

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）

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

研究要旨

本研究においては、がん登録情報のプライバシー保護の方策について検討することを目的とする。がん登録情報は、正確ながん罹患統計を可能にするために同意なく収集されている個人情報であることから、プライバシー保護には細心の注意を払う必要があるため、その具体的な方策の検討のために、まずは、論点整理を行ったうえで、それらを中心に、海外での統計データの利用に際しての方策を中心に情報収集を行った。具体的には11月にヨーロッパ統計局（Eurostat）と、オランダ中央統計局（Central Bureau for Statistics, CBS）を訪問し、それらのデータ提供と、安全管理のプラクティスについて実態をインタビューした。また、日本国内においても、環境省エコチル事業と、独立行政法人統計情報センターの関係者にインタビューをして、前者ではリモートアクセスの仕組みの構築に関する情報収集、後者については、統計表の持ち出しに関するチェックなどを聴取した。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は、5 Safes が良くまわっており、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、リモート解析は、データ管理の方向性としては欠かせないと考えられ、費用が掛かると考えられるものの、試行していくことが必要と思われる。

A. 研究目的

全国がん登録は、法律に基づき同意なく収集しているデータであり、その活用を推進するにあたっては、個人情報の漏洩がないように最大限配慮することが活用に求められている。特にがんの病名は非常に重要なプライバシーともいえる個人情報であるといえる。一方で、収集されたデータの活用を推進することも、必要である。データの秘匿を厳重にすると、データの活用手順に制限が課せられることになり、不便になりがちであり、この両者のバランスをとっていくことが求められる。

本研究においては、我が国における最適のバランスを探るために、

1. 論点整理を行う
2. 海外の事例を収集して学ぶべき原則について整理する。
3. 以上を踏まえて最適のバランスをとるための提案をすることを目的とする。

B. 研究方法

1. 論点整理

これまでのデータ提供実務の経験において課題になったことなどを含めて、データのプライバシー保護と活用に際しての利便性の整理を行う。

2. 海外の事例

11月に研究班員が、欧州ルクセンブルクに位置する・ヨーロッパ統計局（Eurostat）、オランダ・中

央統計局（Central Bureau Statistics）を訪問し、そこから我が国に应用可能な視点、及び対策についてまとめる。

3. 提案

各論点に対して、海外の事例から学んだことを整理し、それをもとに、我が国のがん登録に関して、最適なバランスについての考察を行う。

（倫理面への配慮）

本報告においては公表された情報を元として、検討した事項をまとめるものであり、人を対象とした医学系研究に関する倫理指針の適用外である。特に介入侵襲もなければ、個人情報も扱わない。

C. 研究結果

1. 論点整理

a) 顕名情報と匿名情報

がん登録等の推進に関する法律においては、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。最も異なるのは、複数の都道府県にまたがるデータの利用申し出をした際に、その審査を受けるのが、顕名情報の場合には、厚生労働省の厚生科学審議会であり、匿名情報の場合には国立がん研究センターの合議制の機関となっていることである。

審議の困難さという点では、匿名情報の提供のほ

うが難しい。顕名情報は同意を原則としており、その際の説明が適切に全国がん登録を利用することが理解可能な形で患者が理解できているか、が焦点となる。一方で、匿名情報は同意は要求されておらず、がん登録データは収集時にも同意なく収集されていることから、患者のプライバシー保護はより徹底させなければならない。難しいのは、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ことである。端的な例としては、住所地がわかれば、人が特定される、ということがある。これは特に人口の少ない地域で特定可能性が高くなる傾向にあると考えられるが、さらに地方において、お互いに他の人たちに関して持っている情報も多くなることが予想されるため、何月何日にどの病院を受診したといった情報から、個人が特定されて、病名や進行度がわかるといったことも考えられる。その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) 匿名情報の安全性評価

安全性評価については、個票ベースの安全性と、集計表の安全性評価が考えられる。個票ベースのデータとは、各行が個人（あるいは腫瘍）で、各列に当該個人（あるいは腫瘍）に関する項目の情報が並んでいるものになる。これは、各個人と情報が対応付けられることになるため、より個人識別リスクが高まるものである。ただ、一方で個票データに触れる人員は基本的には解析を担当する研究者等、限られたものであるため、管理方法を規定して個票データが、個人識別を行わない安全な人々の間でのみとすることで、リスクは低減できると考えられる。

c) 集計情報の安全性評価

集計情報については公開する、あるいは広い範囲での共有することが主たる目的となる。そのために、数値とその集計項目そのものの安全性が問われることになる。幸い、集計情報は個票と異なり、多くの項目が対応付けられていくことは少ないため、その点においては比較的安全と考えられる。一方で、公表された情報は、不特定多数が閲覧することや、別々の場所で公表された集計表を組み合わせる可能性があり、その範囲は場所や時間においても蓄積されていく可能性があることから、安全対策はその場で考えられる必要十分な対策よりも、十分なマージンを以て保守的に考える必要がある。特に、少数集計値の報告においては、信頼区間の幅や標準誤差などの統計的な安定性などを含めて、果たして報告の意義があるのか、といったことは慎重に考える必要がある。

d) 安全証明の責任の所在

これまで、統計情報の公表において安全性の基準

は必ずしも確立しているとはいえない。その一因は、「これなら大丈夫」ということが非常に難しいためである。そこである意味、決断した基準で前に進み、決断者が責任をとるという姿勢がどうしても必要になってしまう。また、この基準は上述の通り一定程度保守的な外形基準にならざるを得ないため、外形基準を守っている範囲では、特に責任は問われず、それを越えた公表を希望する場合には、まずは、学術的意義と安全性の比較考量を利害関係のない委員会に対して説得して認められた場合に、希望者の責任において公表を進めるといった手順が必要になると考えられる。この説得基準をどうするのか、が一つの論点であろう。

e) データの活用・管理の方法

これまでがん登録は、データを複製して渡すという原始的な方法のみで提供されてきた。しかし、これでは、データの管理方法などの誓約を取ったとしても、それを検証する方法が、一時的な監査などに限定されてしまう。そこで、おのずと提供できるデータが限られてしまう問題が起きている。

そこで、逆に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る。そのような管理を構築することで、データの散逸を防ぐこともできることから、より詳細なデータを管理下に利用者が使うことが可能になる。もちろん、リモートアクセスにおいては画面の窃視などのリスクや、サイバー攻撃のリスクはあるものの、管理体制が検証不可能な状態での、複製されたデータの提供よりも安全である可能性が高いと考えられる。そのような管理方法と利用可能データのバランスも論点として考えられる。

f) 合成データ、サンプルデータの活用

現状ではがん登録は悉皆データであることを特徴としているが、大元のデータが悉皆データであることの意義は高い一方で、すべてのデータ活用場面において悉皆データが必要なわけではない。特に数が十分存在する対象集団における解析をする場合には、サンプルデータとして悉皆ではなくすることで、データ自体の安全性は向上すると考えられる。これは、サンプル内でたとえ集計値が1（ユニーク）となったとしても、実際には複数人（サンプル確率の逆数の人数）の人がいる可能性が高くなり、また、その人数も不確実となるため、個人識別の程度も同様に確率的となるためである。そのため、サンプルデータであれば、少数値の公表についてもより緩和することが可能と考えられる。

またさらに、合成データを使用あるいは混入させることによりサンプルの集計で出現しているものが現実の個人である確率を調整できるため、安全性

が高まると考えられる。しかし一方で、サンプルデータについてはより細かい解析をする場合に統計的検出力が減退することや、合成データは「ニセモノ」として、真実を追究する研究の姿勢とは逆行するために、研究者からは好まれない可能性が高い。ただ、各変数の分布を保った形での合成データは、実際の解析においても有用であることは知られており、例えば、米国の個票ベースでダウンロードが可能な調査データなどのように、管理方法を気にしなくても良いといった圧倒的な利便性の向上が図られる場合や、データになれる学習目的では、有用と考えられる。

g) わが国のがん登録の特殊性

個人情報保護法における匿名加工情報などは概念的に識別不可能な状態に加工したものと定義されているものの、具体的な加工方法としては、氏名などを加工するなどの一定の基準以上の詳細については個別に検討することとなっている（個人情報保護法施行規則34条第4号：特異な記述を削除する）。これは一般化された記述であるが、がん登録データの取り扱いに関しては、対象となるデータはがん登録と限定されているために、ここで研究を進めることによって安全なデータとは具体的にどのようなものかを同定することは、一般化された加工の方法を検討するよりも具体化しやすと考えられる。そのような試行・研究を進めるべきである。

2. 海外の事例

11月にヨーロッパ統計局（Eurostat）、オランダ中央統計局（CBS）を訪問して、情報収集を行って以下の示唆を得た。

<データ安全性における5 Safesの原則>

データのプライバシーの側面から安全性を確保する上で、5つの側面を確保する必要があるとして整理されている。

- Safe data（安全なデータ）
- Safe projects（安全な研究プロジェクト）
- Safe people（安全な人々）
- Safe settings（安全な環境）
- Safe outputs（安全な出力）

これは健康関連データに限ったことではなく、すべてのデータ管理について作成されている枠組みである。以下に各要素について概説する。

Safe data（安全なデータ）

適切に処置がされて（匿名化・仮名化など）、データ自体が、安全であること。解析目的に必要な範囲でのデータ提供がなされてバランスがとれていること。

Safe projects（安全な分析プロジェクト）

分析がデータの提供者（対象者の利益を代表あるいは責任を持つ主体）によって、分析内容が認められていること。また、個人の識別を目的としない分析であること。

Safe people（安全な人々）

安全管理のための必要な規則を遵守する人物であること。前提条件として規則や手続きを知っていること、また、社会的な立場として安全の脅威になるような立場にないと認められること。

Safe settings（安全な環境）

解析を行う場所が安全であることとは、データに触れる場所が、入室が限定されていることや、ネットワークに接続されるとしたら、それがきちんと管理されていること（パスワードなどのアクセス制限やファイヤウォールの設定など）があげられる。

Safe outputs（安全な出力）

管理された場所から持ち出す、あるいは公表する集計表などの生成物から個人が識別される恐れのない形式（粒度）になっているもの。

これらの原則はEurostat、オランダCBSのいずれでも引用されて、これらの枠組みに沿って、データの扱いに関する規則や審査体制が構築されていた。

<リモートアクセスの方向性>

上記枠組みの中で、Safe settingを確保する一つの方法が、リモートアクセスによる解析である。これは、論点の章で記述した通り、データを利用者の管理下に提供するのではなく、提供者の管理するサーバに設置したままで、解析者がそこにアクセスして解析するものである。そうすることで、利用者が安全な環境を確保する手間を省くとともに、管理体制も徹底することができる。

このような体制を、EurostatもオランダCBSも構築していた。もともとは物理的なオンサイト解析センターとして、特定の場所をそのような形で設置する動きがあったが、コロナ禍のためにオンサイトの運営が難しくなり、また利用者の所在地における不公平もあることから、リモートアクセスの方にシフトしていったとのことである。オンサイトに比べてリモートアクセスは、アクセスする場所の管理は必要になるものの、Safe Peopleに利用を許可していることでカバーされている。また、端末からファイルの出し入れは不可能なシステムとなっており、解析結果は、管理者に依頼して確認をしてから持ち出すこととなっている。これは、オンサイトの運用と同じである。

尚、解析サーバは、現時点ではクラウドではなく、物理的なサーバを使っているとのことだった。

<規則の周知>

規則はせっかく設定しても、それが周知されていないことは問題として以前より挙げられている。これは欧州においても同じであった。特にオランダCBSにおいて対策の話題が上がったが、これは利用に際してテストでチェックすることと、罰則の強化で対応することとなっているとのことだった。

このようなテストは、研究倫理の講習などでも導入されている。同様の守るべき規則を簡単にまとめた動画と、そのポイントについてのテストは有効であると考えられる。(テストはとにかく重箱の隅をついた難問になりがちであることには注意が必要で、周知が目的であることは意識すべきであろう。)

罰則の強化という意味では、施設ごとの連帯責任と取っているということで、何か本当に問題が起きたときには、施設全体が利用を停止するということがあった。そうすることで、施設にも管理責任が生じ、徹底されるとのことである。

<集計値の安全性>

集計値の安全性の評価については、やはり難しいとのこと、外形的な基準と(少数例10例未満は秘匿など)を決めた上で運用しているようであった。この運用の方法としては、基準に抵触していない場合にはそれを確認したうえで、持ち出しあるいは公表を許可するとしているが、基準を満たさない場合には、利用者が安全だと考える理由を主張して、それが説得力があれば(挙証責任が利用者にある。)例外を認める、という方法である。

また、集計値の安全性を利用者自身が自己評価できるようなアプリケーション(SACRO, Semi-Automated Checking of Research Outputs)などを使って、外部評価の前にチェックができる体制も整備しているとのことであった。

<差分プライバシー>

集計表の安全性を評価する際に、少数例が含まれていなかったとしても、ある症例を含む表と、含まない表が公表されて、その差分から個人の識別が可能になるということの評価する技術が、プライバシー保護の研究分野では進められている。その成果を応用していく体制はEuroStatでもオランダCBSでも未構築である。

3. 国内の動き、

リモートアクセスによるオンライン解析を活用することで、データの管理をする計画は、様々な場所で進められている。レセプトナショナルデータベース(NDB)でもシステムが現在構築中という情報もあるし、環境省のエコチル調査においても、現在より簡便なシステムの構築中である

他、総務省管轄の独立行政法人統計センターにおいても、国の統計情報の二次利用についても、オン

サイトセンターの運用と、その持ち出しにおいては統計表の安全性確認を行っている。

D. 考察

データの利用にあたっての、プライバシー保護は利活用の推進とのバランスで考えなければならず、難しい問題である。しかし、がん登録が同意なく強制的に収集している要配慮情報であることを考えると、利用者及び利用を管理する立場からは慎重に検討する必要がある。

もちろん、データはすでに収集されていてそれを十分に公共のために活用すべきであると考え方は重要であるが、Safe projectの原則に基づき、やみくもに利用促進といった姿勢をとることは今後の体制を継続する上で、リスクの高いものになる。

E. 結論

がん登録の安全管理については、最善のバランスを追究すべきであり、特にリモートアクセスによる解析は、デモンストレーションなどを検討する必要がある。

F. 健康危険情報

特になし

G. 研究発表

1. 論文発表

Suzuki S, Nagumo Y, Kandori S, Kojo K, Nitta S, Chihara I, Shiga M, Ikeda A, Kawahara T, Hoshi A, Negoro H, Bryan MJ, Okuyama A, Higashi T, Nishiyama H. The prognostic impact of treatment centralization in patients with testicular germ cell tumors: analysis of hospital-based cancer registry data in Japan. *Int J Clin Oncol*. 2024 Mar;29(3):318-324. doi: 10.1007/s10147-023-02457-0. Epub 2024 Jan 24.

Okuyama A, Mizutani T, Tachibana K, Higashi T, Ogawa A. Treatment patterns and postoperative activities of daily living in patients with non-small cell lung cancer--A retrospective study using nationwide health services utilization data in Japan. *Ann Surg Oncol*. 2024 May;31(5):3409-3416. doi: 10.1245/s10434-023-14798-1. Epub 2023 Dec 21.

Wada K, Higashi T. A retrospective study of sentinel lymph node biopsy for skin cancer in Japan: comparison with breast cancer and evaluation of factors related to its use. *Cancer Medicine* 2023 Dec;12(23):21364-21372. doi: 10.1002/cam4.6677. Epub 2023 Oct 30.

Matsumura K, Nakai H, Kakuwa T, Higashi T Trends in Gynecologic Cancer in Japan: Incidence from 1980 to 2019 and Mortality from 1981 to 2021. *Int J Clin Oncol*. 2024 Apr;29(4):363-371. doi: 10.1007/s10147-024-02473-8. Epub 20

24 Feb 21.

Okuyama A, Kosaka H, Kaibori M, Higashi T, Ogawa A. Activities of daily living after surgery among older patients with gastrointestinal and hepatobiliary-pancreatic cancers: a retrospective observational study using nationwide health services utilisation data from Japan. *BMJ Open*. 2023 May 25;13(5):e070415.

Ogata D, Namikawa K, Nakano E, Fujimori M, Uchitomi Y, Higashi T, Yamazaki N, Kawai A. Epidemiology of skin cancer based on Japan's National Cancer Registry 2016-2017. *Cancer Science* 2023 Jul;114(7):2986-2992.

Watanabe T, Rikitake R, Kakuwa T, Ichinose Y, Nino M, Mizushima Y, Ota M, Fujishita M, Tsukada Y, Higashi T. Time to treatment initiation for six cancer types: An analysis of data from a nationwide registry in Japan. *World Journal of Surgery* 2023 Apr;47(4):877-886. doi: 10.1007/s00268-022-06883-5. Epub 2023 Jan 6.

2. 学会発表

なし

H. 知的財産権の出願・登録状況 (予定を含む)

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

World Journal of Surgery 2023 Apr;47(4):877-886. doi: 10.1007/s00268-022-06883-5. Epub 2023 Jan 6.

2. 学会発表

なし

H. 知的財産権の出願・登録状況 (予定を含む)

4. 特許取得

なし