

厚生労働行政推進調査事業費補助金（がん対策推進総合研究事業研究事業）  
（総合）研究報告書

全国がん登録の円滑な運用のための検証に関する研究

研究代表者 東 尚弘 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 センター長

研究要旨

がん登録等の推進に関する法律に基づき全国がん登録は2016年診断症例以降、全国の病院から義務的届出が開始され、2019年に初年罹患数が初めて発表され、2020年には2017年罹患数が発表されたが、それらの数の動きは制度の変わり目による影響が考えられる。この制度変革による罹患数の影響は今後も追跡して安定を検討すべきである。また、予後情報の精度や利活用における提供データの安全性についても定量的な評価による検討が行われるべきであると考えられる。

本研究では、これらの検討を目標に、それぞれ解析を計画し、最終年度では、データの質評価として制度移行の影響のモニタリング、予後情報の精度について評価を行った。また、データ匿名化の安全性評価の確立を目指し、提供されるデータの安全性について、k-匿名化による評価・検討を行ったほか、医療系マイクロデータであるがん登録情報を対象に、地域情報の匿名化を柔軟に行う匿名化アルゴリズムを開発し、その有用性に関する実証的評価を行った。

これらの研究結果から、診断施設不明例は、制度安定化を評価するための指標の一つになると考えられた。また、k-匿名化及び匿名化アルゴリズムの活用により、より安全かつ有用な全国がん登録情報の提供が可能になると考えられた。

研究代表者 東 尚弘 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 センター長  
研究分担者 祖父江 友孝 大阪大学・大学院医学系研究科・教授  
研究分担者 増田 昌人 琉球大学病院 がんセンター センター長 診療教授  
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  
研究分担者 塚田 庸一郎 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 院内がん登録室長  
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  
研究協力者 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

A. 研究目的

がん登録等の推進に関する法律に基づき全国がん登録は2016年診断症例以降、全国の病院から義務的届出が開始され、2019年に初年罹患数が995,131例発表された。前年の2015年地域がん登録の罹患数903,914例から約9万例増加しており、地域がん登録の毎年数万例程度の増加に比べると急な増加である。これは制度移行の影響と考えられている。

全国がん登録制度の運用の安定化と改善と信頼の

ためには①データの質評価と、②全国がん登録情報の提供におけるデータ匿名化の安全性評価の確立の2点が必要不可欠である。本研究は特にデータの質と安全なデータ利用について、全国がん登録制度の健全な運営を確保するための検証を行い、体制に反映させることが目的である。

B. 研究方法

①データの質評価

全国がん登録の届出数や情報内容の質を評価するため、全国がん登録の運営上で算出される指標を設定し、制度移行の影響などについてのモニタリングを行った。

また、予後情報の精度を評価するために、国立がん研究センター中央病院の院内がん登録の2016年症例、2017年症例（通院継続者を除く）について、住民票照会による追跡等で、生存状況の評価を行った。

②データ匿名化の安全性評価の確立

提供における匿名化個票の安全性確保、データ公表における秘匿性と有用性確保のバランスについて以下のような検討を行った。

1) 匿名化された情報の提供における安全性の検討

2016年、2017年の匿名化された全国がん登録データを用い、提供されるデータの安全性について、k-匿名化による評価・検討を行った。

2) 全国がん登録情報の匿名化指標の開発

がん登録情報の地域情報に国土交通省の位置参照情報を結合し、地域の位置座標に基づき地域領域を柔軟に分割する匿名化アルゴリズムを開発した。この提案手法の有効性を示すため、従来の地域レベルの調整による匿名化アルゴリズムも合わせて実装し、匿名処理で生成されるグループ間の均一性を有用性の指標として両者の比較を実証的に行った。

## C. 研究結果

### ①データの質評価

制度移行の影響のモニタリング指標として、「診断施設不明例」の経過を集計値で追跡したところ、2016年は69,141例(7.0%)、2017年は59,606例(6.1%)、2018年は54,489例(5.6%)、2019年は49,482例(5.0%)と漸減傾向であった。

また、予後情報の精度については、2019年の全国がん登録の死亡情報と突合したところ、院内で生存状況が確認できた16,890名(生存11,327名、死亡5,563名)のうち、死亡が確認されている者で、全国がん登録でも死亡が確認できた者は5,529例(99.4%)、確認できなかった者は34名(0.6%)であった。また、院内で生存が確認された者は全国がん登録でも全て生存が確認でき(100%)、ほぼ実態に近い生死状況を把握できていた。

### ②データ匿名化の安全性評価の確立

#### 1) 匿名化された情報の提供における安全性の検討

基本的な安全性確認のため、ICD-10のみ、ICD-0-3の部位コードのみ、ICD-0-3部位コードと組織型コード、さらに性別、年齢を組み合わせた時のk-匿名化の評価として、ユニーク(k=1)となる症例を集計した。ユニークになるものはICD-0-3の部位分類のみで58件、ICD-10分類では86件、部位組織分類まで含めると4639件であった。

#### 2) 全国がん登録情報の匿名化指標の開発

今回の提案手法をがん登録情報の住所情報に適用したところ、既存の地域レベルを調整する匿名化アルゴリズムと比較して、グループ間の均一性を定量化するDiscernability指標において、5%から16%の改善が確認された。また同一グループに含まれる地域情報の隣接性についても従来手法の結果に比べて際立った改善が実現できることが示された。

## D. 考察

全国がん登録制度の安定化を図るには、データの質評価が重要である。登録数や情報内容の質は、制度としての安定性に関連しており、制度移行における罹患統計への影響を反映した指標としては、診断施設不明例の数・割合、前届出件数、整理症例数割合などが考えられる。本研究では、モニタリング指標として診断施設不明例を用いたが、最新年では5~6%の減少傾向を認めており、今後も精度は向上していくものと考えられた。

また、予後情報の精度については、国立がん研究センター中央病院の院内がん登録症例を用い、住民票照会による追跡等で評価を行った。登録精度については、都道府県によって多少ばらつきがあることに留意する必要がある。

提供されるデータの安全性についてのk-匿名化による評価・検討では、ICD-0-3の部位分類のみ、ICD-10分類のみ、部位組織分類のみの場合においてユニーク(k=1)となる症例を削除しても、全体の件数は200万件以上のためデータの有用性という意味では特に問題ないと思われた。一方で、ICD-0-3の部位・組織分類やICD-10分類、性別、年齢を加えるとユニークな症例が増えるため、必要な項目とその有用性に応じて検討をする必要があると考えられた。

## E. 結論

データ提供における匿名化個票の安全性確保、データ公表における秘匿性と有用性確保のバランスの双方に関して、これまでの検討を踏まえた解析を行った。

これらの研究結果から、「診断施設不明例」は、制度安定化を評価するための指標の一つになると考えられた。また、k-匿名化及び匿名化アルゴリズムの活用により、より安全かつ有用な全国がん登録情報の提供が可能になると考えられた。

## G. 研究発表

なし

## H. 知的財産権の出願・登録状況

なし