

## 米国におけるデータアーカイブの研究利用：現状と課題

研究分担者	大橋靖雄	東京大学大学院医学系研究科
研究分担者	祖父江友孝	大阪大学大学院医学系研究科
研究代表者	玉腰暁子	北海道大学大学院医学研究科
研究協力者	瀬戸口聡子	Duke 大学医学部

### 研究要旨

米国で研究に用いられているデータベースの種類、活用事例とその成果、二次利用を進めるにあたっての留意点等について確認した。今後、日本においても公的研究費を受けた研究を適切に二次利用することが求められているが、そのためには、個人情報保護と研究活用とのバランス、事務手続きの標準化・単純化と必要経費、データの適正使用と質保障のためのサポート、共通化によるデータマニピュレーション、情報のロスとデータ容易使用のバランス、共通化プロセスの透明化と公正なシェアの仕組み等につき、議論を重ねていくことが必要と考えられた。

### A. 目的

国内ではまだ十分に活用が進んでいないデータベースの研究利用に関し、米国の事例と留意点を確認し、今後の方向性を考える一助とする。

### B. 方法

実際に米国でデータベースを用いた研究に従事されている Duke 大学瀬戸口聡子准教授より、情報を得た。

### C. 結果

[米国で比較的高頻度に研究活用されているデータベース]

電子カルテデータ: 病院毎に異なるシステムを導入しており、それぞれ研究用 DB 化を試みているが、統一した全国的なデータベースは存在しない

- EPIC - 近年、このシステムを導入する病院が増加

クレームデータ(レセプトデータ): 保険システム毎に存在

- Medicaid - 貧困層
- Medicare - 高齢者・一部の疾患

患者レジストリデータベース: 主に学会が作成登録

- National Cardiovascular Data Registry - American College of Cardiology、主に手技(ステント、PCI など) 毎に患者を登録
- STS National Database - Society of Thoracic Surgeons、胸部外科の各領域(先天性心臓外科、成人心臓病手術、呼吸器手術)手術の登録
- SEER (Surveillance Epidemiology and End Results) Program - National Cancer Act に基づく州単位のがん登録のうち精度のよいもの

コホート研究データベース:二次利用のために公開されているものもあるが、まだそれほど広がっていない。現在、National Heart, Lung, and Blood Institute (NHLBI)が研究費を受けて作ったデータに関しては3年経過後には提出するよう求めている

- ARIC
- Framingham Study
- Women's Health Initiative

#### [データベースの結合利用事例]

米国では、データベース同士の結合も進められており、例えば ARIC、Framingham Study は Medicare Data とリンクされている。

また、データベース内にある ID、あるいは生年月日や病院 ID、入院日を利用して、患者や医師情報を結合することが可能であり、複数のデータベースを用いた研究を行うことができる。治療法による予後を比較、(埋め込み型除細動器と頸動脈ステント術)する研究では、患者レジストリデータベース (Society for Vascular Surgery's Vascular Registry (SVS-VR)と Carotid Artery Revascularization and Endarterectomy (CARE)) に Medicare Data、さらに有料の病院情報、医師情報 (American Hospital Association)などを結合した。このような方法により、コホート研究を1から始めるのに比べ、少ない労力でデータセットを作ることが可能であるが、データ入手の事務作業等で手元にデータが来るまでに2年かかることもある。なお、患者個人のIDを用いなくても、生年月日、医療手技、入院日とプロバイダーIDを利用すれば、95%以上が正しく突合された。また、米国では保険に加入する際サインする同意書に、データの研究利用の項目があるため、保険者がデータを研究利用するために改めて同意を得ることはされず、またIDを削除したデータが販売も行われているのが現状である(ただし、州により対応が異なる場合がある)。

#### [データ二次利用システム事例]

<https://biolincc.nhlbi.nih.gov/home/>

米国では、NHLBI から研究費を受けた行われた RCT とコホート研究を含めた全てのデータを最大利用することを意図して、Biologic Specimen and Data Repository Information Coordinating Center を事務局としたデータデポジトリが行われている。

データ提供のタイミングは、RCT であれば最後の患者登録から3年後、または主目的の論文が公表されてから2年後、観察型の疫学研究では追跡終了の3年後、または付随的な研究から2年後である。その場合、インフォームド・コンセントに沿った形にすること、個人同意情報は削除すること、地域情報は外すことなどがガイドラインで規定されている。データあるいは生体試料二次利用希望者は、HP 上で各研究の詳細を確認し、自身の用いたい研究に対しリクエストをすることができるが、その際、倫理審査を受けておくなどの規則も定められている。

2014年2月時点で確認したところ、134の研究(臨床研究91件、疫学研究42件、両者1件)が登録されていた。なお、2000~10年(2010年時点では登録は76研究)の間に、498のリクエストが認められ、うち74件(15%)が他国の研究所からであり、224論文が公表されたとのことである。

#### [データベース標準化の方向性]

米国では、各所に散らばるデータを1箇所に集めるのではなく、それぞれの場所に置いたまま共同研究ができるようデータフォーマットを統一する distributed network model という試みが始まっている。例えば、OMOP(Observational Medical Outcomes Partnership)は政府と民間の共同によるクレームデータや電子カルテをつなぐデータベースネットワーク構築とリサーチ手法の研究の試みである。その際、項目名、変数の与え方なども全て統一化する(コモンデータモデル)ことにより、複数のデータベースで同時に同じ解析を行うことを目的として、現在米国内の10データソース、1億3000万人をカバーする規模で進められている。また、FDAのmini-sentinelは特に薬の市販後調査を年頭に、同様の試みをしている。ここでは、18の保険会

社データベースを用いており、1億3000万人がカバーされている。

#### D. 考察

米国では、NHLBI から研究費を受けて実施された臨床研究、疫学研究は研究終了後一定期間が経過した後は、公開される仕組みが整えられている。しかし、日本ではまだその議論が始まったばかりであり、バイオサイエンスデータベースセンター(NBDC)が立ち上がってはいるものの、仕組みそのものに関する議論が十分に行われてはいない。また既存データベース同士を個人単位で連結することで、データ価値が倍増することも考えられるが、そのための仕組みは整っていない。

今後、観察が終了したコホート研究の二次利用を適切に進めるためには、以下の点に関し、議論を重ね、体制を構築していくことが重要と考えられる。

- 個人情報保護と研究活用のためのバランス
- 事務手続きの標準化・単純化と必要経費
- データの適正使用と質保障のためのサポート
- 共通化によるデータマニピュレーション、情報のロスとデータ使用簡易性のバランス
- 共通化プロセスの透明化と公正なシェアの仕組み

#### E. 結論

米国でのデータベース活用とその成果、留意点等について確認した。それらを参考に、今後、日本においても公的研究費を受けた研究を適切に二次利用することが求められているが、そのために検討すべき課題を列記した。

#### F. 研究発表

1. 論文発表
  2. 学会発表
- いずれもなし

#### G. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得
  2. 実用新案登録
  3. その他
- いずれもなし