

厚生労働科学研究費補助金（がん対策推進総合研究事業研究事業）  
研究報告書

全国がん登録の円滑な運用に向けた研究

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、2016年1月以降、全国の病院にがん診断症例の届出が義務となる全国がん登録が開始された。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について追跡し、予後情報の精度についての検証や提供データの安全性についての定量的な評価により、全国がん登録制度の運用の安定化とデータの精度向上を図ることが、今後の全国がん登録において重要であると考えられる。本研究においてはこれらを目標に、今年度は、データの質評価のためのデータの収集、データ匿名化の安全性評価のための実証評価を行った。具体的には、予後情報の住民票照会に基づく検討、また、匿名化アルゴリズムの実装・評価を行った。今後これらの解析を進めることで、円滑な全国がん登録システムの運営を目指す。

研究分担者氏名・所属研究機関名・職名

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

祖父江 友孝 大阪大学  
大学院医学系研究科  
教授

柴田 亜希子 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長

南 和宏 統計数理研究所  
モデリング研究系 教授

このような集計情報の安定性に関する情報は、精度指標の一つと考えられ、データ活用においても重要である。また、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データの質評価が必要不可欠である。さらに、2019年度から始まった全国がん登録情報の提供の②データ匿名化の安全性評価の確立の2点が必要である。以上のことから、本研究は特にデータの質と安全なデータ利用について、今後の全国がん登録制度の健全な運営を確保するための上記検証活動を行い、今後の体制に反映させることが目的である。

A. 研究目的

がん登録等の推進に関する法律に基づき、2016年1月以降、全国の病院にがん診断症例の届出が義務となる全国がん登録が開始された。2019年に初年の2016年罹患数が995,131例発表され、2015年地域がん登録の罹患数903,914例から約9万例増加していたが、2017年罹患数は977,393例と若干少ない値となった。これらは届出の義務化だけでなく届出症例と前年診断例との名寄せの成功率などに影響された結果と考えられている。

B. 研究方法

① データの質評価

細分化すると a.登録数や情報内容の質、及び、  
b.死亡情報の突合確率、の二つが要検討である。

a 登録数については、初年度、全国がん登録の制度的な安定を示す指標について検討した。試算を行うには、全国がん登録情報を利用する必要があるので、初年度の検討は理論的なものにとどめた。

b.死亡情報については、初年度の2016年症例と同様に、国立がん研究センター中央病院の2017年症例で3年予後情報について住民票情報によって得た生死情報と全国がん登録から得た生存確認情

報とを比較した。3年生存率の算出には、全国がん登録の生存確認情報と住民票情報からの生死情報を突合して、その一致度を計算する予定であるが、現時点では、3年予後の算出が可能な生存確認情報の提供は開始されていないため、この前半部分である住民票照会による予後調査のみを実施した。

#### ② データ匿名化の安全性評価の確立

細分化すると a. 匿名化個票の提供における安全性確保、b. データ公表における秘匿性と有用性確保のバランスの2種類の焦点がある。本年度は、2種類のk-匿名化アルゴリズムを実装し、特に外観識別性の高い地域情報に着目して、一般化処理に関する実証評価を行った。1つ目のアルゴリズムは、住所情報を地域レベルに基づき一般化処理を行う方式である。2つ目は、がん登録情報の住所情報にGPS位置情報を紐付け、広域レベルの地域情報の2分割を再帰的に繰り返すトップダウン型の匿名化アルゴリズムである。2つのアルゴリズムが生成するk-匿名化グループのレコード数のばらつきを比較した。

### C. 研究結果

本年度は、3年計画の2年目であり、以下を行った。

#### ① データの質評価

予後情報の精度を検討するために、国立がん研究センター中央病院の院内がん登録2017年症例の通院継続者を除く症例に対し、住民票照会による追跡を行った。4,005人を調査、3,822人に関して住民票照会が可能であり、死亡2,340名、生存1,399名、不明83名（追跡不能77名、除票4名、その他2名）であった。この結果を3年目に2020年の死亡情報が利用可能になった時点で全国がん登録と情報と突合して検討する。

#### ② データ匿名化の安全性評価の確立

地域レベルの一般化による匿名化アルゴリズムは、効率的に動作するものの地域の人口密度を反映した柔軟なグループ化が困難であり、都市部ではk値を大きく上回るグループが多く生成され、人口の少ない地域ではk値を満たすグループが作

成されず、多くのレコードが削除される結果となった。この結果を踏まえ、GPS基準点に基づく動的な領域分割処理を行ったところ、匿名化グループのサイズが均一化について大幅な改善が実現できた。

### D. 考察

本年度は、3年計画の2年目ということで最終年に向けて、データの質評価のための指標の検討及びデータの収集、データ匿名化の安全性評価のための実証評価を行った。

全国がん登録は、正確な罹患数を把握するうえで重要な役割を果たしているが、地域がん登録からの制度移行の影響などもあり少々不安定なところもある。生存率の集計の前提として、生存確認調査の信頼性の検証を行っておくことは非常に重要である。本年度は2年目として、初年度の2016年症例に続き2017年症例に関する住民票調査を行った。検証結果が提供されるのは1年後となる見込みであるが、作業としては順調に進んでいる。

また、GPS位置情報を参照することで、適切な粒度の地域情報の定義が可能となった一方で、市区町村の境界を超えた地域グループも現状定義されている。このような地域情報が分析者のニーズを満たすものであるかの検討は必要であり、匿名化処理された地域情報の可視化方法を検討する予定である。

### E. 結論

全国がん登録の開始から5年が経過し、データの活用が進み始めた。本年度は、データの質評価のためのデータの収集、データ匿名化の安全性評価のための実証評価を行った。全国がん登録制度の運用の安定化とデータの精度向上について、本研究班においてその検討を行っていく。

### F. 研究発表

なし

### G. 知的財産権の出願・登録状況

なし