

厚生労働科学研究費補助金（がん対策推進総合研究事業）
総括研究報告書

全国がん登録情報の利用及び提供における情報の特性と安全管理措置に関する研究

研究代表者	藤下 真奈美	国立がん研究センター	がん対策研究所	がん登録センター	全国がん登録室長
研究分担者	榊原 直喜	国立がん研究センター	がん対策研究所	がん登録センター	全国がん登録分析室 研究員
研究分担者	伊藤 伸介	中央大学	経済学部	教授	
研究分担者	岩隈 道洋	中央大学	国際情報学部	教授	
研究分担者	上原 哲太郎	立命館大学	情報理工学部	教授	
研究分担者	大木 いずみ	埼玉県立大学	保健医療福祉学部健康開発学科	教授	
研究分担者	佐藤 一郎	国立情報学研究所	情報社会相関研究系	教授	
研究分担者	高橋 克巳	国立情報学研究所	オープンサイエンス基盤研究センター	特任教授	
研究分担者	東 尚弘	東京大学医学系研究科公衆衛生学	教授		
研究分担者	南 和宏	統計数理研究所	データ科学研究系	教授	

研究要旨

がん登録等の推進に関する法律に基づき、2019 年以降、全国がん登録情報は提供可否の審査を経て提供されている。利用者が遵守すべき具体的な安全管理措置については、全国がん登録 情報の提供マニュアル別添「全国がん登録 利用者の安全管理措置」に示されているが、適切な安全管理措置については客観的な評価法・根拠がない。また、利便性向上及び安全性の観点等から、リモートアクセスによる情報利用の検討が求められているが、その安全性基準は設定されておらず、現状を踏まえた見直し及び更新が必要になっている。

本研究では、今後の全国がん登録情報の利活用促進に資するため、全国がん登録情報の利用及び提供における課題を抽出し、情報の特性に応じて講じるべき安全管理措置の検討を行った。

ランダムサンプリングは匿名データの安全性強度を高める手法として有望であり、レコード数が多いがん登録情報の場合、低いサンプリング率による情報損失が相対的に少なく、安全性と有用性の両立の可能性を示すことができた。また、国内外の機関から収集した情報は、わが国におけるリモートアクセスの運用可能性や分析結果のチェック方法・基準等の検討に資すると考えられた。

A. 研究目的

がん登録等の推進に関する法律（以下、「法」）に基づく全国がん登録が 2016 年 1 月から開始され、2019 年以降、全国がん登録情報は提供可否の審査を経て提供されている。

法第 33 条では、情報の提供を受けた者に対し、適切な管理のために必要な措置を講じなければならないことが規定されており、利用者が遵守すべき具体的な安全管理措置については、全国がん登録 情報の提供マニュアル別添「全国がん登録 利用者の安全管理措置」に示されている。しかし、適切な安全管理措置については経験も浅く客観的な評価法・根拠がないことから、提供側の懸念と利用側の利便性のバランスが未解決課題となっている。

また、新型コロナウイルス感染拡大により、職場や研究機関等でのリモート環境の整備が進み、利便性向上及び安全性の観点等から、リモートア

クセスによる情報利用の検討が求められているが、その安全性基準は設定されていないのが実情である。

このように、全国がん登録を取り巻く環境の変化や、複雑なデータ加工が必要な利用申出の増加等に伴い、現行の安全管理措置の内容と実情とが一致しない、または不足した部分があるため、現状を踏まえた見直し及び更新が必要になっている。

本研究では、全国がん登録情報の利用及び提供における運用実績から、課題を抽出し、情報の特性に応じて講じるべき安全管理措置の検討を行い、全国がん登録 情報の提供マニュアルの改訂案を提案することにより、今後の全国がん登録情報の利活用促進に資することを目的とする。

B. 研究方法

①現状の把握及び課題の抽出

a) 情報提供窓口における課題の抽出

全国がん登録情報等の提供窓口での、これまで提供申出における安全管理措置の検討を含め、現行の全国がん登録 情報の提供マニュアルの安全管理措置の運用面における課題、問題点、改善すべき点について整理した。

b) 関連する研究班の研究成果を踏まえた課題の抽出

「がん登録等の推進に関する法律の改正に向けての課題に関する研究」(20EA2003)、「がん登録等の推進に関する法律の改正に向けた課題解決に関する研究」(21EA2002) 及び「全国がん登録の円滑な運用のための検証に関する研究」(20EA0701) の研究成果から、差分プライバシーの手法を使った技術的な検討を行い、データ加工の手法等についての整理を行った。

c) データの安全性に関する国内外の知見と対策事例の収集

海外におけるデータの安全性評価について、欧州連合統計局やオランダにおける公的統計及び医療健康情報の二次利用の最新状況について情報収集を行った。また、リモートアクセスやオンサイト解析等の一部の課題把握のため、国内の関連機関から情報収集を行った。

②情報の特性に応じた安全管理措置の検討

a) 匿名性の強度、情報の特性に合わせた安全管理措置のあり方の検討

①で抽出した課題及び解決に向けた提案について検討し、現行の提供マニュアルにおける安全管理措置の変更案の検討を行い、匿名性の強度に応じた安全管理措置の検討を行った。

b) 物理的及び技術的安全管理措置の見直し

「全国がん登録の円滑な運用のための検証に関する研究」(20EA0701) における匿名化情報の安全性評価の検討状況や、他のデータにおける知見・対策を踏まえ、情報の特性に応じた安全管理措置、対策について見直し、物理的安全管理措置の適切な緩和及び利便性の向上について検討を行った。

また、適切に安全管理されたシステム環境の中で全国がん登録情報等の提供及び利用を行うための、リモートアクセスの実現可能性と求められる安全管理措置についての検討を行った。

C. 研究結果

①現状の把握及び課題の抽出

a) 情報提供窓口における課題の抽出

全国がん登録情報等の情報提供窓口における申出から審議、提供、利用、公表それぞれにおける課題を抽出し、整理した。

b) 関連する研究班の研究成果を踏まえた課題の抽出

従来研究では、データの有用性を保持するために選択できるサンプリング率の上限の値を用いていたが、そのパラメータ設定では安全性の要件を緩めても匿名データの有用性が大幅に低減する逆説的な状況が生じた。この有用性低下の問題を解決するため、サンプリング率の値を下げたところ、同程度の安全性を維持しつつ、データの有用性が向上することが確認できた。更にデータサイズが大きいがん登録情報の場合、低いプライバシー予算 ϵ を設定による強い安全性強度とデータの有用性保持が両立可能であることが示された。

c) データの安全性に関する国内外の知見と対策事例の収集

国内においては、同様のやり方を現在実施している、全国健康保険協会(協会けんぽ)の担当者にインタビューを行い、運用上の状況、導入のきっかけ、予算感等の情報収集を行った。また、総務省関連の統計センターにおいて実務担当者にインタビューを行い、国内の公的統計における情報の取扱いについても情報収集を行った。

更に、国内の医療システムにおける事故の報告書を元に検討した結果、いずれも境界線防衛を何らかの形で攻撃者に突破された結果発生させた事件であることが判明した。VPN 機器はリモートアクセス環境の構築でも多用されるため、がん登録データの利用をリモートで行うためには VPN 機器の管理がセキュリティ上重要であることが示唆された。

国外においては、統計作成部局、研究機関、医療健康データのアーカイブ機関の担当者や専門家にインタビュー調査を通じて、公的統計、行政記録情報、更には医療健康データの二次利用に関して、学術研究目的に対する公的統計の個票データ、行政記録情報と医療健康データの利用サービスの現状及び個票データの秘密保護のための技術的な安全管理措置に関する最新の専門的知見を得ることができた。視察先の欧州諸国では種々のデータの取扱いやルールについてデータ利用者自身の知識が豊富であることに加え、認可を受けている機関の職員は、職場だけでなく、自宅からのリモートア

アクセスも可能という利用環境の違いが判明した。

今後は、これらの知見を踏まえて、全国がん登録における応用可能性について検討を行う予定である。

②情報の特性に応じた安全管理措置の検討

a) 匿名性の強度、情報の特性に合わせた安全管理措置のあり方の検討

法において、特に断りのない限り、基本的に情報は、顕名情報である。匿名化された情報は、その都度「匿名化された情報」と記載されている。これらは適用条文が異なり、扱いも異なる。また、個人識別情報を削除したとしても、提供される情報の内容によっては、個人の同定が可能になるため、プライバシー保護については特に匿名化情報の秘匿について知見が必要となることを意識する必要がある。

b) 物理的及び技術的安全管理措置の見直し

プライバシー保護に際しての安全確保のフレームワークとして、5 Safes (Safe Data, Safe Project, Safe People, Safe Setting, Safe Output) を満たした条件でデータ利用を促進していく必要があると考えられた。

また、データ提供にあたっては、データ管理者の管理下にある解析サーバに対して、利用者がオンサイトあるいはリモートでアクセスして解析をするという方法が、管理強化の観点からも有用と考えられた。

D. 考察

国等が保有する公的データベースの一つである全国がん登録について、現時点ではまだ利用実績が少なく、その利用促進が求められている。しかし、がん登録が本人の同意なく収集している要配慮情報であることを踏まえると、データ利用におけるプライバシー保護については慎重に検討し、利活用促進とのバランスで考える必要がある。

また、情報提供時の情報漏洩及び紛失のリスクの軽減、提供作業の効率化や利便性の向上のためには、リモートアクセス等の共通クラウドで作業・利用できるようなシステムの導入が求められている。しかし、リモートアクセス環境の整備をVPN機器によって構築して安全に保つことは、中小規模の施設にとっては困難が伴うほか、提供する側においても、クラウド型リモートアクセスサービスの選定基準や運用形態に関してもセキュリティ上の配慮が必要となることから、次年度以降引き続き形態について検討することとした。

そのほか、サンプルデータや合成データを使用することにより、少数値の公表などにおいて、安全性が高まると考えられた。

E. 結論

全国がん登録情報等の提供における安全管理措置の課題、問題点、改善すべき点について整理・検討した。

ランダムサンプリングは匿名データの安全性強度を高める手法として有望であり、レコード数が多いがん登録情報の場合、低いサンプリング率による情報損失が相対的に少なく、安全性と有用性の両立の可能性を示すことができた。

また、国内外の統計作成部局や医療健康データのアーカイブ機関から収集した情報は、わが国におけるリモートアクセスの運用可能性や分析結果のチェック方法・基準等の検討に資すると考えられた。

G. 研究発表

なし

H. 知的財産権の出願・登録状況

なし