DB)」を対象とし、DB のデータを二次利用する研究 (DB 研究) について述べる。

1.1 本報告書での医療情報データベースの定義

本報告書では「医療情報データベース (DB)」について、特定の目的以外にも応用できる可能性を持つ、すなわち二次データとしての役割を果たすもののうち、電算化され、比較的標準化され、ある程度までクリーニングされ、柔軟なデータ検索や抽出が可能なものと定義する。DB 研究では、対象となる集団を DB の中で得られるデータ項目、例えば診断コードや診断月、治療薬の処方などを組み合わせて定義し、抽出する作業が発生する。ところが、RWD は臨床試験のデータセットとは異なり、標準化もクリーニングもされていない、データ間に不整合がある、検索やデータ抽出ができないなどの状況も珍しくない。このような背景から、本報告書では医療情報 DB を「単なる医療データの集合」としてではなく、特定の項目に基づく検索や抽出を可能とするデータ構造を持つものとして定義した。この定義は本報告書に限るものであり、他の文脈における定義を否定するものでは決してない。例えば PMDA のガイドライン[1]では、「請求書データ、HIS/EMR (Hospital Information System/Electronic Medical Record) データ、レジストリ等の医療情報を含むデータからなるデータベース」と定義されている。

1.2 医療情報 DB に含まれるデータの臨床的妥当性の限界

医療情報 DB に含まれるデータを二次利用して研究を行う際には、医師が患者を診断し、治療した時には周知であった背景情報、いわば文脈が欠けていることを意識する必要がある。また、保険償還目的に加工されている情報（レセプトなど）は標準化される際にもとの臨床情報に比べて粒度が荒くなる傾向がある。ここでは筆者らの経験に基づいていくつかの事例を紹介し、DB 研究ならではの問題点について注意喚起する。

1) 疾病の定義の問題

疾患の定義は、研究の対象集団や曝露に対するアウトカムを決定する上で非常に重要なポイントとなる。国の統計調査や医療機関におけるカルテやレセプトでは、疾病は ICD-10 (International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; 疾病及び関連保健問題の国際統計分類 第 10 版) という国際分類でコード化されている (厚生労働省 疾病、障害及び死因の統計分類 [<http://www.mhlw.go.jp/toukei/sippeij/>])。したがって、DB 研究で疾病を定義する際にもこの ICD-10 コードを使うのが一般的である。しかし、研究の目的によっては ICD-10 コードだけで疾病を定義するには限界がある。

例えば、医療情報 DB を利用してある急性疾患の発症数を推計したいとする。その疾患には特定の ICD-10 コードがあり、ある一年間に限って検索してみると、その DB 内でそのコードを持つ患者は 2,000 人という結果が得られた。この DB は全国のおよそ $x\%$ をカバーしていることから、単純にこれを $100/x$ 倍すると・・・別の方法で推計されている患者数の 5 倍にもなってしまう！なぜか？

この疾患は子供に特有であり、10 歳以上にはほとんどないことが知られているが、DB 内で得られたデータで年齢分布を見てみると、10 歳以上にピークがあることがわかった。この分布から、罹患時に治療を求めて来院した場合だけでなく、治癒した後に合併症が発生していないか確認するために経過観察で定期来院した子供にも ICD-10 コードがつくため区別なく数えてしまい、見かけの発症数が多くなってしまっている可能性が示唆される。しかし、ICD-10 コードだけでは来院が罹患時なのか経過観察なのかは区別できない (つまり特異度が低い)。どうするか？

この疾患には標準的な治療薬がある。ガイドラインでも推奨されており、診断されればまずこの薬剤を投与される。そこで、ICD-10 コードにこの薬剤のコード (薬剤のコードには複数あり、ここでは割愛する。例えば ATC (Anatomical Therapeutic Chemical Classification System) など) を組み合わせ、同じ月にこれらの ICD-10 コードと薬剤コードがある、という疾患定義で再度検索してみると、患者数は 400 まで減少し、他の推計データともよく一致した。また DB 中で一定期間中 (例えば 1 年) にこの ICD-10 コードがなく、その後初めてこのコードがついたときだけをカウントするデザインにすることで対処できるかもしれない。

このように、疾病コードや治療、処置などを組み合わせた疾患の定義をアルゴリズムと呼び、このアルゴリズムの妥当性を検討する研究を validation study と呼ぶ。validation study の詳細は後述する。上記の事例では ICD-10 コードのみを用いるアルゴリズムより ICD-10 コードに特定の治療薬の処方を組み合わせたアルゴリズムの方が妥当性が高いことが示唆されている。また、他の統計データな

どを用いて間接的に妥当性を確認する方法を external validation と呼ぶこともある。

■□コラム□■ ICD-10 (International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; 疾病及び関連保健問題の国際統計分類 第10版) は、世界保健機構(WHO)が設定した国際疾病分類の第10版。様々な病気を分類するための分類表である。統計法に基づく統計調査に使用されるほか、医学的分類として医療機関における診療録の管理等に活用されている[2]。標準化された疾患分類を使用することで、施設間や都道府県、国家間の比較や客観的な統計の作成が可能になる。

2) コーディングの問題

上記事例のように、アルゴリズムには ICD-10 コードをベースに検討するのが一般的であるが、その妥当性を検討する上では ICD-10 そのものの粒度や医療施設でのコーディングの実際を理解する必要がある。

例えば、医療情報 DB を利用してある急性感染症の発症数を推計したいとする。DB 内で検査結果が利用できれば高い特異度・感度が期待できるが、日本の医療情報 DB で標準化された感染症の検査結果を十分に含んでいるものはない。この疾患には複数の関連 ICD-10 コードがあるが、合併症を起こした重症例を示すコードのみでは、軽症例が含まれないため感度が低くなってしまう。一方軽症例は ICD-10 で一般的に使われる 4 桁のコードでは他の感染症と見分けられないため、今度は特異度が低くなってしまう。どうするか？

日本の標準病名マスタ [http://www.medis.or.jp/4_hyojyun/medis-master/]は ICD-10 コードより粒度が細かい。ICD-10 に加えて標準病名を用いることでアルゴリズムの感度・特異度を上げることができる可能性がある。この疾患の場合も、標準病名まで使えば軽症例も拾うことができる。基本的にはコーディングは保険請求のためになされるので、検査や治療にかかった費用の保険請求に当たり、どのような疾病コードが必要かに強く依存することになる。この疾患の場合は、検査に当たり必要な疾病コードは特に定められておらず、また特有の治療法もないため、保険請求上は無理に診断を特定する必要はないが、疾患によってはこの点を考慮する必要がある場合もある。すなわち、これらの診断名が真にその感染症を表しているかはこの段階ではわからないため、カルテや検査結果などを gold standard として validation study を行い、アルゴリズムの妥当性（陽性/陰性適中率、感度、特異度）をさらに検討することになる。

		Gold standard			
		あり	なし		
アルゴリズム	あり	a	b	陽性適中率 $a/(a+b)$	
	なし	c	d		陰性適中率 $d/(c+d)$
		感度 $a/(a+c)$			
			特異度 $d/(b+d)$		

このように、研究目的で収集されていないデータを二次利用するに当たっては、疾患の定義一つをとってもその臨床的妥当性を十分に吟味する必要がある。そのためには DB の成り立ち（本来のデータ収集目的）、データソース（データの由来）などをよく理解する必要がある。次項からは DB の種類別にそ

の特徴の概略を紹介する。

1.3 医療情報 DB の種類と特徴

本報告書では、DB 研究に応用することを念頭に置いて医療情報 DB を定義している。DB 研究は、データの二次利用であり、研究実施計画書に基づいてデータを収集する研究手法と比較して、利用可能なデータ項目も限られるし、データの質にも限界がある。それでも医療情報 DB を研究ツールとして利用することを検討するのは、それなりのメリットもあるからである。例えば、特定の研究を目的として収集されたデータではないので、特定の疾患や薬剤、検査、患者層に限らずデータが網羅的に収集されている（網羅性）。同様に、研究費の継続期間や特定の研究者の熱意によらずデータ収集が継続され（継続性）、一例ずつ症例を集積するのに比べて大きな集団のデータを早く（迅速性）、また比較的安価に収集できる。逆に言えば、これらのメリットが享受できないなら、従来通り医療施設と契約して、信頼できるデータ収集を行えばよいのであり、あえて医療情報 DB を選ぶ理由はないはずである。医療情報 DB 中でまだ標準化されていないデータや、もともと含まれていないデータを無理に求めると、データ所有者側も研究者側にも多大な時間と労力がかかってしまう。それぞれの DB の良さを活かして研究に応用することが肝心である。

上記の利点は、すべての医療情報 DB に共通なわけではなく、それぞれの DB の一次目的（収集されている理由）やデータソースによってさまざまである。例えば、その患者が受診した複数の医療機関にまたがるデータが集積される DB では、単一の医療機関と契約して症例を集積する場合と比較して追跡可能性が高くなる。さらに保険請求目的のデータであれば、データコードが全国的に標準化されているなどの利点がある。

医療情報 DB には大きく分けて臨床 DB（clinical DB）と医療管理 DB（administrative DB）がある。前者はその名の通り診断・治療に関連する記録であり、後者は病院・薬局経営に必要な保険請求申請のために必要なデータなど、主にお金の流れに関する記録からなっている。後者の方がより標準化されており、DB 研究に利用しやすい側面がある一方で、検査値など研究でニーズの高いデータが含まれていないなどの限界もある。

1.3.1 Clinical DB

Clinical DB は多くの人がイメージするであろういわゆる「電子カルテ」に近いものである。しかし、院内で「電子カルテ」として運用されているのは、実は院内の複数の独立した DB をカルテ ID でリンクできるようにしたシステムであり、単一に統合された「電子カルテ DB」があるわけではない。

病院内には多くの異なるシステムに由来する DB がそれぞれ独立に存在している。例として、予約システム、画像システム、オーダーリングシステム、会計システムなどが挙げられ、これらのシステムからそれぞれの目的に沿ったデータ項目のみを含む DB が別々に生成される。いずれの DB にも患者のカルテ ID が含まれており、したがって院内ではこれらの DB をカルテ ID でリンクすることが可能である。自

分が患者として受診したとき、医師が一つのモニタでいろいろな情報を見ることができているのは、あなたのカルテ ID を識別コードとして複数の DB から情報を都度呼び出しているからである。

また、臨床上の一次目的のためには、特定の患者に関するデータを各 DB からカルテ ID で呼び出すことができ、その内容を医師が把握できればよい。院内の DB 間で用語やコードが標準化されているとは限らないし（コラム参照）、DB 研究のためにたとえば「ある薬剤をある一定期間に処方されたすべての人」を検索・抽出できるようなシステムにもなっているとは限らない。

Clinical DB のデータソースは臨床上必要となるあらゆる医療データであるが、院内で利用できるすべてのデータが DB 研究のために利用可能になっているわけではない。院外の研究者も含めて利用できるように、匿名化・コード化（標準化）され、日々データが更新されるようなシステムを持つ医療機関はごく限られているし、そのような施設であっても利用可能なデータは患者基本情報（年齢や性別など）、診断、処方や検査オーダーなど、院内の DB のうちの一部から得られるさらにごく一部のデータに限られる。検査結果や画像診断のシステムから生成されるデータや、患者の自覚症状や生活習慣に関するデータ、また薬剤の有害事象などが標準化された DB として利用可能なわけではない。上述の院内各種 DB のうちもっとも DB 研究への汎用性が高いのは、処方や検査などのオーダリングシステムを利活用するものである。オーダリングは保険請求システムへのフローがあるため、他の臨床データと比較して標準化が容易であると期待される。日本ではある程度の規模の施設であればそのほとんどでオーダリングシステムがすでに導入されている[3]。一方で、電算化されたシステムがあっても、紙ベースのオーダーがゼロになるわけではない。緊急時や特殊な薬剤など、臨床上の必要性から紙運用が一部残っている場合もあり、この場合には会計システムや診療報酬明細（レセプト）には反映されても、オーダリングシステム上には現れないことになるので注意が必要である。

Clinical DB の一次目的は、診断や治療の記録であり、これはどんな医療施設でも一定期間保存することが義務付けられており、特定の研究期間だけでなく継続的なデータ集積がある。特定の疾患や薬剤、検査にとどまらない網羅性があり、DB 研究に活用できる可能性は高い。また多くの場合、ほぼリアルタイム（たとえば毎日など）にデータがアップデートされるため、迅速性もある。Clinical DB の最大の利点は、カルテに戻ることができる点である。研究のためにデータが匿名化されていたとしても、研究用 ID とカルテ ID との対応表があれば、検索・抽出したデータの中から詳細を検討したい症例を選び、診断の確からしさや症状、転帰などの追加情報を得られる可能性もある。一方で、継続的なデータ集積があるが、特定の患者がその特定の施設に通い続ける（あるいは入院し続ける）とは限らないため、追跡可能性は低い。さらに、標準化されたデータを外部の研究者にも利用可能な形で保有する施設はまだ限られているため、データサイズや結果の一般化可能性にも限界がある。分散型研究ネットワーク（各施設の DB を物理的に統合せず、研究ごとに最小限のデータセットやデータサマリだけを抽出する研究の仕組み）を利用するなど工夫が必要である。複数の施設にまたがる DB 研究を円滑に実行できるようにするためには施設を超えた標準化の導入の他に、契約や倫理委員会等の事務手続きの共通化などが今後の課題として残されている。

また Clinical DB には、各施設のデータをもとにしたものの他に、疾患領域によっては疾患レジストリ（患者登録）が存在する（たとえば[4]）。研究目的での二次利用の可能性は、その疾患関連のテーマに絞られるが、その分その疾患特有のデータが豊富であることが多い。多くの場合、そのレジストリに貢

献した研究者チームや学会の所有であり、企業による直接的な利用は難しいかもしれないが、アカデミアとの共同研究により利用できる可能性はある。

■ □コラム □ ■ 臨床データの標準化

「電子カルテ」に入力されている情報の多くは、いわば医師の覚書（メモ）であり、このような自由記載のデータは一般的に標準化が難しく、ひいては情報の検索も当然困難となる。特に日本語は2バイト文字であり、半角と全角のばらつきがあるのに加え、漢字・平仮名・カタカナ・記号などを考えると、この部分を「データ」として利用できるまでにかかる標準化やクリーニングの手間は計り知れない。

さらに、臨床検査結果も一般的には標準化が難しいデータ項目の一つである[5]。レセプトでは保険点数と紐づいた検査コードが用いられているが、ここには検査依頼の記録しかなく、結果はまた別の院内DBに格納されている。この検査結果を格納するDBには検査実施項目も記載されているが、ここでは検査実施項目は標準化コードではなく病院独自の名称が使われていることが多い。したがって、複数の施設から検査結果DB由来のデータを収集しているDBの場合には、その検査をその施設で何と呼んでいるかから調べる必要がある。さらに、一つの検査で複数の結果が出ることもあるし、同じ結果でも施設によって数値、二値データ（+/-、陽性/陰性、pos/neg など）、テキスト（結果の解釈など）など、フォーマットにばらつきがある。

1.3.2 Administrative DB

Administrative DB は、簡単に言えば「医療機関による保険請求の妥当性を担保するための記録」で構築されたDBである。医療機関や薬局は誰にいつどのような診療を行ったので、その月の保険点数の合計はこれだけになる、のような記録を作成し、提出する。これが診療報酬明細書（レセプト）である。レセプトには医科（入院、入院外）、調剤、歯科があり（Figure 1-1）、さらにDPC対象病院（Diagnosis Procedure Combination/Per-Diem Payment System、急性期入院医療の診断群分類に基づく1日当たりの包括評価制度）からはDPCデータも提出される。

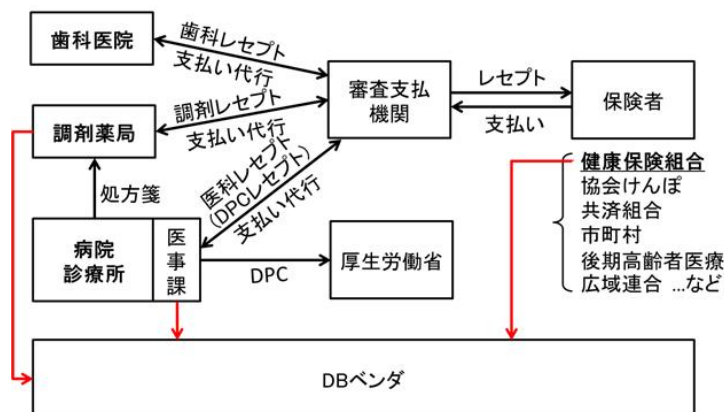


Figure 1-1 種々の administrative DB 間でのレセプトデータの流れ

レセプトは医療機関から審査機関を通じて保険者（保険組合や国など、支払者）に提出される。保険者は、委託先の支払審査機関を介して提出された記録を精査し、支払いの妥当性があると判断すればその保険点数に応じた請求額を医療機関に支払う。健康保険組合など保険者はこのデータに基づいて保険償還を行う

健康保険組合など保険者は、被保険者の健康状況の把握や健康促進施策に必要な解析を専門業者に依頼する場合がある。同様に医療機関も経営支援やベンチマーキングの目的で業者に解析を依頼することがある。これらの業者のうち、データの二次利用許諾を得て調査・研究目的に提供しているのがDBベン

ダである。この図ではレセプトの流れを示しているが、特定健診データや検査データなど、レセプト以外のデータを取り扱う DB ベンダも存在する

■ □ コラム □ ■ DB ベンダってどんな会社？

医療情報 DB を提供する民間ベンダにとって、製薬企業は第一の顧客ではないことが多い。では DB ベンダが誰に対してサービスを提供しているかという点、第一にはデータ提供者である。データ提供者は病院や薬局などの医療関連施設と、健康保険組合などの保険者の 2 種類に大きく分けられる。DB ベンダはこれらのデータ提供者からデータを預かり、種々の分析を実施してベンチマーキングしたり、クリティカルパスを提案したりしてデータ提供者の経営支援を行う会社である。DB ベンダはこの業務を通じてデータの標準化やクリーニングを日々実施しており、それぞれが保持するデータの内容をもっともよく把握している。また、アカデミアや民間企業に調査・研究目的で提供されるデータは、DB ベンダが保有するデータのうち二次利用の許諾を得られたごく一部であり、実際にはもっと大きく、多様なデータを扱っている。

保険請求に関するデータは疾患や薬剤、検査の種類によらず、すべて標準化コード[2]に従って記録され、整理され、支払者への提出のためにまとめられている。逆に言えば支払いが発生しないデータ、例えば検査結果（検査の実施にはお金がかかるが、検査結果が支払いに影響することはない）、死亡（別途保険組合からお悔やみが支払われることはある）、自覚症状（これらは保険償還を受けられる何らかの治療が行われた場合には、実際の診断によらず記録上は何らかの診断名が付される）や、保険対象外の診療行為（例えばワクチン接種など）は、administrative DB に記録されることはない。なお、死亡については入院から死亡に至った場合には入院レセプトの転帰欄に記載されるが、院外での死亡については必ずしも記録が残るわけではない。さらに疾患によるばらつきや、記録上二度死んだり生き返ったり(!)する例があることも知られている[6]。また、健康保険組合の種類により、構成年齢が大きく異なる。例えば社会保険健康組合は、被保険者が定年に達する年齢付近から加入者数は減少し、75 歳以上になると後期高齢者制度に移行するため、75 歳以上のデータはない。

さらに、個人情報保護の観点からすべての人の誕生日が一年のうちのある 1 日に変更されていることもある。大人ではさほど問題にならないが、小児では 1 歳で罹患したとされる子供の約半分は実はまだ誕生日前、すなわち 0 歳であり、見かけ上 0 歳の患儿が少なくなってしまう問題が起こる。誕生日および罹患日に関する季節性を考慮してモデルを作って推計し直してもよいが、元のデータに戻れば真の誕生日はあるので DB ベンダと交渉することで解決できる。

日本で研究に応用可能な医療情報 DB は、公的なものも民間のものもあり、日本薬剤疫学会がその一部を表にまとめ、毎年データを更新している[7]。公の administrative database としては、全国民をカバーするレセプト情報・特定健診等情報データベース[8]、通称 National Database (NDB) がある。現状では企業から直接利用申請をすることはできないが、アカデミアからは次々に研究成果が発表されている[9-13]。同じく公的な administrative database としては全国の DPC 病院の約半分をカバーする DB もある[14-27]。民間の DB ベンダについては、医療データベース協会のウェブサイト[<http://www.amdj.org/>]も参考になる。

この項のまとめとして Table 1-1 に DB の種類別に，その特徴をごく簡単まとめた。

Table 1-1 DB の種類別の特徴

	Clinical DB	Administrative DB
一次目的	診断や治療	保険請求
代表的なデータソース	処方や検査のオーダーリング, 検査結果の DB など	診療報酬明細 (レセプト) やDPC ⁷
ID	カルテ ID	被保険者番号
網羅性 ¹	○	○ 保険の種類に依存する DB は年齢や患者層について△
継続性 ²	○	○
迅速性 ³	○	△ (DB による)
追跡可能性 ⁴	△ (病診連携や病院間連携があってもカルテ ID は施設特有)	○ 処方箋 DB や病院をデータソースとする DB は△
標準化	△ (施設による) ⁶	○
アクセス ⁵	△	○
得意な研究	<ul style="list-style-type: none"> ◆ 元の診療記録に戻って，妥当性を確認する必要のある研究 ◆ 画像診断や検査結果，医師の判断が必要な研究 ◆ 短期間に施設内で曝露もアウトカムも測定できるような疾患に関する研究 (例：救急搬送や手術前後など) 	<ul style="list-style-type: none"> ◆ 個人識別子 (被保険者番号など) を活用した，複数の医療施設，診療科にまたがる研究 (例：処方が必要となった疾患を治療する診療科や施設とは異なるところで発見される可能性のあるアウトカムに関する調査研究) ◆ 重篤なイベントや入院などの「わかりやすい」エンドポイント ◆ 疾患領域や薬剤の種類に寄らない網羅的な研究
不得意な研究	<ul style="list-style-type: none"> ◆ 長期の追跡が必要な研究 (施設を離れたら追跡できない) ◆ 一人の患者につき複数の施設からのデータが必要と考えられる研究 (例：薬剤の併用に関する研究など) ◆ 多くの施設からのデータが必要な研究 (例：施設間のプラクティスにばらつきが予想される場合など) 	<ul style="list-style-type: none"> ◆ DB に含まれるデータ項目では，定義の難しい疾患やアウトカムに関する研究 (例：診断が自覚症状や検査結果に基づいてなされる疾患や，重症度や効果判定に検査値が必要な場合) ◆ (労働者の保険組合由来のレセプト DB では) 働き続けられないような疾患に関する研究
<p>表中の○，△，×は，それぞれ「利点として挙げられる」，「ケースバイケース」，「期待できない」をおよその目安として示している</p>		

1. 疾患や薬剤，検査，患者層に関する網羅性
2. 研究費の継続期間や特定の研究者の熱意によらないデータ収集の継続性
3. データを研究目的で二次利用できるようになるまでの期間の短さ
4. 特定の患者を疾患や受診機関によらず継続的に追跡できる可能性
5. 製薬企業に所属する研究者がデータを利用できる可能性
6. 医療情報学会では各種標準化を進めている [<http://www.jami.jp/jamistd/>]
7. DPC 制度（Diagnosis Procedure Combination/Per-Diem Payment System）は，急性期入院医療の診断群分類に基づく 1 日当たりの包括評価制度で，DPC データは DPC 対象病院から厚生労働省に提出される退院時データからなる

このような特徴を考え合わせて，どの DB（一つとは限らない[28]）がその研究テーマにふさわしいか，検討することになる．Table 1-2 にごく簡単に DB 選択の際に考慮すべき点を国際薬剤疫学会（ISPE）の DB SIG (special interest group)のチェックリスト[29]を意識したものを示す．それぞれの項目について，事前情報からすでに確認済みなのか，DB ベンダに確認してみなければならないのかを考えることで，後述の Request for Proposal 作成の一助となるだろう．

Table 1-2 DB 選択の一助となるチェックリスト

チェック項目	確認済み	DB ベンダに 確認が必要
対象集団はカバーされているか：患者数，網羅性，代表性		
研究に必要な変数が含まれているか：曝露，アウトカム，他の変数が，十分に詳細に，偏りなく含まれており，かつ研究に利用可能であるか		
継続的に，一貫性を保ってデータが收拾されているか：個々人で，またデータ全体として，データ収集が途切れる期間がないか．データ収集や診断・治療基準，保険償還の規定やコーディングの方法が途中で変更されていないか		
追跡期間と最新データ取得の時期：平均追跡期間や曝露が起こってから調査までの期間は，対象となるアウトカムが起こるのに十分に長い期間であるか？		
データベースごとの専門性：そのデータベースを利用するに当たり，社内または社外で専門家が必要か		

2. 医療情報 DB が役に立つ可能性のある場面とは？

DS 担当者の A さんはいぶかしく思います。

「治験で収集しているデータや社内の DB とはずいぶん趣が違うようだなあ。DS 担当としてはデータの信頼性や結果の頑強性が心配だ。ほんとに使えるのかなあ…」

「製薬企業の中でもマーケティングや営業部門では、戦略策定や売上情報の管理などに医療情報 DB を使用してきた歴史があるようだが、臨床開発や安全性領域、ベネフィット・リスク評価や HTA など、規制産業ならではのニーズに関しては、具体的にどのような場面で利用できるのだろうか？」



医療情報 DB が臨床開発や安全性、ベネフィット・リスク（再）評価や医療技術評価（HTA）などの領域で利用可能な場面は多くある。

2.1 臨床開発

まず、開発戦略立案時に医療情報 DB は大きな役目を果たす。例えば、ある領域の疾患を対象に薬剤を開発するかどうかを決定する際には、どの国や地域にどの程度のアンメット・メディカル・ニーズ (UMN) があるのかを調査することになるが、そのとき患者数や現在の治療実態、死亡や入院、手術や ICU 入室などの基本的な疫学データは、UMN の定量化に際し重要な評価項目となる。例えば日本で開発するかどうか決まっていな段階で、調査にそれほど長い時間や大きなリソース、コストをかけるわけにもいかない状況で、医療情報 DB を利用してある程度の情報が得られることがある（迅速性）。また、その領域に踏み込まないことが決まったとしても、次に別の疾患を検討するときに同じ医療情報 DB が役に立つ可能性もある（網羅性）。もちろん、疾病負担は死亡や入院などのようなハード・エンドポイントだけで決まるわけではなく、Quality of Life (QoL) の低下などの主観的な評価項目も重要であるが、そもそも患者がその症状に悩んでいたとしても医療施設を受診していなければ医療情報 DB 上には記録が残らない点には留意頂きたい。1.3.2 でも述べた通り、主観的な評価項目の調査に関しては医療情報 DB はあまり向いていない。また、バイオマーカーなどの検査値データも限られている。

次に、開発が決まった後、具体的な試験計画を立てる段階でも医療情報 DB が役に立つ場面がある。例えば各国の治療実態調査は、治験実施国の決定や試験対象集団（年齢や合併症、併用薬など）・組入れ除外基準などの決定の基礎資料として有用である。また、十分に大きく標準化された医療情報 DB があれば、in silico clinical trials (DB の中で、組入れ除外基準などの影響を調査すること) が事前に行える可能性もある。現状では、各施設の医療情報部や薬剤部、医事課などの協力の下、各施設の DB を検索して当該試験の対象となり得る患者数の調査を行い、組入れ計画に利用することが想像される。これらの治療実態や基本的な疫学情報は、あらかじめ調査し当局へ提出することもあるだろうし、照会事項に対応する形で調査することもあるだろう。

2.2 安全性評価

開発段階の薬剤の安全性評価に医療情報 DB が直接的に役に立つことはあまりないが、安全性シグナルの検討や集積症例検討の評価時には利用可能な場合がある。後期開発と並行して医薬品リスク管理計画

(RMP) の策定が始まれば、背景発生率（当該医薬品に曝露されていない対象患者集団（例えば当該医薬品発売前など）における特定のアウトカムの発生率）の調査のため、文献調査とともに医療情報 DB の活用も検討されるだろう。第 3 相試験が終わってから安全性検討事項が変更になることもあるので、迅速に調査できる手段を確保しておくことは重要である。

安全性評価項目をアウトカムとする DB 研究も、欧米では頻繁に実施されている。これは主にイベント発生率の低さに由来する、莫大な症例数や 10 年を超えるような長期の観察期間を満たす試験デザインの実施可能性によるものであり、また利用できる DB の特徴（例えば、全国民がもれなく登録・追跡されているレジストリがある、カルテに戻ることができる DB を用いた validation study（コラム参照）が実施可能、異なる特徴を持つ複数の DB のリンケージが可能である、など）によるものでもある。日本では本当にそのイベントが発生したのかどうか、あるいはそのイベントの詳細について、カルテに戻って調査することができない状況も考えられるので、企業が実施する安全性評価の根拠資料としては、DB 上でのアウトカム定義の妥当性評価と結果の解釈は慎重に行う必要がある。

すでに特定された重要なリスクがあれば、リスク最小化策を実施することになるが、その方策がうまく機能しているかどうかを評価する際にも医療情報 DB は役に立つ。リスク最小化策については、直接処方者に聞く方法では回答者バイアスの可能性が高く、結果の解釈が難しい。処方（併用や用法用量など）や検査の実施など、医療行為で評価することができるリスク最小化策の評価[30]であれば、実施した診療行為が記録されている医療情報 DB の活用が向いている。一方、最小化したい安全性検討事項そのものを評価する場合には、「安全性評価項目をアウトカムとする研究調査」となり、評価項目の臨床的妥当性評価が不十分という事態が生じ得るので、注意が必要である。

■ □ コラム □ ■ Validation study

DB 研究では対象集団、曝露、アウトカムなどを DB に含まれるデータ項目で定義しなければならない。1.2 で述べた通り、この定義をアルゴリズムと呼んでいる。安全性のアウトカムなど特に critical なアルゴリズムについては、その臨床的妥当性を確認するステップとして事前に validation study を行うのが望ましい。これまでに多くの実例が公表されているので（たとえば[6, 31-40]）、国や地域、施設や DB による差異も考慮しつつ、まずはすでに validation study が行われているかを文献調査する。

本調査の前に validation study を行う必要があるとなれば、何らかの gold standard（例えばカルテや画像診断、検査結果など）を設定し、各アルゴリズムについて、陽性/陰性適中率や感度、特異度を計算し比較検討することになる。アルゴリズムとしては例えば、特定の ICD-10 コードのみ、他のコードも加える、一定期間に複数回のコードありのみにする、標準病名を足す、薬剤治療を足す、検査の実施を足す、入院に絞るなどさまざまな工夫が可能である。陽性適中率を高く保ちつつ、感度も落とさないようなアルゴリズムを探す作業となるが、特に決まった基準はない。例えば陽性適中率が 8 割であれば、そのアルゴリズムで規定される症例のうち、8 割は真にその疾患であると考えられる。また感度が 5 割しかなければ、そのアルゴリズムでは真にその疾患である人のうち、半分しか拾えていないことになる。どこで十分と考えるかはそれぞれのリサーチ・クエスチョン（Research Question, RQ）次第である。

		Gold standard			
		あり	なし		
アルゴリズム	あり	a	b	陽性適中率 $a/(a+b)$	
	なし	c	d		陰性適中率 $d/(c+d)$
		感度 $a/(a+c)$			
			特異度 $d/(b+d)$		

しかし、日本の DB は完全に匿名化されていてカルテに戻れないことも多い。もし validation study を行うことができないなら、臨床的妥当性にある程度の不確実性が残ることを理解し、その不確実性をどの程度許容できるのかについても事前に十分に検討するべきである。

2.3 医薬品の価値の相対的評価

近年では有効性と安全性から評価された医薬品の絶対的価値を示すだけでなく、これまでの治療法と比較した相対的価値も求められるようになってきている。例えば医療経済的価値を検討する HTA 関連の調査解析[41]や、臨床的・主観的な価値を評価するベネフィット・リスク評価関連の調査[42]などがこれに当たる。いずれも対照治療との比較が必要となるが、実薬対照試験の結果が得られる場合以外は、対照薬の情報を何らかの方法で収集する必要がある。この際、医療情報 DB が役に立つことがある。その際、一般化可能性は良く検討しなければならない。また、申請時だけでなく、対象薬剤が市場に出ればしばらく経ってからの Real World での有用性を検討する際にも、医療情報 DB は有用なツールとなり得る。

DS 担当者 A さんの話を聞いたメディカルアフェアーズ担当者の B さんは、

「開発前から市販後まで、さまざまな場面で使えるじゃないですか！この案件も、あの案件も、ここでもあそこでも使えそうです。素晴らしいなあ！」「これからは、何でも DB で解決できるんじゃないですか？もう今までのような研究手法は必要ないんですね（キラキラ）」

A さんは DS 担当ですから、そうでもないんじゃないかなあとぼんやり思います。調査研究は目的が先にあって、そのあとそれにふさわしい手段を決めるはずで、そもそも DB で何でもできるとは言ってないし、DB を使うことありきでは逆に可能な調査の範囲が狭くなってしまふ…



どんな研究でも、なぜその研究が必要なのか、そしてこの研究の結果、何がわかることが期待され、それによって社会にどのようなメリットが期待されるのか（目的）、これまでに何がわかっている、さらに調査しなければならない事項は何なのか（データギャップ）、そのデータギャップを埋めるために、どのような対象集団で、何を曝露要因とし、どのようなアウトカムを調査する必要があるのかを十分に

検討することが重要であり、DB 研究も例外ではない。研究デザインを考えるうえで、目的に沿ったデータを 1 例ずつ集める一次データ収集がふさわしい場合もあれば、既存データの二次利用がよい場合もある。既存データの二次利用でも DB 研究ではなく、文献やその他公知情報の調査で十分な場合も多い。まずは文献調査を十分に行い、Research question (RQ) に磨きをかけて、本当にその RQ には DB 研究がふさわしいのかをよく吟味することが必要である。

それではどのような場合に DB 研究が向いているのであろうか？まず「一般的に医療情報 DB に含まれており、標準化されているデータ項目」のみを使って解決可能な RQ を考える。すなわち患者基本データ（年齢や性別など）、処方や検査・処置の実施、診断名などだけを用いた RQ であれば DB 研究の良さを活かせる可能性がある。バイタルや検査値など、もともと DB に含まれておらず、標準化もされていないデータ項目が研究に必須なのであれば、データソースからそのデータ項目を探し、抽出してクリーニングし、標準化しなければならず、それならもともとから一次データ収集を行う研究デザインを検討した方が効率的かもしれない。

その他にも一般的に DB 研究で調査しやすい薬剤やアウトカムの特徴があり、Table 2-1 に簡単にまとめた。

Table 2-1 医療情報 DB で研究しやすい薬剤やアウトカム

	得意なこと	不得意なこと
薬剤	<ul style="list-style-type: none"> ◆ 外来患者への処方 (DB の選択肢が多い) 	<ul style="list-style-type: none"> ◆ 保険償還されないもの (例: OTC) ◆ 漢方薬 (コーディングが不十分)
アウトカム	<ul style="list-style-type: none"> ◆ ほとんど必ず入院するか死亡に至るような重篤な疾患 ◆ 簡便に、迅速に、正確に診断される疾患 ◆ 急性に発症する疾患 ◆ 比較的少数の疾患コードで正確に定義できる疾患 	<ul style="list-style-type: none"> ◆ いわゆる「保険病名」になりやすい診断名を持つ疾患 ◆ 診断や重症度、効果判定などに画像診断、組織診断や検査値が必要な疾患 ◆ 症状が複雑で正確な診断が難しい疾患 ◆ 保険償還上、分けてコード化する必要のない細分類が必要な疾患 (2 型糖尿病は多くの場合単に糖尿病その他とコードされるなど) ◆ 標準化コードが存在しないような疾患

3. DB 研究の進め方

DS 担当者 A さんは悩んでいます。

上司には「RWD を取り入れていかないとな！」と言われてはいるものの、

まだ社内体制がばっちり整っている状況とは言えません。

でもはじめの一步を踏み出さなければ経験から学ぶこともできない気がします。



DB 研究か否かに関わらず、そもそも社内に「企業で自ら科学的な臨床研究を行い、その結果を論文等で公表し、患者のより良い未来に貢献しよう」という認識がなければ、その重要性を説くところから始めなければならない。また、部署によっては決められた研究デザインの枠組みの中だけで考えるのではなく、「RQ に応じて研究のセッティングやデザインを選択する」という意識改革が必要である。さらに、研究というと「医療機関と契約する前向きな臨床研究」という認識が強い企業では、RWD でのエビデンス創出、DB 研究という新たなフレームワークを理解してもらうことも重要である。

3.1 社内体制

臨床開発をはじめ、メディカルアフェアーズや HEOR (Health Economics and Outcomes Research)、マーケティング部門などでも、疾患領域あるいは各製品の担当者はそれぞれ「こんなデータ/エビデンスがあるといいなあ」という RQ を持っている (はずである)。RQ は個人や単一部署で検討・解決される場合もあるだろうが、調査研究を実施するとなれば、他の研究の例にもれず、DB 研究も様々な専門家から構成されるチームを結成し、検討を進めることになる。医薬品医療機器等法の下で行う治験や製造販売後調査・試験以外の、普通の臨床研究の計画や実施に関する SOP が社内でもまだ整備されていなければ、DB 研究も念頭に置いて部門横断的に検討するのがよいだろう。

DB 研究実施において理想的な社内体制は下記のように考えられる。

1. 経営陣・上層部が RWD によるエビデンス創出の重要性を認識している
2. 社内に DB 研究の専門家または経験者がいる
3. 社内で DB や RQ の検索ツールが利用可能な状況にあり、複数の部門で活用されている
4. 臨床研究・疫学研究が数多く実施されていて、研究予算が確保しやすい
5. SOP や社内レビュー体制など、研究を進めるうえでの基本体制が整っている

DB 研究の前例がない企業では、DB ベンダや DB 研究団体が開催しているセミナーや研修会への参加で、多くの有用な情報・活用事例が得られるはずである。近年は日本の企業や研究者による DB 研究論文も多く公表されている [9, 11, 12, 14-27, 43-66]。また、DB ベンダもニュースレターなどで自社の DB を用いた論文や解析事例を定期的に発信している。お目当ての DB があれば、そういったニュースレターからの情報入手が効率的である。特に、解析事例を日本語で解説してくれているので、自身の RQ に近い事例は大いに参考になるであろう。

外資系企業では、海外本社に疫学やアウトカムリサーチ、医療経済など、以前から RWD を活用してきた専門部署があり、さらには DB 研究に特化した部門もあることが多い。大規模なデータを扱うために IT のサポートが必要だったり、一次データ収集とは異なるデータマネジメント、データ解析の方法があったりするためである。彼らは日々世界中の利用可能な DB を調査し、DB を活用したさらなる研究デザインや解析手法の開発に努めている。

3.2 研究フローと事前準備

ここからは具体的な DB 研究の進め方について概説する。新たに DB ベンダと交渉を始める状況を念頭に置いてごく簡単にフレームワークを説明すると、Figure 3-1 のようになる。

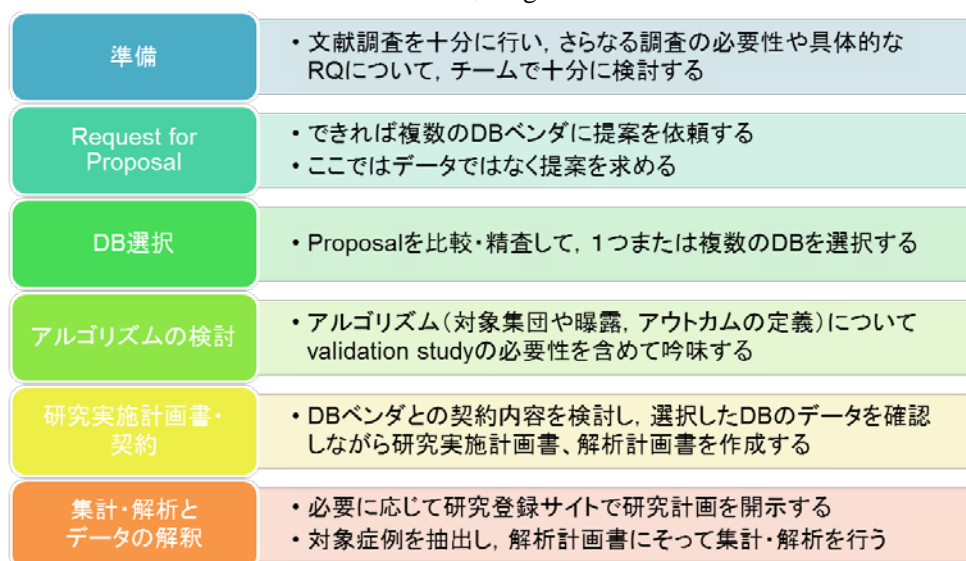


Figure 3-1 DB 研究実施フロー

メディカルアフェアーズ担当の Bさんは薬剤 X を担当しています。

「薬剤 X を服用すると、入院が減るはずですよ。RWD を見てみましょうよ！」

確かに、入院率の低下までは治験では見ていませんし、これまでに世界で論文も出ていないようです。学会等でもこのテーマの研究が進んでいるという話は聞こえてきませんが、国内外の研究登録の DB にも登録されていないようです。入院が減りましたか？というマーケットリサーチには回答者バイアスがかかりそうですし、この疾患に関する入院サーベイランスも実施されていません。市販後調査は実施中ですが、結果が出るまでまだ時間がかかりそうです。

Aさんはこの「薬剤 X の服用で入院が減る」という RQ について真面目に検討し始めます。

「薬剤 X は外来処方されているし、入院というアウトカムは医療情報 DB で研究しやすいみたいだ。でもいきなり自社製品で『何かと比べて入院が抑制される』という仮説を検証するにはリスクがあるし、まずは直接比較を目的とせずに、発売前後の様子を可能な範囲で観察してみることでどうだろう…」



3.3 Request for Proposal

製薬企業によっては、特に外資では、DB を丸ごと購入し、定期的に更新し、社内で標準化して複数のデータソースを一度に検索できるようなシステムを備えているところもあるだろう。しかし、日本の現状では DB ベンダがデータの所有者であることの方が多いと考えられる。DB を保有しているものがそのデータについてもっともよく知っているのは当然であるから、素直に何のためにどんなことが知りたいのかを相談するのがよい。

この時、DB ベンダに対して書くのが Request for Proposal (RfP) である。この際、秘密保持契約が必要となる企業もあるだろう。日本ではまだ RfP 対応の経験の少ないベンダもあるかもしれないので、あらかじめ問い合わせをして、どのように協議を始めるか、RfP を受け付けてもらえるか、その場合どのくらいの期間で提案をもらえそうか、確認しておくことと安心である。RfP は複数の候補ベンダに送ることが肝要である。相見積もりを取る必要がある企業もあるだろうし、比較することによってなぜそのベンダ (DB) を選んだのか、説明しやすくなるメリットもある。

DB ベンダからは各社の DB の特徴を踏まえた提案がなされる。リクエストにあった変数 (たとえば検査値など) が DB に無い場合やデータ数が限られている場合、ベンダは豊富な分析実績をもとに、代替案を提示してくれることもある。また、自社で解析を行うことを予定している場合には、納品されるデータのイメージや具体的な構造を事前に把握するために、数例のサンプルデータの提供を依頼することもできる。なお、ベンダ各社とも継続的にデータの網羅性、データの粒度、継続性を向上させているので、各社の DB の特徴を把握していたとしても、定期的に情報を入手することをお勧めする。

■ □ コラム □ ■ Request for Proposal

RfP (アール・エフ・ピー) は、DB 研究に特有なものではなく、業務委託を行う際に発注先候補の業者に具体的な提案を依頼する文書であり、特定のフォーマットがあるわけでもない。研究概要、特に何のために、何がしたいのか、今回の研究で何が一番重要なのか、依頼者側の思いをできるだけ具体的に伝えることが肝要である。逆に無理に定義を固めすぎた RfP を送ると、ベンダ側としては柔軟な提案ができず、不自由を感じることもある。お互いに最初から完璧な書類の作成を目指すより、わからないことは素直に尋ねながら十分に話し合うことで研究がスムーズに進んでいく。

A さんは、社内の医学専門家や疫学担当者だけでなく、治験で多くの施設のデータを見てきた開発担当者や large data のハンドリングに強そうなプログラマなど、多くの同僚を巻き込みながら研究チームを作り、先行研究や社内資料を調査して研究概要を考え、DB の選択にあたり必要となりそうな検討項目について、Table 4 のようにまとめました。これを Request for Proposal として DB ベンダに送ろう、という計画です。

総合的に判断して、今回は健康保険組合由来のレセプトの DB を提供する P 社と病院ベースの administrative DB を提供する Q 社の 2 社に送ることにしました。

A さんは、DB 研究が初めてで、検査結果や症状のデータがなさそうなのが不安だと打ち明けると、ベンダの担当者は「重症度によって薬剤やその用量など治療法が異なるのであれば、検査値でなく、治療データから重症度を規定することもできるかも知れませんよ。そのあたりも検討してご提案させていただきますね」と言ってくれました。



Table 3-1 Request for Proposal

項目	検討事項
目的：	疾患 Y の重症患者について，薬剤 X 上市前の集団および，上市後の X 曝露群での入院率を調べる
患者数：	調査期間に DB で得られる対象患者数が知りたい
網羅性：	主に外来診療されており，重症患者でも診療所でも治療されているので，その病院と診療所の別を知りたい (病院ベースの DB では他の疾患を合併している患者など，より複雑な背景を持つ群に偏るかもしれない)
代表性：	疾患 Y は高齢者に多い疾患なので，年齢別の分布が知りたい (あまり重症だと働き続けられない可能性もあるため，社会健康保険のレセプトでは一般化可能性が低いかもしれない)
対象集団：	疾患 Y 特定の ICD-10 で定義した場合，それに特定の標準病名を追加した場合など，複数のアルゴリズムを提案してほしい
曝露：	薬剤 X は経口剤で院外処方されることも多い．処方，調剤，処方期間に加え，包括加算などの影響で実際は処方があったのに記録上は見えなくなっている処方がどの程度あるのか知りたい
アウトカム：	疾患 Y 重症患者の増悪時における入院が取りたいが，all-cause hospitalization でも妥協できる
その他変数：	重症度は検査値や症状スコアで分類される．重症度そのものの情報や，特定の検査結果および症状について，どのような情報がどの程度得られるのか知りたい．またこれらの情報が直接得られない場合に，よい代替案があれば提案してほしい (検査値や症状は一般的にはあまり期待できないと聞いているが，どの程度ないのか確認したい)
追跡期間：	薬剤 X は 2013 年 2 月発売で，最長でも 2 年強しか追えないので，最新データがいつ更新されるのか知りたい．また，上市前の重症 Y 患者および X 曝露群それぞれの DB 内での追跡期間 (月) のヒストグラムを示してほしい (普段から病院に通院していたとしても，入院するときにも同じ施設なのかなあ)
() 内は A さんの心の声	

もちろん提案の締切や，おおよそのコストのような事務的な情報や，これまでの同分野での経験や論文一覧など，研究実績を問う質問など，他の分野の RfP と同様に適宜追加すればよい。

3.4 DB の選択

ベンダ 2 社から提案が届きました。A さんはこれらの提案を Table 3-2 のようにまとめてみました。

B さんはウキウキしています。

「どちらの DB でもできそうじゃないですか! 疾患 Y の重症患者さんがこんなにいるなんて，やっぱり DB っ

ていいなあ。どっちにしようかな～、両方にしちゃおうかな～」

Table 3-2 ベンダからの提案例のまとめ

項目	検討事項	P 社)	Q 社*)
目的：	疾患 Y の重症患者について、薬剤 X 上市前の集団および、上市後の X 曝露群での入院率を調べる	保険加入者×期間を分母とし、入院を分子とすることで推計可能と考えます	入院時だけの来院を考慮すると、ある一定期間病院に通院している症例を対象とすることを提案します。この中での入院イベントとして把握可能と考えます
患者数：	調査期間に DB で得られる対象患者数が知りたい	Y を仮に ICD-10 で定義すると、2008 年以降で約 1,500 人	Y を仮に ICD-10 で定義すると、2008 年以降で約 10,000 人
網羅性：	主に外来診療されており、重症患者でも診療所でも治療されているので、その病院と診療所の別を知りたい	両方を受診している場合には病院と定義すると、6 割は診療所となります	診療所のデータはありません
代表性：	疾患 Y は高齢者に多い疾患なので、年齢別の分布を知りたい	(ヒストグラムを見ると高齢者は限られている)	(ヒストグラムを見ると約半数は 65 歳以上)
対象集団：	疾患 Y 特定の ICD-10 で定義した場合、それに特定の標準病名を追加した場合など、複数のアルゴリズムを提案してほしい	標準病名として yy などを試行的に検索しましたが、ICD-10 のみの場合と大きな違いは見られませんでした	検査や治療のためにこのコードを付ける実態はないようですので、ICD-10 での定義でよいかと思えます
曝露：	薬剤 X は経口剤で院外処方されることも多い。処方、調剤、処方期間に加え、包括加算などの影響で実際は処方があったのに記録上は見えなくなっている処方がどの程度あるのか知りたい	院外・院内、処方・調剤ともに記録がありますが、レセプトのみですのでご質問の点は確認できません	薬剤 X について、レセプトおよび明細を比較しましたが、薬剤 X の処方についてはその差は<1%でした
アウトカム：	疾患 Y 重症患者の増悪時における入院が取りたいが、all-cause hospitalization でも妥協できる	入院患者で疾患 Y の ICD-10 がある症例とします	DPC 上で入院契機または主病名に疾患 Y があるものを取得することは可能ですが、過小評価する可能性があるため、入院すべ

			て、Yを含む入院、上記カラムに Y を含む入院など感度分析を提案します
その他変数：	重症度は検査値や症状スコアで分類される。重症度そのものの情報や、特定の検査結果および症状について、どのような情報がどの程度得られるのか知りたい。またこれらの情報が直接得られない場合に、よい代替案があれば提案してほしい	疾患 Y に関する重症度、検査値、症状スコアのデータはありません。重症の場合には薬剤 Z で治療される例が8割以上との文献情報があるので、多少過小評価にはなりますが、Z の曝露で定義することは可能かと思えます	全施設のうち、15%の施設から検査値データセットもお預かりしており、さらにその半数からこの疾患に関する検査結果が取得可能です。施設数は限られますが、
追跡期間：	薬剤 X は 2013 年 2 月発売で、最長でも 2 年強しか追えないので、最新データがいつ更新されるのか知りたい。また、上市前の重症 Y 患者および X 曝露群それぞれの DB 内での追跡期間 (月) のヒストグラムを示してほしい	最新データは6か月前となります。(ヒストグラムを見ると疾患 Y+薬剤 Z の中央値は 3.5 年、疾患 Y+薬剤 X の中央値は 1 年)	疾患 Y を仮に ICD-10 で定義する場合、DB 中での追跡期間の中央値は 3 年、X 曝露群では 1 年でした
*病院経営関連以外のデータセットも一部保有する			

DB ベンダからの提案をチームでよく検討して、今回の RQ に適切な DB を選定する (1 つとは限らない)。ベンダからの提案は、特に指定しなければ様々な形式で送られてくる。データセットの概要や構成要件、調達条件とともに、これまでの当該分野での実績や、ベンダ内の専門家の紹介や研究実施体制などについても紹介があるかもしれない。書類上の条件だけに基づいて DB を選択すると、実際に研究を進めていくうちに依頼者側との理解の齟齬が生じることも多いので、できるだけ実際に面談を行い、インタビュー形式で詳細を確認することをお勧めする。上がってきた提案に DB 選定に必要な情報が全て含まれているとは限らないし、一度で完成に至らなくても、ディスカッションを重ねることで各社の DB の特徴について理解を深めたり、データの限界点に気付いたりすることもある。最終的な提案の詳細は、研究にも大きな影響があると同時に、購買契約時にも重要になる。

さらに一つの研究に対して、使用する DB は一つとは限らない。各 DB に特徴があり、私たちの RQ に完璧に対応できる DB はまず存在しないという前提で考えると、場合によっては一つの研究の中で異なる DB から導かれる結果が共存することは十分考えられる。

以下に、DB 選択におけるいくつかのポイントを示す。

DBに含まれる集団の特性

DB を構成するのはどのような特徴をもった人たちだろうか？それは DB のデータベースに依存する。例としては下記のような DB があり、それぞれの特性から研究デザインに照らし合わせると向き・不向きがある。さらに一般化可能性や情報バイアスにも注意を払う必要がある。

- ◇ 健康保険組合からデータを収集している DB: 基本的には働いている人とその扶養家族からなる集団で、診療行為がない期間も被保険者としての情報を利用できる（分母がある）。健康保険組合の種類により、構成年齢が大きく異なる。社会保険健康組合は、被保険者が定年に達する年齢付近から加入者数は減少し、75 歳以上になると後期高齢者制度に移行するため、75 歳以上のデータはない(2015 年現在)
- ◇ 医療機関からデータを収集している DB: 病院の中にある多くの DB の中からいくつかの DB（多くの場合は病院経営に役立つ保険請求や会計関連の DB）で、登録施設を受診した人の集団。

データの規模

ベンダの提供する DB は全体としては大規模だが、アルゴリズムによっては対象者が数十例になってしまうことも少なくない。そしてそれが契約後に初めて明らかになることもある。契約後でもめないためにも、疾患定義を丁寧に行い、やり取りを繰り返し、慎重に確認する必要がある。例えば Table 3-2 でも、研究期間を通じて疾患 Y の ICD-10 コードが一度でも付与されている患者数ではなく、必要な期間継続して追跡できる患者数を把握する必要があるだろう。さらに、この 1,500 人が今回の研究で解析可能な最大患者数であり、条件をつけていくにつれて減ることはあっても増える可能性はまずない。

得られる情報（データではなく、情報）の確度

DB のデータは、何らかの人為的ミスで欠損することはまずない。しかし、データの解釈には注意が必要である。有名（？）な例として、検査や投薬のためにレセプト上につけられた病名（いわゆる保険病名）があり、その疾患コードが保険病名である可能性を見極める必要がある。逆に保険償還上わざわざ分けてコードする必要のない疾患も多くある。例えば糖尿病は治療者も患者も 1 型か 2 型か把握しているが、圧倒的多数は「糖尿病その他」とコードされており、レコード上は 1 型・2 型の区別がつかない、感染症も呼吸器や胃腸系など症状を主体に診断名（コード）にとどまり、原因菌やウイルスを特定するコードの使用が保険償還上必須ではないことがうかがえる。また薬剤間の比較で、ベースラインの群間差は前提条件として重要な検討事項となるが、自覚症状や重症度、喫煙などの生活習慣や治療に対する積極性などの特性の違いは DB のデータからわかることは少ない。ベンダの DB は原資料には戻れないことが多いが、画像診断、検査結果が一部でも取得可能であれば、これらの確認結果を gold standard とした validation study についても検討する。

バイアスの可能性

他の研究同様、バイアスの可能性に関して事前検討が必要である。情報バイアスの代表的なもの 1 つである診断バイアスは、DB 研究では少し違った意味を持つ。すなわち、診断者間の診断の違いや不正確な診断に基づくのではなく、「DB 内でその疾患をどのように定義するか」に依存する誤分類が起こり得る。例えばある疾患の患者群を「特定の ICD-10 コードあり、同月に同疾患の治療薬あり」と定義する

と、治療されなかった（軽傷な）患者は対象とならないことになる（系統的な誤分類）。結果の一般化可能性に影響する選択バイアスとしては、例えば健康保険組合由来のレセプト DB を使用する場合は、健康労働者効果（Healthy worker effect）が挙げられる。健康保険組合由来のレセプト DB には社会保険組合に加入できる人（それなりの労働条件で勤務している人）およびその扶養家族が含まれる。つまり、基本的には健康な人およびその子供であり、重い病で退職を余儀なくされた患者などは健康保険組合から脱退してしまうかもしれない。また定年を機に保険が変わる人もいる。日本の医療制度上、75 歳以上は後期高齢者医療制度に移行するため、社会保険組合の DB では 75 歳以上の対象者は含まれていないため、DB に存在しているのは、比較的健康的な若い集団ということになる。高齢者に多い疾患や就労条件に影響する可能性のある疾患を研究対象として健康保険組合由来のレセプト DB を使用する場合には、結果の一般化可能性に注意が必要である。一方で大病院由来のレセプト DB には、診療所での患者群と比較して重症・難治の患者が多く含まれている可能性があり、もし一般化可能性の高いデータを求めている場合にこのタイプの DB を選択するならば、同様に結果の解釈に注意が必要である。

また、意図した対象集団や曝露、アウトカムを過不足なく、偏りなく収集できているか、つまり情報バイアスも起こり得る。例えば入院をアウトカムとして取得したい場合に clinical DB を選択すると、カルテ ID を医療機関を超えて連結できないことから、対象として選択した患者が普段通院している病院ではなく、自宅近くの別の病院に入院した場合には、入院のアウトカムを検出することはできず、入院のリスクを過小推定してしまう可能性があるし、逆に入院時にだけ対象施設に来院した患者を拾ってしまつて、過大推定になる可能性もある。

契約上の留意点

提案をよく比較・吟味したのちに DB ベンダとの契約に進む。データセットを丸ごと購入する場合もあるだろうし、対象データセットをサブセットとして購入することもあるだろう。あるいは万一、解析中に自社品の安全性に関する情報を入手してしまう可能性などを配慮して、社内ではデータを極力持たず、集計解析もベンダに依頼する場合もあるだろう。実際に進めていくうちに、RfP にも提案にも含まれていなかった問題点が明らかになったり、目的や対象集団の変更を余儀なくされたりすることもあるので、ベンダと何度かやり取りを重ねながら進められるような、スパイラル的な進め方を契約時にも考慮することを推奨する。また、アカデミアとの共同研究（または委託研究）にする場合には、社内の規定に応じてアカデミアとの契約も必要となる。

3.5 研究実施計画書

今回の RQ にはどちらでも研究ができそうに思いましたが、どちらを選んでも何かが足りない気がしてきました。両社の担当者にいろいろと説明を聞いて、今回の RQ では P 社の DB で研究が出来そうだと考えました。validation study は他国の DB で「ICD-10 の特定コードのみの定義で陽性適中率も感度も十分であった」という報告があり、今回はカルテに戻ることができない DB を選んだこともあり、DB の他の項目での validation も難しそうだったので、限界として認識しつつ、今回は行わないことにしました。ちょうど研究予算的にも可能だし、社内で正式に研究提案してみよう… A さんは、未知の世界に踏み込んでいく感覚を味

わっています。

ベンダとやり取りを続けながら研究実施計画書作成に着手しましたが、DBのことを知れば知るほど、限界が分かってきました。しかしその度にDBベンダの担当者から目からウロコの提案があり、「なるほど!」と進めていくうちに、研究実施計画書は形になっていきました。

DB研究にも研究実施計画書は必要である。基本的には臨床研究と同じと考えてよいが、DB研究では「データに合わせて研究実施計画書を書く」という特徴がある。FDAのガイダンスが記載事項の参考になるだろう[67]。DB研究に特徴的な研究計画書に記載すべき事項を下記に挙げる。

使用するDBの説明

この研究にこのDBを選んだ理由を、データソース、規模、年齢構成、変数（検査値）の有無などで説明する。場合によっては、DBベンダに実施可能性の検討をした結果も書くこともある。

限界

カルテに戻れない特徴をもつDBを使う場合、前述の保険病名問題など限界は多く挙げられる。また、DBのデータソースによる限界もある。たとえば社会保険データをソースにもつ保険レセプトDBでは、65歳以上の年齢層が極端に少なくなる。希少疾患などについてはDB上から疾患コードなどが削除されている場合もある。さらには、保険診療データにはない重要な交絡要因（喫煙状況など）も得られない。これらの特徴が研究結果におよぼす影響（と対応・解釈）を記述する。

対象集団の定義（組入れ・除外基準）

DBでは、ある疾患の初診時からのデータがそろっているとは限らず、また初診以降DBに入るまでの診療状況が不明であるため、その人のデータがDBに入ってから一定期間のデータを観察して研究に組み入れることが重要である。また、過去にある疾患を発症していない、あるいはアウトカムに該当するイベントがない集団であることを定義するために、組入れ前の一定期間アウトカムに定義した疾患コードが付与されていないことなどを考慮することもある。

曝露やアウトカムの定義

DB研究の難所ともいえるところである。アウトカムの定義（アルゴリズム）が複雑であれば図や表などで整理すると理解されやすい。疾患に関しては、ICD-10コードの定義が最も多いが、前述のように保険病名が付与されている可能性が多い疾患では、薬剤処方・処置などのコードを組み合わせて定義することもある。アウトカムは入院・手術・人工透析導入などの保険病名である可能性が少ない診療行為（処置）コードから定義することができればよいが、「重症度の変化」などを観察したい場合には、研究内で定義した重症度が、研究で検討していない他の要因に影響されていないかなどの配慮が必要となる。定義に困ったら、DBベンダでは豊富な研究実績が蓄積されているので、相談すると解決策が得られるかもしれない。また、診療報酬制度に詳しい人や医学専門家に確認してもらうことも考慮する。

サンプルサイズ

一般的には hypothesis generation の目的で行うことが多く、「DB から得られるすべての該当者」を対象とすることが多いが、サンプル数が非常に大きくなることもあり、その場合には臨床的に意味がないようなほんの小さな差異も検出できてしまうので、結果の解釈には十分に注意が必要である。逆に大規模な DB を使用しても、定義の精度を上げようとして多くの制約をつけると、解析可能症例が少なくなることもある。DB 選定は (hypothesis generation 目的の場合でも) 統計的根拠に基づき必要最低限の症例数を確認した上で行うのが望ましい。仮説検証が目的であれば必要なサンプルサイズを通常どおり計算して、十分なデータ数が確保できるかを DB ベンダに確認したうえで、DB 選定の根拠に付け加えるようにする。

昨今では臨床試験と同様、観察研究にも透明性が求められており、研究登録ウェブサイトにて事前にプロトコルを開示する動きが広まっている。特に EMA は、市販後の安全性の研究については登録を義務付けている[68-70]。国内外の代表的な登録サイトを下記に挙げる。企業によっては、自社ウェブサイトを含め、登録サイトを規定している場合もある。

- ◇ 大学病院医療情報ネットワーク研究センター (UMIN: <http://www.umin.ac.jp>)
- ◇ Clinicaltrial.gov (<https://clinicaltrials.gov/>)
- ◇ EU PAS Register (http://www.encepp.eu/encepp_studies/indexRegister.shtml)

A さんは、社内外の協力者と協議を重ねました。時にはアウトカムの定義について医学専門家と熱く語り合い、また、時には「高齢者のデータもほしかったなあ…」というメディカルアフェアーズの担当者を説得し、日夜解析を続けました。DS 部の仲間には、「今後、こういう DB 研究を積極的に実施していく時代になるから!」とレセプトデータについて成り立ちと構成を理解してもらい、データクリーニングやコーディング確認といったノウハウ蓄積をお願いしました。そして、RQ に対する結果が見えてくると、にわかに協力者の間に連帯感が高まりました!

■ □ コラム □ ■ DB 研究実施のための倫理審査

現在は、前向きにデータを収集する臨床研究では倫理審査はほぼ必須となっているが、DB 研究ではどうだろう? 日本の DB ベンダのデータは人を対象とする医学系研究に関する倫理指針でいうところの、「既に連結不可能匿名化されている情報」に該当するが、医療施設や先生方との共同研究・委託研究にならない場合には、誰に倫理審査をしてもらうのか? など、実際に研究に着手する場合には重要な検討事項の一つである。

3.6 解析と結果の解釈

DB 研究では、臨床試験の解析業務とは異なり、ICD-10 コードや、診療報酬制度の仕組み、医療機関内のデータの構造など、DB 研究に特有の背景情報に精通していることが必要となる。例えば administrative DB からデータのサブセットを受け取るような契約をした場合には、ベンダ毎に特有のデータフォーマットでデータが納品される。研究という観点で考えると、社内で DB そのもの、あるいはデータのサ

ブセットを所有して解析できるとスムーズではあるが、社内に専門家・経験者がいない場合や、透明性の観点から、企業の方針として個人レベルのデータ入手を最小限にする場合や、まだ手順書がないためにデータ所有や解析が難しいという場合もあるだろう。その際には、DB ベンダに集計や解析までまとめて依頼したり、DB 研究の解析経験のある CRO に委託することも可能である。DB 研究の経験がない CRO に外注せざるを得ない場合には、データセットの構造や変数の意味、ときには保険診療のしくみについて、サンプルデータなどを用いて説明・指示する必要がある。また、臨床試験と異なり、実臨床データである RWD では様々なデータが混在しているので、想定外のデータにどう対応するかも重要である。いずれにしても社内である程度の解析の知識は必須である。

データの集計や解析はそれぞれの研究計画書、解析計画書に基づいておこなう。その信頼性を担保するためにダブルプログラミングを行ったり、頑健性を確認するために感度分析を行ったり、データの二次利用で可能な限り妥当性を高める工夫が必要である。理由なく都合の悪いデータを除いたり、いい結果が出るまでサブグループを作り続けたりしてはいけけないのは、どんな研究でも同じである。Web-based のサービスでダウンロードしたデータの場合には、再現可能性が担保できないため論文文化には注意が必要である。

また、医療情報を公共のリソースと考え、社内に結果をとどめておくのではなく、すこしでも公共の利益に資するように学会や論文などでの公表を前提に研究を進めるのが望ましい。

とうとう会社初の DB 研究の報告書がまとまり、A さんは上司に最終報告に行きました。リアル・ワールド・データについての解説だけでなく、社内にあった RQ に対して DB 研究の論文文化が進んでいると報告しました。研究実施に関してベンダの担当者に教えて頂いたこと、医学専門家とのディスカッション、データサイエンス部の同僚の協力などは中間報告で熱く語っていましたが、いよいよ論文文化と聞いて、上司は「リアル・ワールド・データについてちょっと調べてもらおうと思って声をかけたのだが、わが社の製品に有用なエビデンスまで生み出せたとは想定外でうれしいぞ!」とこの結果を喜んでくれました。

また、メディカルアフェアーズ部門では、B さんの製品の解析結果を共有し、他製品でも DB 研究の実施可能性を考えようという気運が生まれています。A さんと B さんは、早くこの結果を論文発表してこれからも DB 研究を続けていきたいね〜と打ち上げをしています。



おわりに

「Real World Data の活用」として、医療情報データベース (DB) を二次活用する研究について、その基本と実際をごく簡単に述べた。医療情報 DB は魔法の杖ではないし、かといって単なる情報の寄せ集めでもない。結果は決して決定的なものではなく、解釈には十分な注意が必要である。データホルダである医療情報 DB ベンダを良きパートナーとし、それぞれの DB の特徴や限界をデータに基づいて把握したうえで、適切に利用すれば DB は強力なツールになり得る。本報告書が医療情報 DB の適切な利活用の一助となることを祈っている。

References

1. 独立行政法人医薬品医療機器総合機構, 医療情報のデータベース等を用いた医薬品の安全性評価における薬剤疫学研究の実施に関するガイドライン [<http://www.pmda.go.jp/files/000157340.pdf>]. 2014.
2. 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/standardMenu/>].
3. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP), Guide on Methodological Standards in Pharmacoepidemiology (Revision 3): 3.2. Secondary use of data. 2014.
4. 一般社団法人 National Clinical Database, [<http://www.ncd.or.jp/>].
5. Nakashima, N., [Japanese sentinel project and contribution of laboratory medicine]. *Rinsho Byori*, 2013. **61**(6): p. 501-10.
6. Ooba, N., et al., Claims-based definition of death in Japanese claims database: validity and implications. *PLoS One*, 2013. **8**(5): p. e66116.
7. 日本薬剤疫学会 薬剤疫学とデータベースタスクフォース, 日本における臨床疫学・薬剤疫学に応用可能なデータベース調査 [<http://www.jspe.jp/committee/020/0210/>].
8. 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・特定健診等情報の提供に関するガイドラインの改正等について [<http://www.mhlw.go.jp/stf/shingi2/0000082062.html>].2015.
9. Kubota, K., et al., Epidemiology of psoriasis and palmoplantar pustulosis: a nationwide study using the Japanese national claims database. *BMJ Open*, 2015. **5**(1): p. e006450.
10. 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**: p. 471-481.
11. 飯原なおみ, et al., 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**(2): p. 67-77.
12. 奥村泰之, 野田寿恵, and 伊藤弘人, 日本全国の統合失調症患者への抗精神病薬の処方パターン: ナショナルデータベースの活用. *臨床精神薬理*, 2013. **16**(8): p. 1201-1215.
13. 關真美 and 椿広計, サンプル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*, 2014. **34**(6): p. 293-304.
14. Fukuhara, T., Geographical analysis of aneurysmal subarachnoid hemorrhage in Japan utilizing publically-accessible DPC database. *PLoS One*, 2015. **10**(3): p. e0122467.
15. Masuda, K., et al., Factors affecting the occurrence of pulmonary embolism after spinal surgery: data from the national administrative database in Japan. *Spine J*, 2012. **12**(11): p. 1029-34.
16. Murata, A. and S. Matsuda, Circumstance of endoscopic and laparoscopic treatments for gastric cancer in Japan: A review of epidemiological studies using a national administrative database. *World J Gastrointest Endosc*, 2015. **7**(2): p. 121-7.
17. Isogai, T., et al., Out-of-hospital versus in-hospital Takotsubo cardiomyopathy: analysis of 3719 patients in the Diagnosis Procedure Combination database in Japan. *Int J Cardiol*, 2014. **176**(2): p. 413-7.
18. Isogai, T., et al., Serious cardiac complications in coronary spasm provocation tests using acetylcholine or ergonovine: analysis of 21 512 patients from the diagnosis procedure combination database in Japan. *Clin Cardiol*, 2015. **38**(3): p. 171-7.
19. Isogai, T., et al., Hospital volume and cardiac complications of endomyocardial biopsy: a retrospective cohort

- study of 9508 adult patients using a nationwide inpatient database in Japan. *Clin Cardiol*, 2015. **38**(3): p. 164-70.
20. Niikura, R., et al., Factors affecting in-hospital mortality in patients with lower gastrointestinal tract bleeding: a retrospective study using a national database in Japan. *J Gastroenterol*, 2015. **50**(5): p. 533-40.
 21. Odagiri, H., et al., Factors associated with perforation related to diagnostic balloon-assisted enteroscopy: analysis of a national inpatient database in Japan. *Endoscopy*, 2015. **47**(2): p. 143-6.
 22. Odagiri, H., et al., Hospital volume and the occurrence of bleeding and perforation after colorectal endoscopic submucosal dissection: analysis of a national administrative database in Japan. *Dis Colon Rectum*, 2015. **58**(6): p. 597-603.
 23. Sako, A., et al., Prevalence and in-hospital mortality of gastrostomy and jejunostomy in Japan: a retrospective study with a national administrative database. *Gastrointest Endosc*, 2014. **80**(1): p. 88-96.
 24. Sato, M., et al., Acute liver disease in Japan: a nationwide analysis of the Japanese Diagnosis Procedure Combination database. *J Gastroenterol*, 2014. **49**(3): p. 547-54.
 25. Sugihara, T., et al., Regional, institutional and individual factors affecting selection of minimally invasive nephroureterectomy in Japan: a national database analysis. *Int J Urol*, 2013. **20**(7): p. 695-700.
 26. Suzuki, S., et al., Impact of systemic steroids on posttonsillectomy bleeding: analysis of 61 430 patients using a national inpatient database in Japan. *JAMA Otolaryngol Head Neck Surg*, 2014. **140**(10): p. 906-10.
 27. Takeuchi, M., et al., Descriptive epidemiology of children hospitalized for inflammatory bowel disease in Japan: Inpatient database analysis. *Pediatr Int*, 2015. **57**(3): p. 443-8.
 28. Springate, D.A., et al., Can analyses of electronic patient records be independently and externally validated? Study 2--the effect of beta-adrenoceptor blocker therapy on cancer survival: a retrospective cohort study. *BMJ Open*, 2015. **5**(4): p. e007299.
 29. Hall, G.C., et al., Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf*, 2012. **21**(1): p. 1-10.
 30. Prieto, L., et al., Evaluation of the effectiveness of risk minimization measures. *Pharmacoepidemiol Drug Saf*, 2012. **21**(8): p. 896-9.
 31. Kang, E.M., et al., Evaluating the validity of clinical codes to identify cataract and glaucoma in the UK Clinical Practice Research Datalink. *Pharmacoepidemiol Drug Saf*, 2015. **24**(1): p. 38-44.
 32. Quint, J.K., et al., Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). *BMJ Open*, 2014. **4**(7): p. e005540.
 33. Thomas, K.H., et al., Validation of suicide and self-harm records in the Clinical Practice Research Datalink. *Br J Clin Pharmacol*, 2013. **76**(1): p. 145-57.
 34. Lee, J.K., et al., Development and validation of an algorithm for classifying colonoscopy indication. *Gastrointest Endosc*, 2015. **81**(3): p. 575-582 e4.
 35. Kawatkar, A., et al., Development and validation of algorithms to identify acute diverticulitis. *Pharmacoepidemiol Drug Saf*, 2015. **24**(1): p. 27-37.
 36. Lawrence, J.M., et al., Validation of pediatric diabetes case identification approaches for diagnosed cases by using information in the electronic health records of a large integrated managed health care organization. *Am J Epidemiol*, 2014. **179**(1): p. 27-38.
 37. Thyagarajan, V., et al., Identification of seizures among adults and children following influenza vaccination

- using health insurance claims data. *Vaccine*, 2013. **31**(50): p. 5997-6002.
38. Cheng, C.L., et al., Validation of acute myocardial infarction cases in the national health insurance research database in taiwan. *J Epidemiol*, 2014. **24**(6): p. 500-7.
 39. Cheng, C.L., et al., Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. *Pharmacoepidemiol Drug Saf*, 2011. **20**(3): p. 236-42.
 40. Yamaguchi, T., et al., The Epidemiological Study of Venous Thromboembolism and Bleeding Events Using a Japanese Healthcare Database - Validation Study -. *Jpn. J. Drug Inform.*, 2015. **17**(2): p. 87-93.
 41. 中島章博, 田口奈緒子, and 奥山ことば, 【企画セッション】 データサイエンス担当者のための費用対効果評価講座. *SAS ユーザー総会論文集*, 2015.
 42. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース, ベネフィット・リスク評価入門 [<http://www.jpma.or.jp/information/evaluation/allotment/benefitrisk.html>]. 2013.
 43. 宮川直子, et al., レセプト情報・特定健診等情報データベースを利用した滋賀県における循環器疾患危険因子の有病率, 治療率, コントロール率. *日本公衆衛生誌*, 2014. **7**(61): p. 333-341.
 44. Fujimoto, M., et al., Association between statin use and cancer: data mining of a spontaneous reporting database and a claims database. *Int J Med Sci*, 2015. **12**(3): p. 223-33.
 45. Takada, M., et al., Association of statin use with sleep disturbances: data mining of a spontaneous reporting database and a prescription database. *Drug Saf*, 2014. **37**(6): p. 421-31.
 46. Okumura, Y., et al., Exposure to psychotropic medications prior to overdose: a case-control study. *Psychopharmacology (Berl)*, 2015. **232**(16): p. 3101-9.
 47. Okumura, Y., J. Fujita, and T. Matsumoto, [Trends of psychotropic medication use among children and adolescents in Japan data from the national insurance claims database between 2002 and 2010]. *Seishin Shinkeigaku Zasshi*, 2014. **116**(11): p. 921-35.
 48. Gotoh, M., et al., National Clinical Database feedback implementation for quality improvement of cancer treatment in Japan: from good to great through transparency. *Surg Today*, 2015.
 49. Murata, A., et al., Recent Change in Treatment of Disseminated Intravascular Coagulation in Japan: An Epidemiological Study Based on a National Administrative Database. *Clin Appl Thromb Hemost*, 2015.
 50. Ohisa, M., et al., Estimated numbers of patients with liver disease related to hepatitis B or C virus infection based on the database reconstructed from medical claims from 2008 to 2010 in Japan. *Hepatol Res*, 2015. **[Epub ahead of print]**.
 51. Shima, D., et al., A retrospective, cross-sectional study of real-world values of cardiovascular risk factors using a healthcare database in Japan. *BMC Cardiovasc Disord*, 2014. **14**: p. 120.
 52. Hosaka, A., et al., Clinical and economic burden in patients with diagnosis of peripheral arterial disease in a claims database in Japan. *Clin Ther*, 2014. **36**(8): p. 1223-30, 1230 e1-4.
 53. Murata, A., et al., The recent time trend of outcomes of disseminated intravascular coagulation in Japan: an observational study based on a national administrative database. *J Thromb Thrombolysis*, 2014. **38**(3): p. 364-71.
 54. Tanaka, I., et al., Adherence and persistence with once-daily teriparatide in Japan: a retrospective, prescription database, cohort study. *J Osteoporos*, 2013. **2013**: p. 654218.
 55. Katada, H., et al., Prescription patterns and trends in anti-rheumatic drug use based on a large-scale claims database in Japan. *Clin Rheumatol*, 2015. **34**(5): p. 949-56.

56. Murata, A. and S. Matsuda, Association between ambulance distance to hospitals and mortality from acute diseases in Japan: national database analysis. *J Public Health Manag Pract*, 2013. **19**(5): p. E23-8.
57. Onishi, Y., et al., Psychotropic prescription patterns among patients diagnosed with depressive disorder based on claims database in Japan. *Clin Drug Investig*, 2013. **33**(8): p. 597-605.
58. Murata, A., et al., Time trend of medical economic outcomes of endoscopic submucosal dissection for gastric cancer in Japan: a national database analysis. *Gastric Cancer*, 2014. **17**(2): p. 294-301.
59. Furukawa, T.A., et al., Prescription patterns following first-line new generation antidepressants for depression in Japan: a naturalistic cohort study based on a large claims database. *J Affect Disord*, 2013. **150**(3): p. 916-22.
60. Hori, K., et al., Changes in compliance with Japanese antiemetic guideline for chemotherapy-induced nausea and vomiting: a nationwide survey using a distributed research network. *Support Care Cancer*, 2014. **22**(4): p. 969-77.
61. Kimura, T., et al., Quantitative Evaluation of Compliance with Recommendation for Sulfonylurea Dose Co-Administered with DPP-4 Inhibitors in Japan. *Pharmaceutics*, 2012. **4**(3): p. 479-93.
62. Nakano, T. and K. Shiosakai, Spread of viral infection to family members from influenza patients treated with a neuraminidase inhibitor. *J Infect Chemother*, 2014. **20**(7): p. 401-6.
63. Urushihara, H., et al., Increased risk of acute pancreatitis in patients with type 2 diabetes: an observational study using a Japanese hospital database. *PLoS One*, 2012. **7**(12): p. e53224.
64. Urushihara, H., et al., Oseltamivir prescription and regulatory actions vis-a-vis abnormal behavior risk in Japan: drug utilization study using a nationwide pharmacy database. *PLoS One*, 2011. **6**(12): p. e28483.
65. Miura, M., et al., Intussusception in Japanese infants: Analysis of health insurance claims database. *Open Journal of Pediatrics*, 2013. **3**: p. 311-316.
66. Ooba, N., T. Yamaguchi, and K. Kubota, The impact in Japan of regulatory action on prescribing of dopamine receptor agonists: analysis of a claims database between 2005 and 2008. *Drug Saf*, 2011. **34**(4): p. 329-38.
67. Food and Drug Administration, Best Practices for Conducting and Reporting Pharmacoeconomic Safety Studies Using Electronic Healthcare Data. 2013.
68. European Medicines Agency, Guideline on good pharmacovigilance practices (GVP) Module VIII - Post-authorisation safety studies (Rev 1) [http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129137.pdf]. 2013.
69. European Medicines Agency, FAQ - EU-PAS Register [http://www.encepp.eu/encepp_studies/documents/FAQ_EUPASRegister.pdf]. 2015.
70. European Network of Centers for Pharmacoepidemiology and Pharmacovigilance, EU PAS Register [http://www.encepp.eu/encepp_studies/indexRegister.shtml].

2015 年度データサイエンス部会 TF3

木村 友美	アステラス*
佐藤 恵子	グラクソ・スミスクライン
小林 典弘	塩野義製薬*
阪口 元伸	武田薬品工業*
中村 正樹	メディカルデータビジョン (日本医療データベース協会)
石川 秀一	エーザイ
大倉 征幸	ファイザー
米田 茂広	大塚製薬工場
長野 敦	バイエル薬品
惟高 裕一	塩野義製薬
河口 裕	田辺三菱製薬
鳥居 友紀子	サノフィ
國富 悠司	大鵬薬品工業
中島 章博	帝人ファーマ
湯浅 美幸代	大正製薬
瀧田 厚	第一三共

*TF リーダー

担当副部長

酒井 弘憲 田辺三菱製薬

Special Thanks:

青木 事成 (イラストレーション)

