

難病疫学研究におけるビッグデータ活用の可能性

研究分担者 氏名 川村 孝（京都大学健康科学センター）

研究要旨

近年、国や学会の主導で診療報酬や診療録情報をベースにした大規模データベースが整備されつつあり、それぞれに特徴がある。難病は稀発性のものが多いがゆえに、その疫学研究には広域性や悉皆性が重要となり、大規模データベースの活用の余地が大きい。一方で個々の難病の病名が登録されていないなどの問題がある。病態やテーマによっては、大規模データベースが難病疫学研究に利用できる可能性があると思われる。

A. 研究目的

いわゆる難病は、その稀発性ゆえに個々の疾患名が必ずしも国際疾病分類（ICD）に登録されておらず、既存の大規模データベース（いわゆる「ビッグデータ」〔後述〕）において特定の難病を同定することができなかった。そのため、疫学研究を行う際には疾患ごとあるいは疾患群ごとに実態調査やコホート研究、症例対照研究が行われるのが通例であった。しかし、揮発性の疾患こそ、いわゆるビッグデータが力量を発揮する余地が大きい。

一方、平成28年6月の「日本再興戦略2016」の閣議決定に基づいて“医療等情報の利活用”が推進されることとなり、平成30年5月に「医療分野の研究開発に資するための匿名加工医療情報に関する法律（次世代医療基盤法、通称：医療ビッグデータ法）」が施行されて、ビッグデータ活用の機運が高まってきた。

本試論では、難病研究においていわゆるビッグデータが利用できるかどうか、利用できるとすればその利用法と注意点は何か、について論考することとした。

B. 研究方法

公表された情報、自らの経験、あるいは研究班員の班会議での報告に基づき、概念的に分析・考察を行った。

（倫理面への配慮）

本分担研究は公開済みの一般的資料に基づいて論理的に考察するものであって、人を直接の対象としておらず、また個人情報を取り扱わないため、各種の法令や倫理指針の対象にはならず、特段の倫理的配慮を要しない。

C. 研究結果

1. 大規模データベースの種類と利用可能性

大規模データベースには医療データベースと行政データベースがある。

医療データベースの第一は、医療機関が作成した診療報酬明細書（いわゆるレセプト）の内容を直接あるいは健保組合を通して集約し統合したもので、（株）JMDCの「JMDC Claims Database」がその嚆矢である。近年では厚生労働省の「NDB」、国保中央会の「KDB」など構築されてきた。第二は、入院時医療の包括払い制度（DPC/PDPS）のデータをとりまとめたもので、厚生労働省の「DPCデータ」、メディカル・データ・ビジョン（株）の「MDV診療データ」がある。第三は、学会等が構築した症例レジストリで、外科手術所例を登録する（一社）National Clinical Databaseの「NCD」などがある。本研究で構築した症例レジストリもこの範疇に入る。第四は、法律に基づく症例の登録で、国立がん研究センターの「全国

がん登録」)がある。

行政データベースには、官庁が通常業務として行っている行政事業で入手した情報を集約して電子化したものや、法令に基づいて行政機関があらたに調査を実施したものがある。

通常業務ベースのデータベースには、出生や死亡等の届出をもとにした人口動態統計ならびにそれに基づく生命表、自治体の救急隊が院外心停止症例の発生から1ヶ月予後までを国際標準の様式で集約したウツタイン統計(医療データベースにも分類可)などがある。今後は母子健康手帳も電子化されて研究に利用できるようになる可能性がある。

2. 各データベースの特性

1) 診療報酬明細書データベース

「NDB」「KDB」「JMDC Claims Database」が該当する。

「NDB」は「高齢者の医療の確保に関する法律」に基づいて厚労省が全国の保険者から収集する診療報酬情報および特定健診のデータを収集し、匿名化処理を施したものである。申請に基づいてこのデータベースから必要部分を切り出し、研究者に提供される。退職に伴って保険者がかわったり、75歳になって後期高齢者医療制度に移ったりするので、追跡が不完全な症例が出てくる。複数の保険者番号を持っていたり改姓した者のデータが十分に連結できない。また、病名がいわゆる「保険病名」と言われるように便宜的なものが付与されている可能性がある。行われた行為はほぼ確かだが、診断や治療の意図や過程はわからない。なお、従来問題となっていた医療と健診の個人同定情報の記述方式の差異による連結不全はかなり克服されている。

「KDB」はNDBの国保版である。市町村が扱う介護保険情報が加わっているのが特徴である。

「JMDC Claims Database」は民間会社が医療機関や健保組合の経営分析のために多数の健保組合から提供を受けたものである。大企業の健保組合が多いため、対象者の年齢が18~65歳に偏っている。NDCと同様に転退職で所属する保険者が変わると連結はできない。

2) DPC/PDPSデータベース

厚生労働省の「DPCデータ」とメディカルデータ・ビジョンの「MDV診療データ」が該当する。前項のレセプトデータと類似した特性を持つが、病名が「入院の契機となった病名」や「医療資源を最も消費した病名」などとなっていること、体重や意識レベル、がんのTNM分類、脳卒中のModified Rankin Scaleなど一部の臨床情報が含まれていて、患者の基本的な臨床像がわかるのが特徴である。DPC制度を利用している病院は1730施設(平成30年4月現在)で全体の2割にとどまり、大規模な急性期病院が中心である。医療機関ベースのため、転医すると追跡不能となる。

3) 人口動態統計(死亡統計)

市区町村への死亡届に基づいて厚生労働省で作成される。原死因について知ることができる。したがって、自前の診療や健診データと突合して疫学研究の転帰データとして利用することが出来、すでに多くの利用実績がある。ただし、データの利用には特別な申請が必要であり、許可が下りるまでに相当の期間を要する。

4) がん登録

発がんにつながる難病であれば、がんの併発があるかどうかを確認することができる。ただし、法制化されたのは最近であり、現時点では登録漏れが稀ではなく、「登録がないからがん発生がない」とは言えない。また、入力1患者1回のため、長期予後はわからない。

5) 学会・省庁等による症例レジストリ

NCDやウツタイン統計は規模も大きく、患者の病態を示す多数のデータを内包しているが、1患者1回入力のため、長期予後はわからない。

3. データベース研究の利点と問題点

既存のデータベースを利用した研究の利点と問題点を列記する。

1) 規模が大きい

全国レベルで情報が得られることから、データベースの規模が大きくなる。疫学研究では数が武器となるので、稀発性の疾患を課題とする研究ではこの

スケールメリットはきわめて大きい。

2) 労力が削減できる

日常業務で得られるデータを国や学会が音頭を取って集約するので、個々の研究者はデータ収集の設計や運営を自ら行う必要がなく、労力が大幅に削減できる。

3) 疾患把握漏れのおそれ

反面、揮発性の疾患は独立した病名が付与されているとは限らず、包括的な病名（類縁の高頻度疾患もしくは「その他」）にまとめられたりすることもあり、目的の疾患が把握できない可能性がある。

4) データの成り立ちが十分にわからない

研究者が自分で作成したものではないため、個々のデータがどのように処理されて記載されたかが十分にわからない。たとえば、診断の基準、数字の丸め方、選択肢複数選択時の処理、矛盾回答の処理などである。

5) 解析したいデータが含まれていない

日常業務ベースであることが多いため、研究としてほしい項目が含まれていないことが稀ではない。また、日常業務で扱っている事項でもデータとして記載されていないこともある。また、全国がん登録、外科手術例のレジストリ(NCD)、院外心停止登録(ウツイン)では、登録作業が1回「だけなので、登録時以降の長期予後は把握できない。

6) 他のデータベースとの連結が困難

登録あるいはデータベースが構築される段階で匿名化されるので、個人の同定が原則としてできない。個人情報の保護の観点から当然の措置であるが、匿名化状態のまま他のデータベースとのリンケージを行うにはハッシュ関数など特別な仕組みを組み込まなくてはならない。

4. データベース研究の候補

平成30年度の本研究班の班会議で発表された研究課題のうち、既存のデータベースを利用することでメリットを発揮しそうな研究には、以下のようなものがある。

1) 血栓塞栓症発症までの女性ホルモン剤の使用期間 (浜松医大健康社会医学教室)

女性ホルモン剤使用中の血栓塞栓症の関連を明らかにするため、全国疫学調査が行われた。その結果、女性ホルモン剤の使用期間が1ヶ月以内が31%、1~3ヶ月が17%、4~6ヶ月12%、7~12ヶ月が10%、1~2年が5%、2年超が15%であった。その幾何平均は低用量ピルで116日、中容量ピルで29日、高容量ピルで13日、経皮剤で76日であり、有意な差が認められた。

この研究はNDBなど診療報酬明細書データベースもしくはDPC/PDPSデータベースの処方歴と血栓塞栓症を示す病名の付いた疾患の発症日を用いて実施可能である。ただし、前述のように、前者では高齢者の把握漏れ、後者では転医時の追跡漏れが生じうる。

2) 特発性大腿骨頭壊死症の地域分布 (大阪市立大学公衆衛生学教室)

2015年実施の全国疫学調査で特発性大腿骨頭壊死症と確定診断された症例の都道府県別発症率を算定した。その結果、都道府県間に6~15倍の発症率の差異が認められた。福岡県、宮崎県、佐賀県、山口県、島根県、石川県などが高率で、九州・中国地域に集積していた。発症と生活習慣との関連についても検討されたが、関連は希薄であった。

この研究は、都道府県格差だけであれば、診療報酬明細書データベースで分析しうる。ただし、疾患の捕捉が十分にできないおそれがある。生活習慣との対比については、他資料との個人レベルのリンケージができないため、国民健康栄養調査結果を用いた生態学的研究(地域相関研究)にとどまる。

3) 定点モニタリングによる特発性大腿骨頭壊死症の記述疫学 (大阪市立大学公衆衛生学教室)

研究班で構築した定点モニタリングシステムにより、2011~2018年に報告された新患の診断時年齢を分析した。その結果男性では30歳代から40歳代に集積が認められ、顕著な経年変化は認められなかった。女性では30歳代から60歳代に広く分布し、近年は40歳代が相対的に減少していた。

病名の正確さの問題はあるが、有病患者の経年変化も診療報酬明細書データベースで分析できよう。

ステロイドとの関連もこのデータベースが得意とするところである。

D. 考察

既存データベースを用いた研究には種々の限界があるが、日常業務のデータがほぼ丸ごと記録されており、新たな情報収集の手間が省略できることが最大の強みである。また悉皆性が高く、スケールメリットで微小な関連を検出しうる点も大きな特徴である。

反面、診療報酬明細書データベースにおける病名の正確性、症例登録における追跡情報（転帰）の不備、また全体に症状や発症前生活習慣、検査結果の欠如などがあり、本質的な問題が残る。

しかし、仮説形成のためのスクリーニングを行ったり、介入試験の症例設計をしたり、介入試験結果の実社会における妥当性の確認など、使い道も広い。今後、難病研究における活用法の検討が必要と思われる。

E. 結論

難病の疫学研究においても大規模データベースを活用する可能性がある。

F. 研究発表

1. 論文発表

- (1) Sawada H, Oeda T, Kawamura T, et al. Early use of donepezil against psychosis and cognitive decline in Parkinson's disease: a randomised controlled trial for 2 years. J Neurol Neurosurg Psychiatry 2018; 89: 1332-1340.
- (2) Takeda M, Yamamoto K, Kawamura T, et al. Clinical characteristics and postoperative outcomes of primary aldosteronism in the elderly. J Clin Endocrinol Metab 2018; 103: 3620-3629.
- (3) Umakoshi H, Tsuiki M, Kawamura T, et

al. Correlation between lateralization index of adrenal venous sampling and standardized outcome in primary aldosteronism. J Endocr Soc 2018; 2: 893-902.

- (4) Ohno Y, Sone M, Kawamura T, et al. Prevalence of cardiovascular disease and its risk factors in primary aldosteronism: a multicenter study in Japan. Hypertension 2018; 71: 530-537.
- (5) Akehi Y, Toshihiko Yanase T, Kawamura K, et al. High prevalence of diabetes in patients with primary aldosteronism (PA) associated with subclinical hypercortisolism and prediabetes more prevalent in bilateral than unilateral PA: a large multicenter cohort study in Japan. Diabetes Care (in press).
- (6) Katabami T, Kawamura T, et al. Clinical and biochemical outcomes after adrenalectomy and medical treatment in patients with unilateral primary aldosteronism. J Hypertens (in press).
- (7) 川村 孝. 臨床研究のデザインとデータ分析のための疫学の手法. 第37回日本社会精神医学会雑誌 2018; 27: 207-211.
- (8) 川村 孝. 研究方法論: 疫学を用いたエビデンスの作り方. 理学療法京都 (in press).

2. 学会発表

- (1) 川村 孝. 診療・保健事業での疑問から研究計画へ. 第54回日本循環器病予防学会: 2018年6月: 札幌.
- (2) 川村 孝. 研究方法論: 疫学を用いたエビデンスの作り方. 日京都府理学療法士会: 2018年7月: 京都.
- (3) 川村 孝. エビデンス確立のための臨床研究. 第16回日本機能性食品医学会: 2018年12月: 新潟.

G. 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
特記事項なし

H. 共同研究を行った他の難病研究班

重症型原発性アルドステロン症の診療の質向上に資するエビデンス構築（17ek0109122h0003）. 研究代表者：成瀬光栄 国立病院機構京都医療センター臨床研究センター内分泌代謝高血圧研究部長