厚生労働科学研究費補助金（がん対策推進総合研究事業）
（分担）研究報告書

Federated Learningによるデータシェアリング

研究分担者　シャルヴァアドリアン　順天堂大学　国際教養学部　講師
研究代表者　松田智大　国立がん研究ｾﾝﾀｰがん対策研究所国際政策研究部　部長

研究要旨　希少がんや小児がんの研究では、十分な症例数を確保するため国際共同研究が不可欠であるが、GDPR をはじめとする各国の個人情報保護法制により、患者レベルデータの国際移転には大きな制約が存在する。本研究では、個別データを施設外へ持ち出さずに解析可能な Federated Learning（FL）を中心とした分散型解析手法を導入し、大阪国際がんセンター、佐久医療センター、放射線影響研究所、国立がん研究センター、およびオランダ IKNL を接続した国際共同解析基盤を構築した。RARECAREnet Asia において頭頸部がんを対象とした解析を開始し、FL を研究手法として用いた国際共同研究成果が学術誌に掲載された。さらに、G7 Cancer WG1 や WHO WPRO と連携し、GDPR、HIPAA、匿名化・仮名化、Federated Querying、仮想合成データ等の課題整理を進めるとともに、WPRO 加盟国を対象としたデータ共有環境調査を計画し、国際的ながんデータ共有基盤形成に向けた実装可能性を検討した。

A.　研究目的

希少がんや小児がんは症例数が少なく、単一施設あるいは単一国のみで十分な解析規模を確保することが困難である。そのため、発生率、生存率、治療成績、医療格差などを適切に評価するには、国際共同研究による大規模データ解析が必要となる。一方で、近年は GDPR（EU 一般データ保護規則）をはじめとする個人情報保護規制の強化により、患者レベルデータを国外へ移転する従来型の中央集約型研究が実施しにくくなっている。特に欧州では、加盟国内でも GDPR の解釈が異なり、北欧諸国を中心に厳格な運用が進められている。さらに、米国の HIPAA、日本の個人情報保護法など、各国法制との整合性確保も課題となっている。

本研究では、このような国際的課題に対応するため、個別データを施設外へ持ち出さずに解析を可能とする Federated Learning（FL）や Federated Querying（FQ）などの分散型解析手法を導入し、安全性と研究効率を両立した国際共同解析基盤の構築を目的とした。また、RARECAREnet Asia や G7 Cancer WG1 との連携を通じて、実際の国際共同研究への応用可能性を検証するとともに、WHO 西太平洋地域（WPRO）加盟国におけるデータ共有の法的・技術的・文化的制約を把握し、将来的な国際データ共有ガバナンス形成に資する基盤整備を目指した。

B.　研究方法

台湾がん登録において既に導入されている Federated Learning の仕組みを参考に、大阪国際がんセンター、佐久医療センター、

放射線影響研究所、国立がん研究センター、およびオランダ IKNL を接続した分散型解析環境を構築した。各施設は患者レベルデータを自施設内に保持したまま、共通プロトコルに基づくローカル解析のみを実施し、その解析結果やモデルパラメータのみを中央解析側へ送信する構成とした。

RARECAREnet Asia においては、頭頸部がんを対象とし、発生率推定および生存率解析を実施した。発生率解析では Bayesian model を用いて各施設データを統合し、生存率解析では likelihood 情報のみを共有する分散型生存解析手法を採用した。また、患者データそのものを移動させないことにより、各国・各施設の法規制や倫理要件への適合を図った。

さらに、G7 Cancer WG1 の WS4 に参画し、小児がんデータ共有に関する国際的議論へ参加した。IKNL を中心とする欧州側ネットワークと連携し、Federated Learning の実用化に向けた運用上の課題、標準化、法的整合性などについて検討を行った。

法制度面では、GDPR、HIPAA、日本の個人情報保護法について比較整理を行い、匿名化、仮名化、非識別化の概念差異を検討した。また、データビジティング、Federated Querying、仮想合成データ、PET（Privacy Enhancing Technologies：暗号化、差分プライバシー等）などの関連技術についても整理し、国際共同研究への適用可能性を検討した。

加えて、WHO Collaborating Centre 指定に向けた活動の一環として、WPRO 加盟国におけるデータ共有状況調査を計画した。調査では、研究利用に関する法制度、データローカライゼーション要件、匿名化基準、国外移転規制、技術基盤等について情報収集を行うこととした。

C. 研究結果

Federated Learning を用いた国際共同解析基盤を構築し、RARECAREnet Asia において頭頸部がんを対象とした解析を開始した。各施設が個別患者データを保持したまま解析を行い、解析結果のみを共有する方式により、国際的な個人情報保護規制に配慮した共同研究が可能であることを示した。また、Federated Learning を研究手法として用いた国際共同研究成果が学術誌に掲載され、分散型解析手法の実用性が国際的に認知された。

G7 Cancer WG1 の WS4 においては、小児がんデータ共有に関する国際的枠組みづくりが進められており、IKNL を中心に Federated Learning の導入・実装が検討されていることが確認された。また、日本側もデータパートナー候補として連携を進めることとなった。

法制度面では、GDPR の解釈が EU 加盟国内でも大きく異なり、研究実務に混乱を生じさせていることが整理された。特に北欧諸国では患者レベルデータ共有が困難となっており、EU 外へのデータ移転には妥当性認定など厳格な条件が必要であることが確認された。一方、日本は EU から妥当性認定を受けているが、米国では十分な制度整備がないため、EU から NCI へのデータ提供が停止する事例も発生していることが把握された。

また、匿名化・仮名化・非識別化について、GDPR と HIPAA の概念差異を整理した。さらに、OMOP CDM や NAACCR など標準

化規格の重要性が確認され、多施設共同研究ではコア項目と詳細項目を区別した階層的データ構造が有効であることが示唆された。

WHO Collaborating Centre 指定に向けた準備も進展し、WPRO 加盟国に対するデータ共有状況調査について WHO 側から協力が得られる見込みとなった。

D. 考察

本研究により、患者レベルデータを施設外へ移転することなく国際共同研究を実施可能な分散型解析手法の有効性が示された。特に希少がんや小児がん分野では、単独国では解析困難な課題に対し、多施設・多国間で安全に症例を統合解析できる点で大きな意義がある。

近年の GDPR 強化により、従来型の中央集約型データ共有モデルは持続困難になりつつある。そのため、Federated Learning や Federated Querying のような「データを動かさず解析を動かす」手法は、今後の国際研究基盤として重要性が増すと考えられる。特に、各施設にデータ主権を残したまま共同解析できる点は、各国の法制度や倫理規制への適合性が高い。

一方で、これらの手法には高度な技術基盤が必要であり、計算環境、暗号化技術、認証管理、通信セキュリティなどの整備が不可欠である。また、入力情報や出力情報からの再識別リスクにも注意が必要であり、PET 技術を含めた包括的対策が求められる。さらに、各国で匿名化や同意取得に関する定義・運用が異なるため、技術のみならず国際的ガバナンスやルール形成が重要となる。

加えて、データの標準化・相互運用性も大きな課題である。OMOP CDM 等の標準化規格を活用しなければ、分散型解析を導入しても国際比較可能性を十分に担保できない。そのため、将来的には解析技術だけでなく、登録定義、コード体系、品質管理を含めた国際標準化の推進が必要と考えられる。

E. 結論

本研究では、Federated Learning を中心とした分散型解析手法を導入し、患者レベルデータを施設外へ持ち出すことなく国際共同研究を実施可能な解析基盤を構築した。RARECAREnet Asia における頭頸部がん解析や、G7 Cancer WG1 における小児がんデータ共有議論への参画を通じて、実際の国際研究環境における実装可能性を示した。

また、GDPR、HIPAA、日本の個人情報保護法など国際的法制度の差異や、匿名化・非識別化の概念整理を行い、国際データ共有における主要課題を明確化した。さらに、Federated Querying、PET、仮想合成データ等の関連技術についても検討を行い、今後の安全な国際共同研究基盤形成に必要な方向性を整理した。

今後は、WHO WPRO 加盟国調査を通じて地域ごとの法的・技術的課題を把握し、国際標準化、ガバナンス形成、相互運用性向上を進める必要がある。特に希少がん・小児がん分野において、分散型解析技術を基盤とした持続可能な国際研究ネットワークの構築が期待される。

F. 健康危険情報

なし

G.　研究発表

　なし

H.　知的財産権の出願・登録状況

１．特許取得

該当なし

２．実用新案登録

該当なし

３．その他

該当なし