

厚生労働科学研究費補助金 政策科学総合研究事業

(臨床研究等 ICT 基盤構築・人工知能実装研究事業)

分担研究報告書

リアルワールドデータの二次利用加速にむけた多施設データ処理方式の導入の試行研究

(国立がん研究センター東病院における検証)

研究分担者：青柳吉博 国立がん研究センター東病院 臨床研究支援部門 臨床研究推進部 システム管理室

A. 研究目的

わが国では AMED 事業において、臨床研究中核病院による RWD 利活用に向けた基盤構築の取り組みである「臨中ネット」を進めている。2024 年 3 月時点の「臨中ネット」では、各施設に臨中ネット共通 DB を構築し、研究者が SQL を作成し、各施設のデータ提供者に SQL を配布することで、施設ごとにデータ抽出を実施し、データ収集を行う。この方法には、研究者が SQL を作成するという高いハードル、各施設のデータ提供にかかる工数の負荷、安全にデータ授受を行う仕組みの欠如といった問題がある。これらの問題点を解決するために、臨床研究中核病院である京都大学、大阪大学、国立がん研究センター東病院の 3 施設で、研究者が SQL を利用せずとも安全にデータ抽出ができる仕組みについて施行を行い、将来的に全ての臨床研究中核病院での導入可能性についても検討していく。

B. 研究方法

国立がん研究センター東病院では、研究対象のデータを Microsoft 社が提供するクラウド基盤である、Azure 上に格納した。環境構築に関しては、富士通社提供の Healthy Living Platform サービス（以下、HLPF という。）を利用した。分析環境は他機関と同様に Google クラウド上に構築した。研究対象データは電子カルテ系のネットワークにある電子カルテ DWH からのデータを HLPF 上の臨中ネット共通 DB へ手動で転送した。HLPF 上の臨中ネット共通 DB から Google クラウド上の臨中ネット共通 DB へは ODBC 接続することでデータ抽出を可能とした。

なお、ODBC 利用環境を構築するために、以下の設定を行った。

- HLPF 内の臨中ネット共通 DB 環境と Google クラウドの間にバーチャル FW を設置する。
- HLPF 内の臨中ネット共通 DB と Google クラウドとの通信に VPN を利用する。
- 臨中ネット共通 DB と Google クラウド解析環境を ODBC 接続する。

今年度は、京都大学が実施する臨床研究において、電子カルテ DWH から実臨床データを HLPF 上の臨中ネット共通 DB へ転送し、京都大学および大阪大学のデータと共にリアルワールドデータを用いた解析が行えるか否かを確認した。研究者は国立がん研究センター東病院にある端末から、Google クラウド上の計算リソース（Airflow、Kubernetes モジュール）に接続して、データ抽出のためのユースケースを策定し、データ抽出および解析を行った。

また、令和 6 年度はさらに、各施設で利用されるクラウド環境を安全に接続するためのセキュリティ対策について調査を行った。調査方法は Azure において利用されているクラウド間接続に関するベストプラクティスを中心に要件の概要を公開情報から収集した。調査は HLPF を提供する富士通社の支援のものと行った。

C. 研究結果

国立がん研究センター東病院では、HLPF 上に格納されている臨床データに対して Google クラウ

ド環境から ODBC 接続でデータ抽出が可能であることを確認した。本研究を通じて、実患者におけるデータの提供・取得が可能であることが明らかになった。また、クラウド間の接続におけるセキュリティ対策の調査では、トークンを用いたクラウド型情報交換技術仕様の確認やその実装状況を確認することができた。

D. 考察

データ分析では、複数の施設の大量データ抽出によるデータベースの負荷等、RWD ならではの課題が発生すると考えている。国立がん研究センター東病院では、京都大学が実施する臨床研究を通じて、京都大学および大阪大学のデータを国立がん研究センター東病院のデータを統合して解析することができた。この結果は、今後、臨床研究中核病院で行われている臨中ネットのような分散型の研究用データ統合プロジェクトにおいて有用であると考えられた。また、クラウド間接続におけるセキュリティ対策調査の結果、クラウド型情報交換技術仕様を適切に各施設環境に取り入れていくことで、より安全に各施設のデータを共有できる環境ができる可能性が示唆された。この結果は、電子カルテ情報共有サービスで利用される環境と ARO 等の研究機関で利用されるクラウドサービス間を連携する際のセキュリティ対策として発展することが可能ではないかと考えられた。

本研究における課題は各施設を増やした際にデータベースへの負荷が十分に検討できなかったことである。参加施設を増やすことにより、より大量のデータを研究に利用できることが可能となるが、その際に各施設の環境に与える負荷については十分把握することができなかった。国立がん研究センター東病院では、データベースをクラウド上に設置しており、負荷の度合いによってフレキシブルに性能を上げることができるが、その分だけ運用費用等への影響がある。今後は負荷に対する対応などを検討していくことにより大規模な研究を実施することが可能になると考えられた。

E. 結論

臨床研究中核病院である京都大学、大阪大学、国立がん研究センター東病院の3施設で、実患者を利用した分析ができることが確認できた。本研究の成果を基に多規模研究を行う際の負荷等の課題を検討することでリアルワールドデータの二次利用加速にむけた多施設データ処理環境を構築する可能性が示唆された。

F. 健康危険情報

該当なし。

G. 研究発表

該当なし。

H. 知的財産権の出願・登録情報

該当なし。