

疫学研究におけるリアルワールドデータ

Real World Data in Epidemiological Study

増成 直美¹⁾ 水津 久美子¹⁾ 畔津 忠博²⁾

MASUNARI Naomi¹⁾, SUIZU Kumiko¹⁾, AZETSU Tadahiro²⁾

要旨

情報通信技術の発展により、大量のデータ、いわゆるビッグデータが容易に集まるようになり、加えてコンピュータの計算パワーが飛躍的に向上したことで、大量のデータを現実的な時間で解析できるようになった。日常の臨床現場で記録され蓄積されている患者データの総称といわれるリアルワールドデータを疫学研究に利用することで、桁外れのサンプル数の妥当性の高い解析が低コストで可能となる。ビッグでオープンなデータに触れる機会として始めた、本学公衆衛生学研究室および栄養教育研究室での取り組みを紹介する。

キーワード：疫学研究、リアルワールドデータ、NDBオープンデータ、二次利用、データサイエンス

Key words : Epidemiological study, Real world data, NDB open data, secondary use, data science

はじめに

近年、大量のデータから科学的に意味のある情報を見つける学問である「データサイエンス」の重要性が注目され、大学教育においても広く組み込まれ始めている。データサイエンスは、統計学や数学、コンピュータサイエンスをベースとし、さらに情報工学やデザイン情報学など複数の分野を掛け合わせた学問といわれている¹⁾。インターネットの登場により、大量のデータ、いわゆるビッグデータが容易に集まるようになった。加えて、コンピュータの計算パワーが飛躍的に向上し、大量のデータを現実的な時間で解析できるようになった。

康永秀生教授²⁾、山下武志医師³⁾のご研究によれば、これまで疫学研究のゴールド・スタンダードは、無作為化比較試験 (Randomized Control Trial : RCT) であった。RCTは、介入群と対照群を無作為に割り付けるものであり、適応交絡を除外するうえで最も優れており、さらにプラシーボ効果の影響も除外できる。他方で、RCTは、倫理的・費用的な問題などから実施困難であることが多い。また、膨大な資金を要するため、RCTを数多く実施することが不可能となってきた。さらにRCTは、日常臨床とは異なる実験的な条件下で行われ、しかも参加する意欲のある患者に対してだけ行われる。RCTに参加できるような患者は、臨床現場ではむしろ少数派である。その結果、リアルワールドの臨床とはかけ離れた状況となっている。すなわち、被験者は制限され、高齢者や小児、妊産婦は除外され、追跡期間は短い。このようなRCTの欠点や限界を背景として、いわばRCTの反対命題として、近年は大規模なリアルワールドデータを用いた観察研究への注目が集まりつつある^{2,3)}。

1) 看護栄養学部栄養学科

2) 国際文化学部文化創造学科

1 リアルワールドデータ

リアルワールドデータ (Real World Data : RWD) とは、日常臨床現場で記録され蓄積されている患者データの総称である²。米国の薬事規制当局である食品医薬品局 (FDA) によれば、「さまざまな情報源から日常的に収集された患者の健康状態や医療の供給に関するデータ」と定義されている⁴。RWD研究は、臨床研究のゴールド・スタンダードとされるRCTのような特殊環境ではなく、まさに現実の世界を反映したデータであり、臨床現場のデータを用いた、広義の「観察研究」に属する³。

RCTは、厳格に管理された理想的な医療環境下で実施され、高齢者や合併症のある患者は対象から除外される。RCTで評価されているものは、効果が最大限に発揮できそうな環境で、しかも効果が高そうな対象に絞った場合の、つまり理想の世界における「有効性 (efficacy)」である。一方、RWDを用いた研究は現実の世界における「有用性 (effectiveness)」を評価している。言い換えれば、RCTは “Can it work?” を問うており、RWDは “Does it work?” を問うている⁵。

RCTはランダム化によって、未測定交絡を含めて、交絡の影響を排除できる可能性が高い。一方でRWD研究は、治療の選択がランダムでなく、交絡の影響が避けられない。RWD研究は、観察研究のための応用的な統計手法をいかに駆使しようと、交絡を完全に排除することはできず、その点ではRCTにおよぶところではない²。

RWD研究の利点として、多様な背景をもつ多数の患者集団を対象にした、現実の世界で実際に利用されている治療法をありのままに観察したデータが得られることが挙げられる。データベースによっては、桁外れの症例数が確保でき、サンプル集団以外にも一般化できる、外的妥当性の高い研究になりうる。そうした研究がRCTとは比較にならないくらいに低コストで実現できる²。

観察研究は、交絡の影響を受けやすく、これまでRCTよりもワンランク下に見られてきた。しかし、大規模なRWDを用いて厳密な統計解析を施した観察研究は、RCTに準じる結果をもたらしうる⁵。RWDを用いた質の高い観察研究は、RCTの役割を十分に補完し、医療の質の向上に大きく貢献することが期待できる²。

RCTは倫理的、費用的な問題などから実施困難であることが多い。そこで、RCTを補完する手段として、大規模なRWDを用いた観察研究デザインによる質の高い臨床研究が世界的に急増している。特殊環境のもとで限られた答えしか得られないRCTでは、「エビデンスの隙間」を埋めきれないためといわれる²。

以上のように、疫学研究においても、RWDを利用した研究は増加している。そこで、教育現場においても、これからの疫学研究手法に、RWDは避けて通れないのが現状である。

2 NDBオープンデータ

RWDの一つとして、大規模なデータの収集・蓄積・管理するインフラである保険データベースの一つとされる、厚生労働省のレセプト情報・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 : NDB) がある。

今日の医療の高度な発展の下で、きわめて低くなったイベント発生率を生物統計学的に解析するためには、患者数、イベント数が足りないという状況が生まれている。医療が進歩すればするほど、RCTの限界が明らかになってきた。そのような状況でも使えるデータとして、国民の保険データベースが挙げられた。国が医療に要する予算を管理するために保有しているデータベースであり、新たに膨大な研究資金を要さない。患者数とイベント数は膨大であるので、非常に小さなイベント発生率に対しても生物統計学的な解析が可能となる。国が管理しているので、患者の転居や転院も把握することが可能であり、追跡不能者は極端に少なくなる²。

NDBは、2008年4月から施行されている「高齢者の医療の確保に関する法律」に基づき、医療費適正化計画の作成、実施および評価のための調査や分析などに用いることを目的として、厚生労働省がレセプト情報、および特定健診・特定保健指導情報を収集・構築したデータベースである⁷。

NDBに格納されているレセプトデータは、全国の審査支払機関から匿名化処方された2009年以降の電子

レセプトを収集したものであり、紙レセプトは含まれていない。また、公費負担部分については、紙も電子レセプトも含まれない。

特定健診等情報については、全国の特定健診等実施機関で行われた特定健診と特定健診指導の情報が各保険者に提出され、匿名化処理後に社会保険診療報酬支払基金に集約されたものが、NDBに格納されている⁸。特定健診とは、生活習慣病の予防のために、40～74歳を対象に、メタボリックシンドロームの早期発見および予防に着目した健診のことである。また、特定健診指導とは、特定健診の結果に基づき、生活習慣病の発症リスクが高い対象者に対し、専門スタッフ（医師、保健師、看護師、管理栄養士）が生活習慣の見直しをサポートするものである⁹。NDBでは、特定健診については問診結果や生活習慣病に関連した測定項目の結果が含まれ、特定保健指導については保健指導内容についての項目が含まれている。NDBは、世界でも有数の規模と悉皆性を誇るデータベースとなっており、国民の医療の動向や健康状態を把握できるため、有用なデータとして研究機関はもちろん、さまざまな分野で利活用が進められている。

NDBオープンデータは、厚生労働省が2009年より電子化されたレセプト情報ならびに特定健診・特定保健指導情報を収集し、構築したNDBを誰もが利用できるように、日本の医療におけるビックデータをあらかじめ定式化された基礎的な集計表として、厚生労働省のホームページに公開しているものである¹⁰。

NDBオープンデータには、医薬品の処方件数をはじめ、NDBに含まれているさまざまな各種情報の分布が掲載されている。厚生労働省の公表データは、①医科診療報酬点数表項目、②歯科診療報酬点数表項目、③歯科傷病、④薬剤データ、⑤特定保険医療材料、⑥特定健診検査項目、⑦特定健診質問票項目に関し集計表が作成されており、また一部項目はグラフが作成されている。①～⑤の集計表では「都道府県別」および「性・年齢別」の集計を、⑥、⑦の集計表では「都道府県別／性・年齢別」のクロス集計が行われ、集計対象①の集計表では一部の診療行為に対し「都道府県別／性・年齢別」のクロス集計が行われている。

3 本年度の取組み

本学栄養学科公衆衛生学研究室および栄養教育研究室においても昨年度から、NDBのオープンデータを利用して、地域相関の視点から山口県民の健康課題の抽出を試みている¹¹⁻¹³。

生活習慣病予防、改善のための新しい取り組みとして始まったのが特定健診であり、その結果を基に、個々人に合わせたやり方によって生活習慣を見直す支援を行うのが特定保健指導である。その際、対象者の検査結果を基に、問題点の分析や改善への道筋を示す面談をし、食生活や運動等のアドバイスを行う特定保健指導員には、医師、保健師、看護師、管理栄養士のいずれかの資格が必要とされる。保健師の資格を有する者が特定保健指導員になった場合は保健師の資格の範囲内で指導を行い、管理栄養士の場合は栄養管理の面での指導をするといった具合である。すなわち、各方面のエキスパートが集まって相談者の健康管理を手助けするしくみの運用を目指している。このようなしくみの中で、まさに管理栄養士、看護師、保健師の養成機関である本学看護栄養学部において、NDBオープンデータの解析技術を習得することは、近年のビッグデータ社会において重要な意義を有する。

2020年度、筆者らは初めてオープンデータを利用して、卒業研究「NDBオープンデータを活用した平均寿命と健康寿命に関連する要因の解析」を行った¹¹。その過程において、オープンデータ活用のメリットを十分経験することができた。他方で、使い難さも体験した。2020年度の解析は、4年生という比較的経験豊かな学生が1年半という時間を費やして行ったものであった。

量的研究において、本学学部生では栄養学科の場合には1年半という研究期間内に、研究計画を立て、データを収集し、解析・検討するのは、時間的に厳しいものがある。ところが、厚生労働省や各府省が公表する統計データの一つにまとめ、統計データの検索をはじめとした、さまざまな機能を備えた政府統計のポータルサイトである「政府統計の総合窓口 (e-Stat)¹⁴」が提供しているビッグでタイムリーで正確なオープンデータを利用すれば、効率よく量的研究を体験することができる。

オープンデータには、倫理面での問題が少なく使用しやすいという利点がある。反面、オープンデータを利用した地域相関研究では、年齢調整など、別途データマイニング作業が必要になる等の課題もある。昨年

度の末永ら論文作成の過程で¹¹、年齢調整に非常に多くの時間がかかることを経験した。

既存のデータベースに集約された医療情報データをRWDとして二次利用する場合は、これらの医療情報データは研究目的や研究計画とは無関係に作成・収集されるデータであり、研究への利用を想定したデータの入力や質の管理が行われていないことも多い。研究計画の段階で、研究に必要なデータ項目が含まれていることを確認したうえでデータベースを選んだ場合であっても、そのすべてのデータを即座に解析に用いることは適切でない¹⁵。多くの場合において、その結果も踏まえて解析ができるデータとなるよう、研究目的に応じてデータを加工する作業を実施する必要がある。

昨年度の末永ら論文では¹¹、データセット作成のために、学生と教員5人でおおよそ2週間を費やした。そこで、本年度は、プロジェクトチームを結成し、年齢調整を半自動化するプログラムを開発し¹⁶、多くの学生にビッグデータに触れる機会を経験してもらうことを計画した。実際に、本学看護栄養学部栄養学科4年生に開発段階から参加してもらい、ある程度形になってから同栄養学科3年生、看護学科2年生にも、当該プログラムの使用を体験してもらうことができた¹⁶。当該年齢調整半自動化プログラムを利用することで、今年度は、学生1人おおよそ数日で、解析用データセットを用意することができた。

おわりに

NDBの研究利用上の利点として、日本全体で実施された保険診療のほぼすべてが網羅された悉皆に近いデータベースであり、母集団代表性に優れていることが挙げられる。しかし、NDBデータには、患者のリスクや重症度を調整するデータが不足している、アウトカムデータも不足している、といったデータそのものの限界も指摘されている¹⁵。

慎重に集積されたRWDとその解析手法の発展は、従来の観察研究の限界を打破する可能性がある³。無作為化しにくい対象・介入・エンドポイント、医療の進歩による低いイベント発生率を科学的に吟味するために必要となる膨大な患者数、観察期間、イベント数等々、これらを同時に解決する科学的な研究手法を考えた場合、RWDを使った手法の展開が期待される³。

引用文献

- 1 特集「AI・データサイエンスの未来」AERA' 21.9.13 No.39.
- 2 康永秀生『超入門！スラスラわかるリアルワールドデータで臨床研究』金芳堂2019年.
- 3 山下武志『リアルワールドデータの真っ赤な真実 一室の山か、ゴミの山か―』南江堂、2017年.
- 4 U.S. Food and Drug Administration. Framework for FDA' s Real-World Evidence Program December 2018. <https://www.fda.gov/media/120060/download> (最終閲覧日 2022.1.10) .
- 5 Issa J Dahabreh, David M Kent. Can the learning health care system be educated with observational data? JAMA. 2014; 312(2): 129-30.
- 6 Hiam Chemaitelly, Patrick Tang, Mohammad R Hasan, Sawsan AlMukdad. Waning of BNT162b2 Vaccine Protection against SARS-CoV-2 Infection in Qatar. NEJM 2021 Oct 06; doi: 10.1056/NEJMoa2114114.
- 7 厚生労働省保険局、レセプト情報・特定健診等情報の提供に関するホームページ https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/iryuu/hoken/reseputo/index.html (最終閲覧日 2022.1.10) .
- 8 内閣官房社会保障制度改革推進本部、第2回医療・介護情報の活用による改革の推進に関する専門調査会、参考資料2 医療・介護情報の活用(厚生労働省提出資料)、平成26年11月11日https://www.kantei.go.jp/jp/singi/shakaihoshoukaikaku/chousakai_dai2/sankousiryuu2.pdf (最終閲覧日 2022.1.10) .
- 9 厚生労働省保険局 医療介護連携政策課 医療費適正化対策推進室、特定健診・特定保健指導について <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000161103.html> (最終閲覧日 2022.1.10) .

- 10 厚生労働省 NDBオープンデータ (mhlw.go.jp) <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000177182.html> (最終閲覧日 2022.1.10) .
- 11 末永瑤葉、周田紗里奈、照屋裕菜、中島海桜、増成直美「NDBオープンデータを活用した平均寿命と健康寿命に関連する要因の解析」山口県立大学看護栄養学部紀要14号9-23頁、2021年.
- 12 兼重美沙季、畔津忠博、水津久美子、寺田亜希、増成直美「死亡率に関する要因についての検討 — NDBオープンデータを用いた解析—」山口県立大学看護栄養学部紀要15号1-15頁、2022年.
- 13 濱口裕太、畔津忠博、寺田亜希、水津久美子、増成直美「リアルワールドデータを活用した健康寿命に関連する要因の解析」山口県立大学看護栄養学部紀要15号17-23頁、2022年.
- 14 政府統計の総合窓口 (e-stat.go.jp) <https://www.e-stat.go.jp/> (最終閲覧日 2022.1.10) .
- 15 佐藤俊哉 = 山口拓洋 = 石黒智恵子編『これからの薬剤疫学 —リアルワールドデータからエビデンスを創る—』朝倉書店、2021年.
- 16 畔津忠博、増成直美、水津久美子「NDBオープンデータを利用するための教材作成の試み」基盤教育紀要2号97-101、2022年.