

厚生労働行政推進調査事業費補助金
(地域医療基盤開発推進研究事業)
令和元年度～令和2年度(令和3年度への繰越)総合分担研究報告書

医療データベースを活用した診療ガイドラインの推奨度決定手法に関する研究
(19IA2024)

各種医療データベースからデータを収集するための経費や作業、得られる臨床データの性質や位置付けについての調査、比較解析

研究分担者 藤永 潤

(公財)大原記念倉敷中央医療機構 倉敷中央病院 集中治療科
岡山大学大学院医歯薬学総合研究科 疫学・衛生学分野 客員研究員

研究要旨：2019 年度は国内に複数存在している大規模医療データベースをレビューし、それらに基づく研究結果をランダムに選び検討した。Randomized controlled study との関係やエビデンスとしての位置づけは、用いられたデータベースや情報の切り出し方に由来するバイアスが関わってくると考えられた。

2020 年度は国内に複数存在している大規模医療データベースを用いて行われた研究結果を検討した。それぞれのデータベースにより実績のある研究領域、研究の形式は異なっており、前年度の検討で指摘したそれぞれのデータベースの特徴に關係していると考えられた。

2021 年度においては総括として、行政請求データベース (administrative claims database) を用いた各研究における種類、デザイン、研究領域を評価しデータベースの適合性を検討した。4 つの行政請求データベース (NDB: the Japanese National Database of Health Insurance Claims and Specific Health Checkups, DPC: the Diagnostic Procedure Combination database, JMDC: Japan Medical Data Center, MDV: Medical Data Vision) からの研究をレビューした。PubMed は 2015 年 1 月から 2020 年 10 月まで検索した。653 件の研究が同定された。研究数は 2015 年の 60 件から 2020 年の 176 件に増加していた。記述研究は NDB からの研究の 63.2% (43 件) を占め、治療効果研究は DPC データベースからの研究の 81.7% (215 件) を占め、JMDC と MDV に最も多かった。コホート研究は NDB では 16%に過ぎないが、DPC では 41.1%を占めていた。研究分野は「Medicine, general and internal medicine」が 8.2% (53 件) と最も多かった。実施された研究の種類やデザインはデータベースごとに異なり、使用するデータベースの特徴に影響されることがわかった。今後、様々な RWD データベースを統合し、網羅性を高めることが必要である。

研究1. 国内大規模医療データベースの特徴と各データベースを用いた研究への影響

A. 研究目的

国内における大規模医療データベースそれぞれの強いところ、弱いところに関して検討を行う。

B. 研究方法

1) 臨床研究の実績があるものを中心に、国内における医療データベースを抽出し、複数の観点からの分類を行った。

2) レセプト情報・特定検診等情報データベース (NDB)、Japan Medical Data Center (JMDC)、Medical Data Vision (MDV)、Diagnosis Procedure Combination database (DPC)、National Clinical Database (NCD) を選択し、それぞれランダムに論文を選び、使用されたデータベースに由来する特徴にに関して検討した。

(倫理面への配慮) すでに発表された研究のみを用いたため、倫理承認は不要であった。

C. 研究結果

分類方法としては情報取得方法から見た分類（リアルワールド系と非リアルワールド系）と、一次目的による観点（clinical databaseとadministrative database）による分類が主であった。データベースを用いた研究に共通する強みとして、稀な疾患・アウトカムを検討可能、無作為化を行えない（倫理的、費用など）内容が検討可能、短時間に行えることや、一般化可能性が検討可能であることがあげられた。一方、観察研究一般に共通する弱みとして、逆因果や交絡因子の検討を要していた。また医療データベース（特にadministrative claims database）を用いた研究に特有と考えられる弱みとして、情報の内容によっては精度が低く、情報を抽出するコードセットに工夫が必要である、曝露/アウトカム情報の感度・特異度や選択基準、除外基準における陽性/陰性的率が生み出すバイアスや臨床情報、検査データ、画像などの情報が得られないデータベースが多い、などの点が指摘できた。

E. 結論

検証したい仮説に適した医療データベースを用いているか、用いられた医療データベースや情報の切り出し方に由来するバイアスはどのようなものであるかがエビデンスの確実性に対する評価に関わってくると考えられた。

研究2. 各種医療データベースから得られる臨床データの性質や位置付けについての調査、比較解析

A. 研究目的 それぞれの医療データの強いところ、弱いところに関して指摘された点を、これまでに各データベース(DB)を用い

て発表された研究において影響が見られるか検討を行う。

B. 研究方法

1) 対象: 2015. 01-2019. 01の期間に発表され、かつレセプト情報・特定検診等情報データベース (NDB)、Japan Medical Data Center (JMDC)、Medical Data Vision (MDV)、Diagnosis Procedure Combination database (DPC) のいずれかを用いて行われ、PubMedに収載された研究。除外基準: 本文が英文以外である、Review article, editorial など個人レベルのデータを使用していない研究。データ収集の方法: JMDC, MDVに関しては各DBのホームページ (HP) (<https://www.jmdc.co.jp/>, <https://www.mdv.co.jp/>)、DPCに関しては「 Diagnosis Procedure Combination 」でPubMedを用い検索を行った。NDBに関しては「 NDB & japan, OR National Database of Health Insurance Claims and Specific Health Checkups of Japan 」でPubMedを用い検索した。また厚生労働省 HP (<https://www.mhlw.go.jp/index.html>) を参照した。各研究の領域の分類は研究が掲載された雑誌に基づいて分類した。分類は、1. Science Citation Index Expanded (SCIE), 2. Emerging Sources Citation Index (ESCI), 3. Social Sciences Citation Index (SSCI) (<https://mjl.clarivate.com/home>), 4. Directory of Open Access Journals (DOAJ) (<https://doaj.org/>) の順に適用した。研究の形式は記述疫学、探索的研究、治療効果研究、その他の4つとした。(倫理面への配慮) すでに発表された研究のみを用いたため、倫理承認は不要であった。

C. 研究結果 対象となった研究の内、NDB: 48件、JMDC: 127件、MDV: 114件、DPC: 217件を検討した。治療効果研究の内、多くがDPCデータを用いており、またDPCを用いた研究の内70. 5%が治療効果研究であった。NDBは記述研究が多い傾向にあった。研究の領域としてはPharmacology & Pharmacyに次いでMedicine, General & Internalの論文が多く認められた。これらの領域での研究は NDB, JMDC, MDVを用いたものが多く見られた。DPCではSurgeryの領域での研究が多くみられた。

治療効果研究は DPCを用いて多くが行われていた。DBごとに発表の実績のある研究領域、研究の形式は異なっていた。このことは前年に指摘した各データベースの特徴に由来していることが考えられた。

研究3. 日本における行政請求データベースからのデータを用いた研究調査のレビュー

A. 研究目的 日本では、リアルワールドデ

ータ (RWD) を用いた大規模医療データベースが開発され、リアルワールドエビデンスが構築されている。米国食品医薬品局では、リアルワールドデータを、従来のRCTの文脈ではなく、電子カルテ、請求・課金活動、製品、疾病登録などから日常的に収集される患者の健康状態および/または医療データとして定義している。日本薬剤疫学会「薬剤疫学とデータベースに関するタスクフォース」では、臨床疫学や薬剤疫学に応用可能な日本国内の医療データベースの特徴を調査しており、ここで挙げられている大規模医療データベースには、National Clinical Databaseなどのレジストリ、DPC (Diagnostic Procedure Combination) データベース、NDB (Japan National Database of Health Insurance Claims and Specific Checkups)、JMDC Claims database、Medical Data Vision (MDV) databaseといった行政請求データベースがある。レジストリは、特定の目的-例えば、研究の実施、希少疾患のレジストリ、医療提供者の指標のモニタリング、医療の質の向上-を前提に作成されるものである。一方、行政請求データベースは、診療報酬を得るために日常的に収集される医療情報で構成されている。研究を目的としたデータではないが、行政請求データベースはアウトカム研究のための貴重な情報源であり臨床研究への応用が進んでいる。しかし、研究者は行政請求データを利用する際にいくつかの点に注意しなければならない。第一に、不完全性やコーディングの正確さのばらつきなど、固有の限界がある。同じコードを使用しても、異なる環境の異なる集団では精度が異なり、結果の妥当性や一般化可能性に影響を与える可能性があることが指摘されている。研究者はまた、研究の目的に適したデータベースを選択する必要がある。症例対照研究やコホート研究では、参加者の追跡が必要なため、保険者ベースのデータベースが適しており、入院中に観察されるアウトカムには病院ベースのデータも適用可能である。患者の背景を調整し、適応バイアスを制御するために用いられる傾向スコア分析では、参加者の基礎疾患や疾患の重症度に関する情報が必要となる。さらに、研究に選ばれた集団の特性も、研究結果の一般化可能性を損なう可能性がある。したがって各データベースの特徴を把握し、研究目的に応じた最適なデータベースを選択することが必要である。各データベースの特徴はすでに報告されているが、各データベースで行われた調査研究がすでに指摘されている特徴に対応しているかは不明である。

日本の行政請求データベースで実施された研究の種類を検討した研究は、NDBを対象と

したもののみであり、各データベースで実施された研究の種類とデザインを比較した研究は存在しない。そこで、大規模医療データベース、特に日本の行政請求データベースを用いてこれまでに実施された研究の種類とデザインを評価し、各データベースにおける研究の特徴を明らかにすることを目的とした。

B. 研究方法

国内行政請求データベースからの研究のレビューを実施した。NDB、DPCデータベース、JMDC、MDVのいずれかを用いた英文研究を対象とした。NDBは2013年に第三者へのデータ提供を開始し、上記データベースの中では最も新しい。そのため、2015年1月から2020年10月までの研究を対象とした。検索にはPubMedを用いた。

除外基準は以下の通りである。1. 英語以外の研究、2. レビュー論文や論説など、個人レベルのデータを使用していない研究。

データ収集

JMDC、MDV は各データベースのホームページ (<https://www.jmdc.co.jp/>, <https://www.mdv.co.jp/>) の研究一覧から、DPC は PubMed の "Diagnosis Procedure Combination" から検索した。NDB については、PubMed で "NDB & japan, OR National Database of Health Insurance Claims and Specific Health Checkups of Japan" で検索を行った。また、厚生労働省のホームページ (<https://www.mhlw.go.jp/index.html>) を参照した。さらに、NDB に関する過去のレビューと各研究の参考文献を検索し、研究を特定した。

分類

行政請求データベースを用いた研究は、主要アウトカムについて、記述的研究、治療効果研究、その他に分類した。治療効果研究は、手続き的アプローチの違いにより、探索的治療効果研究と仮説的治療効果評価 (HETE) 研究の2つに分類し、4つのカテゴリーに分類された。HETE研究は、先駆的に明示された仮説を持つ研究と定義された。主要転帰の明示がない場合は、最初に記載された結果を主要転帰とした。2つ以上のグループを比較したが、起こりうる交絡因子について調整が行われていない場合、その研究は記述的研究と分類した。また、研究をデザインによって以下のように分類した。1. 記述研究、2. コホート研究、3. 横断研究、4. 症例対照研究及びマッチドコホート研究、5. 傾向スコア解析、逆確率治療重み付け (IPTW) 、marginal structural models、g-formula、6. Case-cross over及びSelf-Controlled Case Series、7. 回帰不連続、差分の差分法、操作変数法、8. 予測モデル、臨床スコア、9. 機械学習、10. 費用対効果分析、費用最小化分

析、11. その他である。一次アウトカムを調べるために主に使用された研究デザインを分類した。各研究領域の分類は、研究が掲載されたジャーナルを基準とした。1. Science Citation Index Expanded (SCIE)、2. Emerging Sources Citation Index (ESCI)、3. Social Sciences Citation Index (SSCI)、そして4. Directory of Open Access Journals (DOAJ) のデータベースを用いて、研究が掲載された雑誌を分類した。SCIEはその分野で最も影響力のある科学雑誌、SSCIは最も影響力のある社会科学雑誌、ESCIはSCIEに掲載されていないが一定の資格を満たす科学雑誌を含む、新たに導入されたデータベースである。DOAJは、オープンアクセスの査読付き学術雑誌を索引付けしている。ジャーナルが上記のデータベースの複数に掲載されている場合、影響力の大きい順 (SCIE、ESCI、SSCI) に優先した。複数の研究分野にまたがるジャーナルが分類されている場合は、最初に分類された研究分野を使用した。

統計分析

データベース別に研究の種類とデザインについて記述的な分析を行った。データベース間の特徴を比較するために、フィッシャーの正確検定を用いた。分析はすべてStata version 16.1 software (Stata, College Station, TX) を用いて行った。p値<0.05 (両側) を統計的に有意とみなした。すでに発表された研究のみを用いたため、倫理承認は不要であった。

C. 研究結果

合計812件の研究を検索した (NDB : 53件、DPCデータベース PubMedから341件、JMDCとMDVからそれぞれ227件と183件、NDBは厚生労働省のHPから8件)。タイトルと抄録をスクリーニングした結果、632件 (NDB : 40件、JMDC : 185件、MDV : 147件、DPC : 260件) が本レビューに含まれた。NDBの過去のレビューから28件、マニュアル検索から7件が追加された。このうち、JMDCとMDVの両方を用いた研究は14件であり、重複としてカウントされた。合計653件の研究が対象となった。各年度の発表された研究数は年々増加傾向にあり、特に2017年から2020年にかけては8件から176件と増加が著しかった。

研究の種類は、半数以上がHETE研究に分類され、特にDPCデータベースを用いた研究では70% (184件) を占めた。データベース別の研究分野は「Medicine, General & Internal Medicine」が53件 (8.2%)、次いで「Cardiac & Cardiovascular Systems」が40件 (6.2%) となっている。しかし、NDBではInfectious Diseases (n=10, 14.7%)、Pharmacology & Pharmacy (n=8, 11.8%) が多く、DPCではSurgery (n=25, 9.6%) が最も多く

なっている。JMDCではPharmacology & Pharmacy (n=21, 12.1%)、MDVではEndocrinology & Metabolism (n=15, 11.5%) が最も多く実施された。

D. 考察

2017年以降に研究が大きく伸びた理由は、データベースの整備だけでなく、法改正によるところも大きい。2017年の個人情報保護法の改正、2018年の次世代医療基盤法の制定により、RWDが適切な場面で活用されやすくなった。

HETE研究がDPCデータベースを中心に行われたのは、治療や処置だけでなく、基礎疾患や重症度に関する情報を比較的容易に収集できるからだと推測される (Kumamaru 2020)。また、近年多くの観察研究で用いられている傾向スコアマッチングや治療逆確率重み付けなどの準実験デザインにおいても、これらの情報は必要とされる。(Bärnighausen 2017) NDBには基礎疾患や重症度に関する情報はない。そのため、DPCデータベースでは準実験デザインが多く用いられているが、これはこれらの情報を比較的容易に入手できることに起因すると考えられる。DPCデータベースの欠点は、その対象が入院患者、特に急性期病院の患者に限定されていることである。この欠点は、研究の範囲を限定し、長期的な転帰を追跡することを困難にするだけでなく、一般化可能性の欠如や選択バイアスにつながる可能性がある。

NDBは他のデータベースと比較すると記述的研究の割合が最も高く、NDBを用いた研究の60%以上が記述的研究であり、これは先行研究と同様である (Hirose2020)。重症度や他の要因の調整が難しいことが示唆された。NDBの長所は、日本のほぼ全人口をカバーしていること、ほぼ全ての保険医療をカバーしていること (Okumura2017)、保険者からデータを収集しているため、患者が異なる医療機関を受診してもトレーサビリティが確保されていることである。一方、NDBデータの利用にはいくつかのデメリットがある。申請制度が厳格で、研究者は厚生労働省が定めた個人データを用いた研究の手続きに従う必要があり、申請の妥当性を専門家委員会で承認されなければならない。データ取得に時間を要し、申請からデータ提供まで180日以上と長期に渡ることもある

(https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryouhoken/reseputo/index.html)。これに対し、JMDCは5~30営業日 (<https://www.jmdc.co.jp/en/jmdc-claims-database/>)、MDVは2週間で取得できる。(<https://en.mdv.co.jp/about-mdv-database/>) また、死亡率に関する情報がなく、データの構造や量から、患者のリスクや重症度を調整するための情報が不足

しており、データの取り扱いが困難である（Hiramatsu2021）。そのようなことから、記述研究以外の研究を行うことは難しく、他のデータベースと比較して研究数が限られる可能性がある。

また、JMDCは保険者ベース、MDVは病院ベースであり、実施される研究の種類やデザインも異なることが予想されたが、両データベースとも、研究の種類は比較的HETE研究が多く、デザインはコホート研究が多かった。両データベースともレセプトデータだけでなくDPCデータも使用しており、DPCデータベースと同様に基礎疾患や重症度の調整が可能するために両データベースは同様の結果であった可能性がある。

大規模医療データベースを用いて政策・ガイドライン変更の効果を評価するには、回帰不連続デザイン（RD）または差分の差分法（DID）が適切であると思われる（Nishioka 2021, Zogg2017, Magnani2019）。しかし、本研究でレビューした研究のうち、RD や DID

が適用されたのはごく一部であった。医療における準実験デザインの少なさは以前から指摘されており、改善のため次のような方法が提言されている。準実験デザインを用いる研究能力の開発支援、実施と報告に関するガイドラインの開発、医療プログラムの管理データとレジストリが、公的資金を受けるすべての医療機関で利用可能にすること、検証済みのバイアスリスク評価法を拡張して系統的レビューに準実験的デザインに含めること、交絡変数に関するデータ収集を標準化すること、など（Tugwell 2017）。とはいっても、本研究においてデータベース研究のデザインはほとんどがコホート研究であり、高度な統計手法がなくても研究が可能であること、また、データへのアクセスやデータの大きさなど、他の障壁が課題となりうることを示唆している（Hiramatsu2021）。さらに縦断データの利用がコホート研究の実施を容易にすることを示唆するものである。

また、大規模医療データベースを利用して、広い範囲の研究分野で研究が実施されていることがわかった。薬理学・薬学分野の研究が2番目に多かった（45件、7%）が、大規模データセットの利用は、薬剤疫学においてますます一般的になってきている。RWDを用いた医薬品の有効性・安全性の研究である薬剤疫学におけるデータベースの利用は、今後さらに重要性と割合が高まると予想される（Takahashi2012）。

E. 結論

各データベースで実施された研究の種類とデザインは、データベースの特徴に対応していることが示唆された。DPCデータからなるデータベース（DPCデータベース、JMDC、

MDV等）では、個人に関する情報を用いたHETE研究が多く実施されていた。一方、NDBを用いた研究は記述研究が中心であり、データの入手に関するハードルや網羅性の影響があると考えられる。傾向スコア分析は多く実施されたが、その他の準実験デザインは十分に実施されていない。縦断的データの入手が可能であったため多くのコホート研究が実施された。今後、RWD のデータベースを幅広く収集・統合し、包括性を高め、価値あるエビデンスを導出することが必要である。

F. 研究発表

a. 論文発表：

Body Mass Index and Ventilator Dependence in Critically Ill Subjects in Japan: A Cohort Study Using a Nationwide Database

Jun Fujinaga, Etsushi Suzuki, Hiromasa Irie and Mutsuo Onodera
Respiratory Care March 2021, respcare. 08660; DOI: <https://doi.org/10.4187/respcare.08660>

b. 学会発表：

- 藤永潤 集中治療領域におけるreal-world evidenceとガイドラインへの活用の状況 第49回日本集中治療医学会学術集会 シンポジウム8(SY) ビッグデータとデータサイエンス2022.3.18 仙台
- 藤永潤, 栗山明, 西川貴史, 梅田武英, 大竹孝尚, 小野寺睦雄 人工呼吸患者における去痰剤と人工呼吸期間との関係-a propensity score matched study- 第49回日本集中治療医学会学術集会 2022.3.18 仙台
- 国内大規模医療データベースと各データベースを用いて行われる研究の特徴 第58回日本腹部救急医学会総会 特別企画「厚労省研究班からの報告 ビッグデータをどう使うか：課題と未来」2022.3.24

G. 知的財産権の出願・登録状況

（予定を含む。）

特になし

(2) 研究実施日程

2019年

研究実施内容	実 施 日 程											
	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月
各種医療データベースからデータを収集するための経費や作業、得られる臨床データの性質や位置付けについての調査、比較解析 (藤永 潤)			調査		第1回会議			調査	第2回会議		検討	

2020年

研究実施内容	実 施 日 程											
	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月
各種医療データベースからデータを収集するための経費や作業、得られる臨床データの性質や位置付けについての調査、比較解析 (藤永 潤)		調査		第1回会議	調査	第2回会議	調査		第3回会議	検討		

2021年

研究実施内容	実 施 日 程											
	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月
各種医療データベースからデータを収集するための経費や作業、得られる臨床データの性質や位置付けについての調査、比較解析 (藤永 潤)		調査			第1回会議		調査		第2回会議	検討		発表

国内大規模医療データベース を用いて行われる研究の特徴

藤永 潤

大原記念倉敷中央医療機構
倉敷中央病院 集中治療科

日本腹部救急医学会 Medical ethics

筆頭演者氏名: 藤永 潤

私の発表内容は、倫理指針カテゴリー A に該当するので、倫理審査は不要だが、倫理指針に則った発表である。

- 症例報告
- 連結不可能匿名化データのみを扱った研究
- 培養細胞のみを扱った研究
- その他の倫理審査が不要な発表

該当する項目の□に✓を入れて提示してください

内容

- 1.国内における大規模医療データベースにはどのようなものがあるか
 - NDB, DPC, MDV, JMDCとNCDなどレジストリとの違い
薬剤疫学会DB
- 2.データベースを用いた研究の型にはどのようなものがあるか
 - 記述、探索など
- 3.各データベースに適していると考えられる研究の型
 - 過去文献での記述の紹介 分析対象の考察
- 4.実際にどのような研究が行われているか
 - タイプ、デザイン、分野の実際の検討

医療における大規模医療データベース

医療系データベース

医療現場で行われている情報を
改変なく取得

- Administrative データベース（レセプト、
DPC由来）、電子カルテ由来の診療情報
データベース etc.

自発的に入力あるいは客観的に
取得も精度管理に限界がある

- レジストリ：NCD、JIPADなど
- 特定の疾患、状態、または曝露によって定義される集団の特定の結果を評価するため、均一なデータを収集するシステム

国内のデータベース

- 日本薬剤疫学会 データベース一覧 (google.com)

- 日本薬剤疫学会 薬剤疫学とデータベーススクウォース. 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.
- Available: <http://www.jspe.jp/committee/020/0210/>



- 研究実績の多いデータベース※

- NDB; レセプト情報・特定健診等情報データベース (National Database of Health Insurance Claims and Specific Health Checkups of Japan)
医療費適正化計画の作成、実施及び評価を目的とした電子レセプト及び特定健診・特定保健指導データのデータベース
保険者ベース Administrative data
- DPC; Diagnostic procedure combination database.
全国DPC病院が収集する入院患者情報データ、診療報酬明細データ
病院ベース Administrative data
- JMDC; 株式会社日本医療データセンター(JMDC保険者データベース)
全国の保険者から寄せられたレセプト・健康診断結果・加入者台帳からなる
保険者ベース Administrative data
- MDV; メディカル・データ・ビジョン株式会社(EBM Provider)
全国449の急性期医療機関を対象とした、入院、外来の診療データベース
病院ベース Administrative data
- NCD; National Clinical Database. 外科系学会を中心に創設された多領域のレジストリ基盤
病院ベース レジストリ

※日本薬剤疫学会 データベース一覧で取り上げられているもの。その他、ウツタイン統計、J-ROADなども実績が多い

Real-world dataを用いた研究の型

- 記述疫学
- 治療効果に関する研究
 - 探索的研究(exploratory treatment effectiveness studies)*仮説生成型
 - 治療効果(評価)研究(hypothesis evaluating treatment effectiveness studies, HETE)*仮説検証型

*Berger ML et al: Value Health. 2017;20:1003-1008.

Original Report

Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making

Marc L. Berger, MD^{1,2*}, Harold Sox, MD³, Richard J. Willke, PhD⁴, Diana L. Brixner, PhD⁵, Hans-Georg Eichler, MD⁶, Wim Goettlich, PhD⁷, David Madison, PhD⁸, Amr Makary, MSc⁹, Sebastian Schneeweiss, MD, ScD¹⁰, Rosanna Tarricone, MSc, PhD¹¹, Shirley V. Wang, PhD, ScM¹², John Watkins, MPH, PharmD¹³, C. Daniel Mullins, PhD¹⁴

¹New York City, NY, USA; ²Patient-Centered Outcomes Research Institute, Washington, DC, USA; ³Healthcare and Medical Society for Pharmacology and Therapeutics, Lansdale, PA, USA; ⁴University of Utah, Salt Lake City, UT, USA; ⁵European Medicines Agency, London, UK; ⁶Zorginstituut Nederland and University of Utrecht, Utrecht, The Netherlands; ⁷Columbia University, New York City, NY, USA; ⁸B Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; ⁹Brown University, Miln, Italy; ¹⁰Premera Blue Cross, Mountlake Terrace, WA, USA; ¹¹University of Maryland, Baltimore, MD, USA



	NDB	DPCデータベース	JMDC	MDV
カバー範囲	ほぼ全人口	急性期病院の入院患者	健康保険加入者の28% 主に65歳未満	急性期病院の26%
主体	公的	民間	民間	民間
データ	DPC データ	×	○	×
	医科レセプト	○	○	○
	調剤レセプト	○	×	○
	電子カルテ	×	×	×
	健康診断	特定検診	×	○
記述研究	◎	○	△	△
探索的研究	-	○~△	○	○
治療効果研究	○	○	○	○
構成家族の分析	-	-	○	-
高齢者の分析	○	○~△	△	○
入院患者の分析	○	◎	○	○
外来患者を対象	○	-	◎	○
地域を考慮	○	○	-	△
重症度調整	-	○	○	○

日本薬剤疫学会 データベース一覧 (<https://sites.google.com/view/jspe-database-ja2020/%E3%83%9B%E3%83%BC%E3%83%A0>)
廣居 伸藏ら, 薬剤疫学 Jpn J Pharmacoepidemiol 2018; 23(1): 41-47.

実際にどのような研究が行われているか

- データベース（DB）の特徴は研究の種類、内容に影響を与える可能性がある。分析するデータの特徴を考慮して研究デザインを選択する必要がある¹
- 例)
 - コホート研究、ケースコントロール研究；縦断データが必要
 - プロペンシティスコア解析；重症度や患者背景に関連するデータが必要
 - アウトカムの種類；急性期のアウトカムなら病院ベースのDB
- これまでに国内のDB毎にどのような研究が行われているか比較、検討されていない

1. Imai S. [An Introduction to Real-world Data and Tips for Analysing It]. *Yakugaku Zasshi*. 2021;141(2):169-174. doi:10.1248/yakushi.20-00196-2

目的

- 国内DBを用いてどのような研究が行われているか、DBごとに実施された研究の特徴を明らかにする



方法

・対象

- ・期間；2015.01-2020.12.31
- ・レセプト情報・特定検診等情報データベース（NDB）、Japan Medical Data Center（JMDC）、Medical Data Vision（MDV）、Diagnosis Procedure Combination database（DPC）のいずれかを用いて行われた
- ・PubMedに収載された研究

・除外基準

- ・英文以外
- ・Review article, editorialなど個人レベルのデータを使用していない研究
- ・データベース作成時にDPCデータを用いたレジストリは除外した
 - ・※研究に用いることを想定し作成されており、今回扱ったデータベースと異なる性質を持つことが想定される



方法 2

- ・データ収集：JMDC, MDVに関しては各DBのホームページ(HP), ^{1,2} DPCに関しては「Diagnosis Procedure Combination」、NDBに関しては「NDB & japan, OR National Database of Health Insurance Claims and Specific Health Checkups of Japan」でPubMedを用い検索した。また厚生労働省のHP³と先行研究⁴を参照した
- ・研究の領域の分類：掲載雑誌に基づいて分類した
 - ・1.Science Citation Index Expanded(SCIE), 2.Emerging Sources Citation Index (ESCI), 3.Social Sciences Citation Index (SSCI) (<https://mji.clarivate.com/home>), 4. Directory of Open Access Journals(DOAJ) (<https://doaj.org/>)の順に適用した
- ・研究の形式の分類：プライマリアウトカムについて、「記述疫学」、「探索的研究」、「治療効果研究」、「その他」に分類した。
 - ・プライマリアウトカムが明示されていない場合、一番初めに記載されている目的を対象とした
 - ・各群の単純な比較を行った場合は、記述疫学に分類した
- ・研究デザインの分類：主として用いられた研究デザインを分類した

1. <https://www.jmdc.co.jp/>
2. <https://www.mdv.co.jp/>
3. <https://www.mhlw.go.jp/index.html>
4. Hirose N, et al. *Annals of Clinical Epidemiology* 2020;2(1):13-26

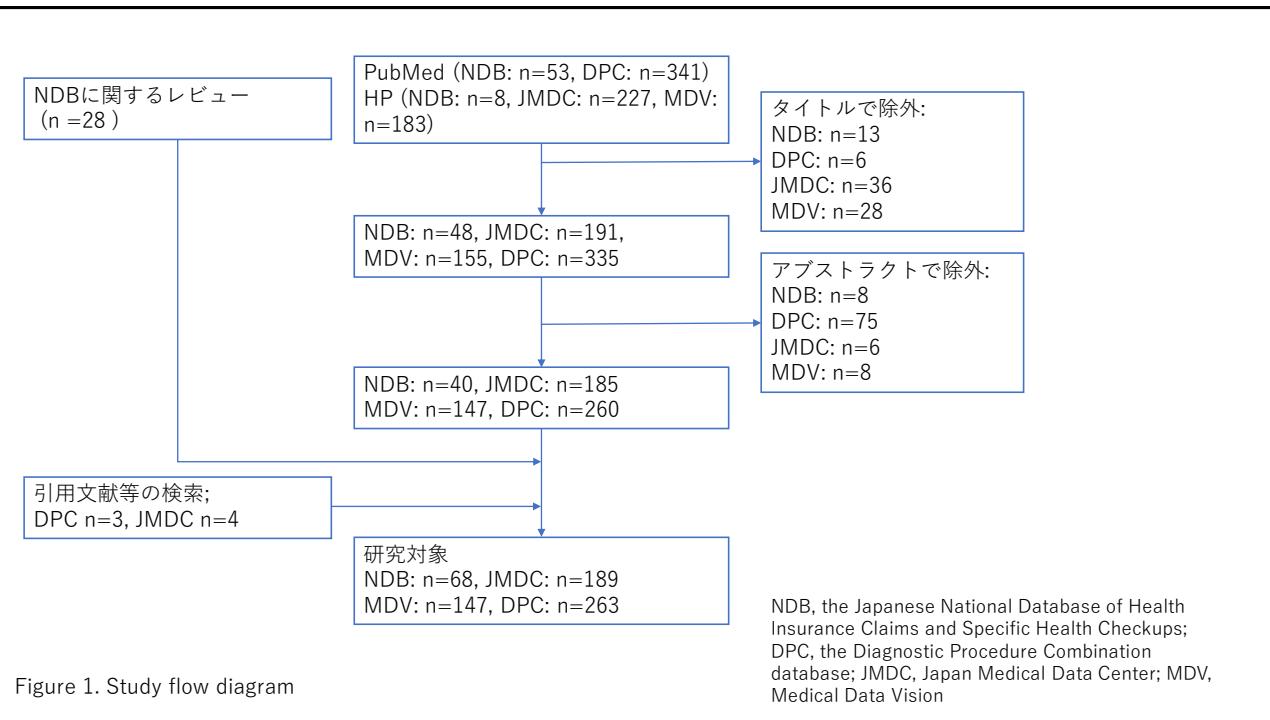


Figure 1. Study flow diagram

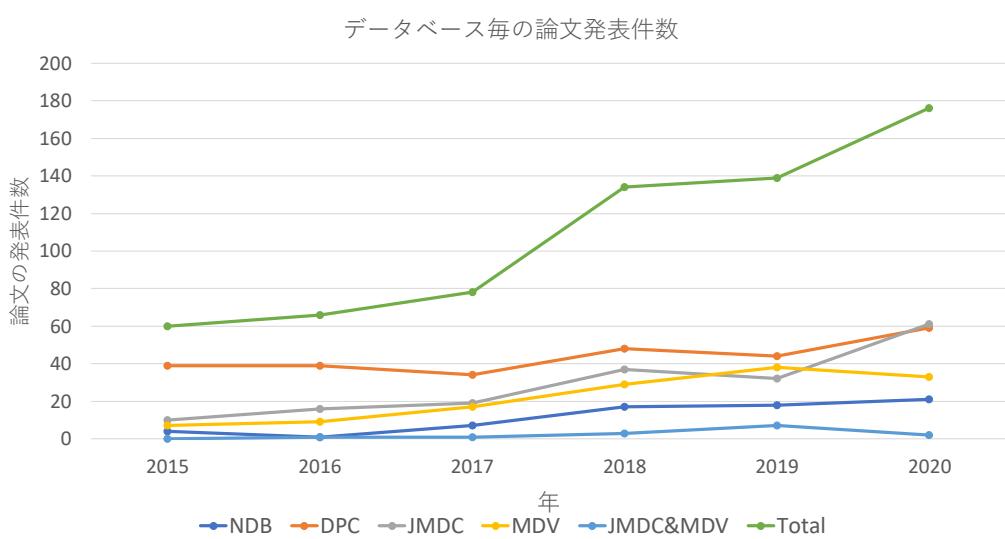


図 2. 各年の論文発表件数

NDB, the Japanese National Database of Health Insurance Claims and Specific Health Checkups; DPC, the Diagnostic Procedure Combination database; JMDC, Japan Medical Data Center; MDV, Medical Data Vision

表 1. 各データベースにおける研究の形式

	計 n=653	NDB n=68	DPC n=263	JMDC n=175	MDV n=133	JMDCとMDVの組み合わせ n=14	p-value
記述疫学	218(33.4)	43(63.2)	35(13.3)	73(41.7)	56(42.1)	11(78.6)	
探索的研究	78(11.9)	6(8.8)	31(11.8)	22(12.6)	19(14.3)	0(0.0)	
治療効果研究	332(50.8)	19(27.9)	184(70.0)	71(40.6)	55(41.4)	3(21.4)	
その他	25(3.8)	0(0.0)	13(4.9)	9(5.1)	3(2.3)	0(0.0)	<0.001

NDB, the Japanese National Database of Health Insurance Claims and Specific Health Checkups; DPC, the Diagnostic Procedure Combination database; JMDC, Japan Medical Data Center; MDV, Medical Data Vision

表 2. 各データベースにおける研究領域 (n ≥ 10)

Research areas, n(%)	計 n=645	NDB n=68	DPC n=260	JMDC n=173	MDV n=130	JMDCとMDV組み合わせ n=14
Medicine, General & Internal	53(8.2)	6(8.8)	14(5.4)	15(8.7)	14(10.8)	4(28.6)
Pharmacology & Pharmacy	45(7.0)	8(11.8)	1(0.4)	21(12.1)	15(11.5)	0(0.0)
Endocrinology & Metabolism	41(6.4)	7(10.3)	2(0.8)	14(8.1)	15(11.5)	3(21.4)
Cardiac & Cardiovascular Systems	40(6.2)	0(0.0)	17(6.5)	8(4.6)	13(10)	2(14.3)
Health Care Sciences & Services	32(5.0)	3(4.4)	4(1.5)	13(7.5)	10(7.7)	2(14.3)
Peripheral Vascular Disease	31(4.8)	2(2.9)	17(6.5)	3(1.7)	6(4.6)	3(21.4)
Multidisciplinary Sciences	29(4.5)	2(2.9)	8(3.1)	12(6.9)	7(5.4)	0(0.0)
Clinical Neurology	28(4.3)	2(2.9)	14(5.4)	7(4.1)	5(3.9)	0(0.0)
Infectious Diseases	28(4.3)	10(14.7)	8(3.1)	1(0.6)	9(6.9)	0(0.0)
Surgery	28(4.3)	3(4.4)	25(9.6)	0(0.0)	0(0.0)	0(0.0)
Public, Environmental & Occupational Health	23(3.6)	2(2.9)	9(3.5)	9(5.2)	3(2.3)	0(0.0)
Respiratory System	22(3.4)	1(1.5)	15(5.8)	4(2.3)	2(1.5)	0(0.0)
Critical Care Medicine	20(3.1)	0(0.0)	20(7.7)	0(0.0)	0(0.0)	0(0.0)
Gastroenterology & Hepatology	19(3.0)	0(0.0)	11(4.2)	4(2.3)	4(3.1)	0(0.0)
Urology & Nephrology	16(2.5)	2(2.9)	10(3.9)	1(0.6)	3(2.3)	0(0.0)
Oncology	15(2.3)	0(0.0)	11(4.2)	1(0.6)	3(2.3)	0(0.0)
Anaesthesiology	14(2.2)	0(0.0)	6(2.3)	2(1.2)	6(4.6)	0(0.0)
Paediatrics	14(2.2)	2(2.9)	6(2.3)	6(3.5)	0(0.0)	0(0.0)
Rheumatology	14(2.2)	2(2.9)	3(1.2)	8(4.6)	1(0.8)	0(0.0)
Emergency Medicine	11(1.7)	0(0.0)	10(3.9)	0(0.0)	1(0.8)	0(0.0)
Geriatrics & Gerontology	11(1.7)	1(1.5)	4(1.5)	3(1.7)	3(2.3)	0(0.0)
Medicine, Research & Experimental	11(1.7)	1(1.5)	6(2.3)	3(1.7)	1(0.8)	0(0.0)
Orthopaedics	11(1.7)	1(1.5)	8(3.1)	0(0.0)	2(1.5)	0(0.0)
Psychiatry	10(1.6)	2(2.9)	2(0.8)	6(3.5)	0(0.0)	0(0.0)

表 3. データベース毎の研究デザイン

研究デザイン, n (%)	Total n=653	NDB n=68	DPC n=263	JMDC n=175	MDV n=133	JMDC&MDV* n=14	p-value
記述疫学	218(33.4)	43(63.2)	35(13.3)	73(41.7)	56(42.1)	11(78.6)	
コホート研究	195(29.9)	11(16.2)	108(41.1)	42(24.0)	33(24.8)	1(7.1)	
横断研究	14(2.1)	5(7.4)	2(0.8)	6(3.4)	1(0.8)	0(0.0)	
症例対照研究, マッチドコホート研究	23(3.5)	1(1.5)	5(1.9)	11(6.3)	6(4.5)	0(0.0)	
Propensity score, IPTW, MSM, g-formula	134(20.5)	2(2.9)	96(36.5)	16(9.1)	20(15.0)	0(0.0)	
自己対照研究 (Case-cross over, Self-Controlled Case Series)	6(0.9)	1(1.5)	0(0.0)	4(2.3)	1(0.8)	0(0.0)	
回帰不連続デザイン, 差分の差分法, 操作変数法	11(1.7)	3(4.4)	3(1.1)	3(1.7)	2(1.5)	0(0.0)	
予測モデル, 臨床スコア	11(1.7)	0(0.0)	8(3.0)	2(1.1)	1(0.8)	0(0.0)	
機械学習	5(0.8)	0(0.0)	1(0.4)	1(0.6)	3(2.3)	0(0.0)	
Cost-effectiveness analysis, Cost-minimization analysis	12(1.8)	1(1.5)	2(0.8)	2(1.1)	6(4.5)	1(7.1)	
その他	24(3.7)	1(1.5)	3(1.1)	15(8.6)	4(3.0)	1(7.1)	<0.001

NDB, the Japanese National Database of Health Insurance Claims and Specific Health Checkups; DPC, the Diagnostic Procedure Combination database; JMDC, Japan Medical Data Center; MDV, Medical Data Vision

* Studies using JMDC and MDV together.

結果のまとめ

- 653件の研究を検討した
- 発表件数増加は2017年以降に著しかった
- 形式：治療効果研究はDPCデータベースにおいて実績が多く、NDBでは記述研究が6割以上を占めた
- 領域：“Medicine, General & Internal” が最も多かった
- デザイン：記述研究、コホート研究が多かった



考察

- 2017年以降の研究の増加は同年の法改正を背景としているか
- 治療効果研究がDPCにおいて多く行われている理由としては、患者背景や重症度の調節が行いやすいことが考えられる¹
- (おそらく同様の理由で) Propensity scoreやそれを用いた重みづけ (IPTW) など、近年観察研究で多く用いられる手法が、DPCを利用したデータベース (JMDC, MDVを含め) においても用いられていた
- NDBを用いた研究において記述研究が多い。悉皆性が高いが²重症度などの調整が難しい¹ことが背景

- 
1. Hiramatsu K, et al. *Real World Outcomes*. 2021;8(4):459-480.
 2. Okumura Y, et al. *J Epidemiol*. 2017;27(8):373-380.

考察

- JMDCは保険者ベース、MDVは病院ベースであり、実施される研究の種類やデザインも異なることが予想された
- しかし両データベースとも比較的HETE研究が多く、デザインはコホート研究が多い
- 両データベースともレセプトデータだけでなくDPCデータも使用しており、DPCデータベースと同様に基礎疾患や重症度の調整が可能であり両データベースは同様の結果となった可能性がある



考察

- ・大規模データを用いて政策・ガイドライン変更の効果を評価するには、回帰不連続デザイン（RD）または差分の差分法（DID）は適切である。¹⁻³
しかし適用されたのはごく一部であった
- ・医療における準実験的研究の少なさは以前から指摘されており、以下が提案されている⁴
 - ・準実験的デザインを用いる研究能力の開発支援
 - ・実施と報告に関するガイドラインの開発
 - ・医療プログラムの管理データとレジストリが公的資金を受けるすべての医療機関で利用可能にすること
 - ・バイアスリスク評価法を拡張して系統的レビューに含めること、
 - ・真の交絡変数に関するデータ収集を標準化すること、など
- ・研究デザインの多くがコホート研究であり、高度な統計手法がなくても研究が可能なこと、データへのアクセスやデータの大きさなど、他の障壁が課題となりうることを示唆している⁵
- ・縦断データの利用がコホート研究の実施を容易にすることを示唆している

1. Nishioka D, et al. *J Epidemiol Community Health*. 2021;jech-2021-217296.
2. Zogg CK, et al. *Ann Surg*. 2017;265(4):734-742.
3. Magnani CJ, et al. *J Natl Compr Canc Netw*. 2019;17(7):795-803.
4. Tugwell P, et al. Big-5 Quasi-Experimental designs. *J Clin Epidemiol*. 2017;89:1-3.
5. Hiramatsu K, et al. *Drugs Real World Outcomes*. 2021;8(4):459-480.

考察

- ・様々な研究領域で大規模医療データベースを用いた研究が実施され、Medicine, General & Internal (53件、8.2%)、Pharmacology & Pharmacy (45件、7%) の領域で多く行われていた
- ・大規模データセットの利用は薬剤疫学においてますます一般的になってきている。RWD を用いた医薬品の有効性・安全性の研究である薬剤疫学におけるデータベースの利用は、今後さらに重要性、割合が高まる予想される¹



1. Takahashi Y, et al. *Eur J Clin Pharmacol*. 2012;68(2):123-129.

結論

- 各DBで実施された研究の種類とデザインは、DBの特徴に対応していることが示唆された
- DPCデータからなるDB（DPCデータベース、JMDC、MDV等）では、個人に関する情報を用いたHETE研究が多く実施されていた
- 一方、NDBを用いた研究は記述的研究が中心であり、データの入手に関するハードルや網羅性の影響もあると考えられる
- 傾向スコア分析は多く実施されたが、その他の準実験的デザインは十分に実施されていない
- 縦断的データの入手が可能であったため多くのコホート研究が実施された
- 今後、RWD データベースを幅広く収集・統合し、包括性を高め、価値あるエビデンスを導出することが必要である

