RO-006

長期追跡研究のための複数機関にある匿名化データの共有における セキュリティ対策の検討

Security Measures to Share De-identified Data among Multi-centers for Longitudinal Studies

白石 善明† **中井 敏晴**‡ **毛利 公美**†† **福田 洋治**‡‡ **廣友 雅徳**††† **森井 昌克**†
Yoshiaki Shiraishi Toshiharu Nakai Masami Mohri Youji Fukuta Masanori Hirotomo Masakatu Morii

1．まえがき

病気の早期発見や発症前の予防法の確立のためには，大規模な集団の長期観察によって個人の健康に関わる多様な情報や生体試料を蓄積し，分析する大規模な追跡研究が必要とされている[1]．臨床研究では，提供者の同意を得た上で，個人情報保護のために匿名化された試料・情報（**臨床研究用データ**）が取り扱われる．匿名化とは，試料・情報から個人を特定できないように，氏名，住所などといった情報（**個人識別情報**）を取り除き，新たに識別子などを付すことである．

匿名化の方法には連結可能匿名化と連結不可能匿名化の 2 つがある．連結可能匿名化は，臨床研究用データと匿名化で取り除いた個人識別情報の 2 つに，必要な時に提供者を識別できるように当該提供者と新たに付した識別子の対応情報（**連結情報**）を加えた 3 つに分ける方法である．連結不可能匿名化は，個人が識別できないように連結情報を残さない方法である．追跡研究では，臨床研究用データの提供者を識別する必要があるため，試料・情報は連結可能匿名化されて取り扱われる[2].

複数の機関に提供された試料・情報を共有し，分析する多施設研究では，単一の研究機関よりも研究に必要な試料・情報を確保しやすく，国内外を問わず研究成果が示されている[3], [4], [5], [6]．多施設研究で，複数機関に保管されている試料・情報を一元管理して共有する方法論[7]が示されており，国内でも社会保障・税番号制度で個人に割り振られる番号を利用し，臨床研究用データを個人と紐づけた形で一元管理し共有するシステムが想定され，セキュリティ対策が検討されている[8]．ただし，DNA などのセンシティブな情報を扱う場合，個人と紐づいた形で提供元機関の外部に一元管理されることは望ましくない[9].

本研究では，多施設研究に参加する各機関が保管する試料・情報に連結可能匿名化を施し，連結情報を各機関の外部に秘匿しつつ臨床研究用データを共有するシステムのセキュリティ対策を構築することを考える．図 1 に示すように，連結情報が複数の機関に分散して保管されていても，保管先を把握し仮想的に参照できる機関があれば，複数の機関にある同一提供者の臨床研究用データを研究機関の要求に応じて共有できるようになる．本稿では，連結情報を利用し臨床研究用データの提供者を識別する場合に生じる脅威を分析し，その脅威に対抗するために連結情報と個人に紐づいた関連情報に対するセキュリティ対策を検討する．

2．複数機関にある臨床研究用データの共有の流れとその脅威分析

安全な情報システムを構築するためのセキュリティ構築方法論[10]に従ってシステム設計すると次のようになる．セキュリティ機能を除く外部仕様が決定した段階からスタートし，（1）決められた外部仕様をセキュリティの観点から分析する，（2）システムの保護資産を決定し，脅威を網羅的に洗い出す，（3）洗い出された脅威に対策を施す，といった手順で行われる．本章では，（2）の脅威分析までを，次章では（3）の施すべき対策について述べる．

2.1 システムの構成要素

システムの構成要素を説明するにあたり，まず次の用語を定義する．

検索クエリ：複数機関にある特定の提供者の臨床研究用データの検索の要求に使うデータ．
検索ワードは何かわからないようにしてある．

次にシステムの構成要素を図 2 に示し，その役割を説明する．

提供者：研究に利用する試料・情報の提供者．
提供元機関に試料・情報を提供する（0.）．

† 神戸大学, Kobe University
‡ 国立長寿医療研究センター, National Center for Geriatrics & Gerontology
†† 岐阜大学, Gifu University
‡‡ 愛知教育大学, Aichi University of Education
††† 佐賀大学, Saga University

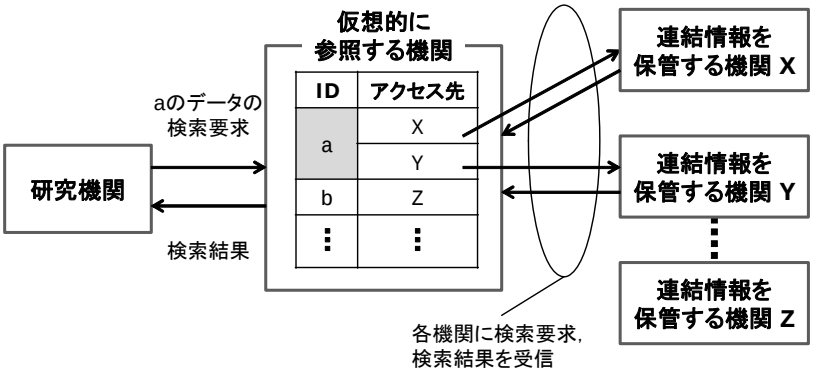

図 1　連結情報の仮想的な参照
Figure 1　Access to re-linking information.

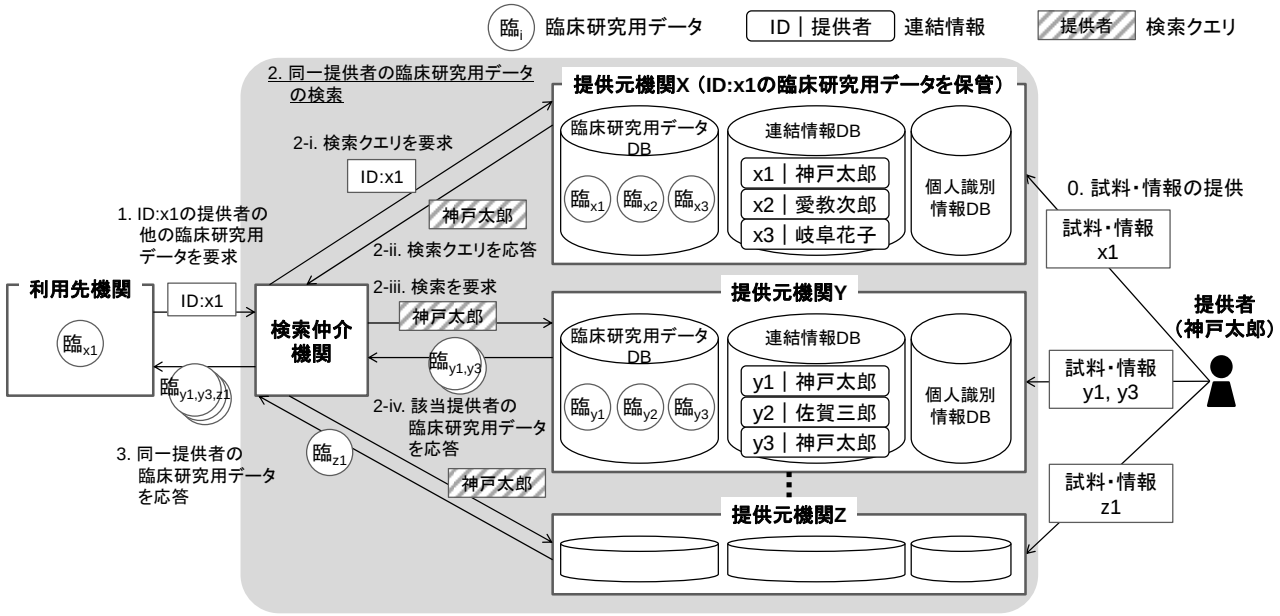

図 2　複数機関での同一提供者の臨床研究用データの共有
Figure 2　Sharing de-identified data among multi-centers.

提供元機関：提供者から試料・情報の提供を受け，連結可能匿名化を施し，臨床研究用データ，連結情報，個人識別情報を別々に保管する機関．検索仲介機関の要求を受けて，臨床研究用データに対応する連結情報を参照し提供者を識別したうえで，検索クエリを作成して返す（2-ii.）．また，検索仲介機関からの要求を受けて，検索クエリを用いて臨床研究用データを検索する（2-iv.）．

検索仲介機関：利用先機関が提供元機関に保管されている臨床研究用データを検索する際に，2 つの機関を仲介する機関．利用先機関からの要求を受けて，提供元機関に検索クエリを要求する（2-i.）．また，複数の提供元機関を対象に，検索クエリに対応する臨床研究用データを検索（2-iii.）し，利用先機関に検索結果を返す（3.）．

利用先機関：臨床研究用データを研究に利用する機関．臨床研究用データに付された識別子を指定し，検索仲介機関に同一提供者の臨床研究用データを要求する（1.）．

2.2 保護資産の決定と脅威の分析

悪意の第三者や内部不正者によって匿名化した臨床研究用データの提供者が間接的に特定されることで，提供者個人に不利益が生じてはならない．そこで，提供者の特定につながる恐れのある，連結情報と個人に紐づいた関連情報をここでの保護資産とする．

表 1　セキュリティ要件と対策
Table 1　Requirements and solution.

セキュリティ要件		セキュリティ対策
提供元機関の個人の特定を行う端末以外で連結情報の参照が行われないこと	DB から情報が漏えいしないこと	・保護対象データの暗号化（*）
	DB への問い合わせで情報が漏えいしないこと	・通信路の暗号化（*） ・検索可能暗号
	必要な情報以外は秘匿すること	・検索可能暗号 ・個人の特定を行う端末からの個人の特定後の連結情報の削除
縦断的データから個人を推定できないこと		・縦断的データからの情報の適切な除去

（*）文献[8]に示されているセキュリティ対策

連結情報と個人に紐づいた関連情報を取り扱う次の 5 つのユースケースを想定し，脅威を分析した．

・試料・情報の登録
・臨床研究用データの更新・原資料の確認
・特定の個人や集団の臨床研究用データ（縦断的データ）の追加取得
・提供者へのフィードバック
・同意の撤回

分析には脅威を網羅的に洗い出す手法の 1 つである 5W1H 脅威分析法を用いた．保護資産（what）を「連結情報」，「個人に紐づいた関連情報」とし，脅威を引き起こす可能性のある主体（who）を「悪意の第三者」，「内部不正者」とし，保護資産を扱うところ（where）や攻撃手段（how）などから次の 4 つが洗い出された．

・連結情報の漏えい
・特定提供者の臨床研究用データの名寄せ
・提供者の疾患・経過の漏えい
・連結情報の未削除

以上の脅威が，提供者の特定につながる恐れがある．例えば，特定提供者の臨床研究用データが名寄せされると，各提供元機関の匿名化で取り除く情報が異なる場合，医療機関の利用履歴や病歴などといった残された情報（行動データ）を組み合わせて提供者を推定されるといったことが挙げられる．

3．複数機関にある臨床研究用データの共有のためのセキュリティ対策の検討

2 章で複数の提供元機関にある同一提供者の臨床研究用データを共有する場合に生じる 4 つの脅威が洗い出され，その過程で保護資産を扱うところや攻撃手段などがわかった．攻撃手段がわかれば，その攻撃を成功させないための要件を導くことができる．複数の機関に保管されている同一提供者の臨床研究用データを仮想的に参照できるように，洗い出された脅威に対抗するセキュリティ要件として a.と b.の 2 つを導いた．a.の要件は 3 つの条件を同時に満たすことに細分化される．

a. 提供元機関の個人の特定を行う端末以外で連結情報の参照が行われないこと
  a-1. データベース（DB）から情報が漏えいしないこと
  a-2. DB への問い合わせで情報が漏えいしないこと
  a-3. 必要な情報以外は秘匿すること
b. 縦断的データから個人を推定できないこと

導いたセキュリティ要件を満たす対策を検討した．表 1 に検討した結果を示す．表 1 のそれぞれの対策とその効果を説明する．“保護対象データの暗号化”と“通信路の暗号化”は，文献[8]で検討されているセキュリティ対策と同様である．

保護対象データの暗号化：第三者が保護対象資産にアクセスできた場合でも，データ自身を暗号化することで情報漏えいを防止できる．

通信路の暗号化：ネットワークを流れる情報の盗聴を防止できる．

検索可能暗号：検索対象を復号することなく検索できる暗号化技術である．暗号化された連結情報を参照する際，暗号文のまま検索できることから該当する連結情報以外の復号が不要である．検索クエリも暗号化されるため，DB に問い合わせる際，検索に関係する情報の漏えいを防止できる．

個人の特定を行う端末からの個人の特定後の連結情報の削除：具体的にはキャッシュの削除

や，メモリに展開したデータを解放することである．データを残さないことで，連結情報を参照する端末に不正アクセスされても，参照中でない限り連結情報の漏えいを防止できる．

縦断的データからの情報の適切な除去：具体的には，縦断的データに含まれる行動データを適切に取り除くことである．複数の行動データを組み合わせた提供者の推定を防止できる．

検討したセキュリティ対策と文献[8]のセキュリティ対策を組み合わせることで，多施設研究において連結情報を各提供元機関の外部に秘匿しつつ臨床研究用データを共有する際の保護資産に対する脅威に対抗できる．

4. おわりに

本稿ではセキュリティ構築方法論[10]に従って，複数の提供元機関にある連結情報を外部に秘匿しつつ，同一提供者の臨床研究用データを検索する場合に生じる脅威を分析した．連結情報と個人に紐づいた関連情報を保護資産とし，5W1H 脅威分析法により，“連結情報の漏えい”，“特定提供者の臨床研究用データの名寄せ”，“提供者の疾患・経過の漏えい”，“連結情報の未削除”の 4 つの脅威が生じることがわかった．その脅威に対抗するための保護資産に対するセキュリティ要件を導き，対策を検討した．検索可能暗号や個人の特定を行う端末からの個人の特定後の連結情報の削除などのセキュリティ対策を施せば，利用先機関が複数の提供元機関にある同一提供者の臨床研究用データを共有するシステムに対する脅威を排除できるようになる．システムのセキュリティ対策を構築するにあたり，今後の課題として縦断的データに含まれる行動データから提供者を特定できないように，個人の特定につながる情報を適切に取り除く匿名化技術を確立することが挙げられる．

参考文献

[1] 日本学術会議第二部ゲノムコホート研究体制検討分科会：提言「100 万人ゲノムコホート研究の実施に向けて」（2013）．

[2] 星野隆之：大量かつ複雑な非構造化データを扱う解析基盤の仕組み，ユニシス技報，Vol.31, No.4, pp.69-77（2012）．

[3] Ulmsten, U., Falconer, C., Johnson, P., et al. : A Multicenter Study of Tension-Free Vaginal Tape (TVT) for Surgical Treatment of Stress Urinary Incontinence, Int Urogynecol J Pelvic Floor Dysfunct., Vol.9, No.4, pp.210-213 (1998).

[4] Fasano, A., Berti, I., Gerarduzzi, T., et al. : Prevalence of Celiac Disease in At-Risk and Not-At-Risk Groups in the United States : A Large Multicenter Study, Arch Intern Med., Vol.163, No.3, pp.286-292 (2003).

[5] Hurria, A., Togawa, K., Mohile, S. G. , et al. : Predicting Chemotherapy Toxicity in Older Adults With Cancer : A Prospective Multicenter Study, J Clin Oncol., Vol.29, No.25, pp.3457-3465 (2011).

[6] 尾長谷靖，金廣有彦，谷本安ほか：吸入ステロイド治療を継続中の喘息患者の吸気流速と背景因子の関連性調査－中国，四国地区多施設研究，アレルギー，Vol.60，No.12，pp.1621-1629（2011）．

[7] Thomas, S.S., Buckon, C.E., Russman, B.S., et al. : Methodology for Developing a Multi-center Clinical Research Study, Pediatric Gait, 2000. A new Millennium in Clinical Care and Motion Analysis Technology, pp.8-15 (2000).

[8] 坂崎尚生，側高幸治，長谷部高行ほか：社会保障・税番号制度の民間利活用における課題の整理と解決策の検討，情報処理学会論文誌，Vol.53，No.9，pp.2194-2203（2012）．

[9] Cambon-Thomsen, A., Rial-Sebbag, E. and Knoppers, B.M. : Trends in ethical and legal frameworks for the use of human biobanks, Eur Respir J., Vol.30, No.2, pp.373-382 (2007).

[10] 秋山浩一郎，鬼頭利之，梅澤健太郎：セキュリティ構築方法論とその支援ツール，東芝レビュー，Vol.60，No.6，pp.40-43（2005）．