

共同企画 | 第40回医療情報学連合大会（第21回日本医療情報学会学術大会） | 共同企画

共同企画1

臨床データベースから見えてくること：ビッグデータを医療に活かす

2020年11月19日(木) 09:20 ~ 11:20 A会場 (中ホール)

[2-A-1-04] 臨床データベース・ビッグデータを活用した臨床研究 - 期待と課題

*石田 博¹（1. 山口大学大学院医学系研究科）

*Haku Ishida¹（1. 山口大学大学院医学系研究科）

キーワード：clinical database, real world evidence, nephrology

近年、臨床データベースは、公的機関、学会、あるいは、大学等の施設が中心となり、国全体、あるいは、地域限定などの拡がりの中で、多くの活用の期待をもとに様々な目的を持った構築がなされている。

目的によってそれぞれの症例データベースの対象疾患や調査項目は異なるが、薬剤等を含む様々な医療技術介入に対し RCTができない、あるいは、補足するような診療のリアルワールドにおける有効性やアウトカム評価、経済性等の効率性評価、希少疾患や難病を含む疾病の頻度や自然歴、QOL評価、地域特性の評価や広域での診療の質評価、さらには、AI等を用いた診断支援、意思決定支援などにつながる臨床研究への活用が期待されている。

一方、それらの臨床データベースを用いた研究により生み出されるエビデンスやファクトの質は、蓄積されるデータの品質や広域性、および解析プロセスに依存する。そのため、データの取得や収集、その維持管理などのデータ管理、および、集計や解析などの手法等のデータ活用のあり方が問われる。具体的な課題として、症例登録における負担軽減とそれに伴う電子カルテ等の既存の診療情報の活用を可能とする質の高い正確な情報の作成、および、連携のあり方、診療データにおける欠損等の少ないより正確な現場での登録、および、収集後のデータの縦断的な連続性を含む品質管理、さらには、これらの収集、維持管理における費用、活用に向けたガバナンスのあり方などが挙げられる。

臨床データベースは、その正確性の高いデータが適切に用いられることがその使命であり、そのデータ品質の検証などを公開すると言った透明性を持ち、可用性を高める恒常的な向上が望まれる。

臨床データベース・ビックデータを活用した臨床研究 - 期待と課題 -

石田 博^{*1}

^{*1} 山口大学大学院医学系研究科

Clinical studies using health care database. - Viewpoints from Medical Informatics -

Haku Ishida^{*1}

^{*1} Department of Medical Informatics & Decision Sciences, Yamaguchi University Postgraduate school of medicine

In recent years, the ministry of health and welfare, academic societies, and other organizations generated the health care databases based on the expectation of the usage for various clinical studies. Although the target diseases and accumulated data of each healthcare database differ depending on the purpose, the users can obtain the real-world evidence of the effectiveness or patients related outcomes of various medical technologies or interventions in the real world using the usual statistics or artificial intelligence. The representative examples of the epidemiological study, the supplemental study reinforcing the evidence of the randomized controlled study, and those developing new algorithms will be presented. As the evidence and facts produced through the research using those health care databases depend on the quality of the accumulated data, specific issues are focused on high-quality and accurate registry data with the utilization of various clinical data in the electronic medical records and with the least number of missing data. Besides, the appropriate and transparent methodological procedure to create new real-world evidence should be necessary. The highly accurate and qualified health care databases should have transparency disclosing verification of the data quality and continually improving the availability.

Keywords: health care databases, real-world evidence, nephrology, CDSS

1. はじめに

近年、臨床データベースは厚生労働省や公的機関、学会、あるいは、大学等の施設が中心となり、日本全体、あるいは、地域限定などの拡がりの中で、多くの活用の期待をもとに様々な目的を持って構築がなされている。

その目的によってそれぞれの症例データベースの対象疾患や調査項目は異なるが、薬剤等を含む様々な医療介入に対し Randomized Control Trial(RCT)ができない、あるいは、補足するような診療のリアルワールドにおける有効性や、アウトカムおよび QOL 評価、経済性等の効率性評価¹⁾、希少疾患や難病を含む疾病の頻度や自然歴、地域特性の評価や診療の質評価、さらには、AI 等を用いた診断支援、臨床意思決定支援など実臨床に直接、貢献する臨床研究への活用が期待されている。

本稿では、臨床データベースの代表的な研究活用例と今後の課題について取り上げる。

2. 臨床データベースを活用した臨床研究

2.1 疫学研究

観察研究においては、まず、記述統計により地域や国全体における疾患頻度、関連する要因の違いなどを明確にでき、また、それらが仮説樹立のきっかけとなることも期待される。この中で地域別、あるいは、様々な分類や重症度等から層別化した頻度による発症率差、リスク比、オッズ比などが求められるが、相関は示されても疑似相関であって原因と結果の因果を示すものとは言えないものであることが多い。一方、もともと診療報酬を目的に作られたデータをもとにする NDB などを用いて実施されることで、疾患の誤分類につながることもある。これらのデータソースの違いによっては、異なる結果をもたらすことは容易に理解できる。異なる 2 つの症例データベース

をもとにした慢性腎臓病 (CKD) の病期ステージ別の貧血頻度の割合の違いを示す。²⁾³⁾

2.2 RCT を補完する観察研究

観察研究における結果は、RCT、あるいはそのシステマティックレビュー研究に比べてエビデンスレベルで劣るといったエビデンスピラミッドが浸透し、また、多くがその効能(efficacy)の根拠として臨床試験の結果を用いている。しかし、RCT は事象の因果関係を明らかにする上での内的妥当性の高さ、頑健性が強みである一方、その症例選択における厳密性や介入方法、評価の厳格性が高いことから、一般臨床における対象患者層とは異なり、外的妥当性が低いなどの問題がある。また、観察期間の短さにより頻度の少ない副作用等の事象の検出に限界がある、さらには、RCT が倫理的に困難である場合なども含めて、長期に様々な年代の患者、あるいは、患者背景の異なる対象に対する有効性 (Effectiveness)を確認する目的で、日常臨床における Real World Data (RWD)として症例データベースが用いられることが多い。⁴⁾糖尿病性腎症における SGLT2 阻害薬の腎機能悪化抑制について検討した RCT および、そのエビデンスを確認した観察研究事例を取り上げる。⁵⁾⁶⁾

2.3 予測モデルの構築と臨床意思決定支援

RWD をもとにしたビックデータに対する昨今の Artificial intelligence (AI) の活用はその有用性を示すことで、ビックデータの価値そのものを示すものとなっており、また、臨床意思決定支援システム(CDSS)との発展にも寄与している。AI の応用は診断や予後予測等を目的に様々な試みがなされているが、その課題には、いわゆる Black-Box 問題があり、意味解釈が不可能なロジックがある。現状においては、AI 等による臨床的な意思決定支援の位置付けは、医療者による決断をあくまでも支援すると言ったものであり、そのためには説明可

能な AI 開発が重要である。事例として血液透析患者への死亡予測について取り上げる。⁷⁾

3. 今後の課題

症例データベースは、物理的および機能的なネットワークの進歩と拡大によって、データリソースとしての活用がさらに進むとともにその価値も高まるものと考えられるが、その中で、改善、あるいは解決すべき課題も多い。

3.1 蓄積されるデータの品質とその保証

様々な活用が期待される症例データベースは、その蓄積される情報の品質により生み出される Real World Evidence (RWE)に対する信頼性の度合いも左右されることは容易に理解されるものである。

日本の於ける診療報酬情報や電子カルテ情報に基づく SS-MIX2 の情報を活用した症例データベースにおいて最も課題となるのは病名データである。元々、診療報酬請求に必要な情報であることによってその精度の低さが指摘されている。特に、アウトカムや併存症や合併症などの情報の欠落は、研究結果そのものに大きく影響するものである。⁸⁾

また、検査結果値においても検査法の標準化は進んでいるが、多施設間、あるいは経時的な検査法の変更などが含まれるデータにおける活用には、施設単位でのデータクレンジング、さらには、それら施設をまとめる際のキュレーションの作業は重要である。さらには電子カルテから SS-MIX2 へのデータ出力精度を保証するシステムバリデーションも不可欠である。

3.2 臨床情報を含めた症例データの重要性

現時点での医事情報や SS-MIX2 情報は、検査値や処方・注射情報を除いてカルテに記載されるような具体的な臨床情報に欠けるため、これらの情報をどのように活用可能とするかは重要な課題である。National Clinical Database (NCD) など多くの症例登録においては、これらの診療情報の登録が求められている。一方、多くの診療情報は電子カルテ上に登録されている情報であるものの Narrative な記載であることが多く、将来的な自然言語解析による取得が実現するまでには、その情報取得を目的として電子カルテ上で入力様式が構造化された様々な書式(Template)との連動が望まれる。また、複数の症例データベース間のリンケージにおける相互の情報の利用も課題であるが、個人情報やセキュリティにおける課題の解決が先決である。その一方で次世代医療基盤法による匿名加工データの活用も期待されるが、現行では患者の同意取得への対応が必要なことから、実現には時間を要すると考えられる。

3.3 データ活用における適切なプロセス(図 1)

臨床研究の一つの目的は、事象の因果およびその程度を明らかにすることである。RWD をもとにした観察研究では、まず、そのデータ内容やその質がその研究に目的に適したものかを検討する必要がある。次に、その解析においては、バイアスの影響や交絡を最小限にする必要がある。それらの質の高い因果推論を行うための統計的手法としては、操作変数法、回帰不連続デザインやプロペンシティスコアなどの活用などが求められる。一方、機械学習においても、その導出プロセスの透明性を明確にするとともに、臨床意思決定支援システムへの活用を目的とする場合には Black box を避けるこ

とやその領域における知識との関連性と洞察が不可欠である。⁹⁾

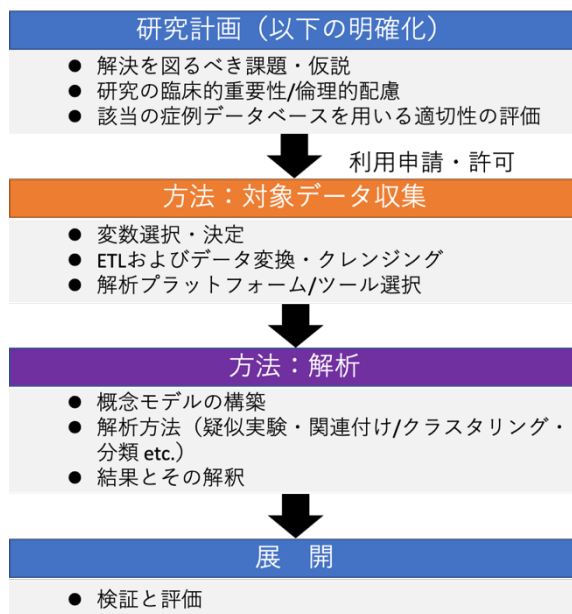


図 1 研究手順

4. まとめ

症例データベースとその研究は、そのデータ量やデータ取得における悉皆性などから、これまでの研究では確認できなかった新たな知見を生み出すことを可能とさせるものである。そのため、医療情報の電子化の拡大とともに、今後もより幅広く、また、長期間に渡るデータ蓄積がなされていくものと考えられる。しかし、その有用性は蓄積されるデータの質の向上と維持のもとに、その適切な活用プロセスに委ねられている。そのような認識、および、その具体的、かつ継続的な対応が重要であり不可欠である。

5. 参考文献

- 1) Pergola PE, Pecoits-Filho R, Winkelmayer WC, et al. Economic Burden and Health-Related Quality of Life Associated with Current Treatments for Anaemia in Patients with CKD not on Dialysis: A Systematic Review. *Pharmacoecon Open*. 2019;3:463-78.
- 2) Kimura T, Snijder R, Nozaki K. Diagnosis Patterns of CKD and Anemia in the Japanese Population. *Kidney Int Rep*. 2020;5:694-705.
- 3) Sofue T, Nakagawa N, Kanda E, et al. Prevalence of anemia in patients with chronic kidney disease in Japan: A nationwide, cross-sectional cohort study using data from the Japan Chronic Kidney Disease Database (J-CKD-DB). *PLoS One*. 2020;15:e0236132.
- 4) Koretsune Y, Yamashita T, Yasaka M, et al. Usefulness of a healthcare database for epidemiological research in atrial fibrillation. *J Cardiol*. 2017;70:169-79.
- 5) Heerspink HJL, Karasik A, Thuresson, et al. Kidney outcomes associated with use of SGLT2 inhibitors in real-world clinical practice (CVD-REAL 3): a multinational observational cohort study. *Lancet Diabetes Endocrinol*. 2020;8:27-35.
- 6) Wanner C, Inzucchi SE, Zinman B. Empagliflozin and

Progression of Kidney Disease in Type 2 Diabetes. *N Engl J Med*. 2016;375(18):1801-2.

7) Kanda E, Epureanu BI, Adachi T, et al. Application of explainable ensemble artificial intelligence model to categorization of hemodialysis-patient and treatment using nationwide-real-world data in Japan. *PLoS One*. 2020;15(5):e0233491.

8) Yamana H, Moriwaki M, Horiguchi H, et al. Validity of diagnoses, procedures, and laboratory data in Japanese administrative data. *J Epidemiol*. 2017;27(10):476-82.

9) Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. *Jama*. 2018;320(21):2199-200.