

## 追跡終了後コホート研究を用いた共通化データベース基盤整備と その活用に関する研究

研究代表者 玉腰暁子(北海道大学大学院医学研究科・教授)

### 研究要旨

疫学研究により得られたデータを広く共有化するためのシステムであるデータアーカイブ化に向けた検討を行った。統計法の規定上、死因情報を付加してのデータ公開・二次利用は認められていないことなどを踏まえて、現制度下での運用方法を2つ提案した。また、実データを用いて、一意性に関する検討を行い、10万人規模のコホート集団はもちろん、100例規模のコホート集団でも分類数が全対象者数の40%程度で、ユニークセルの割合が80%に達していたことから、コホートの規模にかかわらず、80%程度のレコードは一意性があるものとして対応する必要があることを示した。これら研究成果や他国・他領域での実例を参考に、データアーカイブの利活用を進めるため、データ提供の際に従うべきガイダンスに含める必要がある項目をA対象者の個人情報保護、Bインフォームド・コンセントと倫理審査、C知的財産権の帰属、D寄託する項目/しない項目の判断、Eデータ提供先の制限、の5点に整理した。

一方で、エンドポイントとしての死因情報は非常に価値が高いものであることから、追跡が終了したコホート研究の二次利用を進めるために、死因情報のソースとなる人口動態統計調査の有効活用の方策を模索することが望まれ、その際の論点を2つ提示した。統計法の従来解釈経緯などといったコホート研究とは異なる議論の方向性が必要と考えられ、医学分野の知見だけではない法学分野からの知見に照らし、今後、より視野を広げた説得力ある議論の展開が不可欠である。なお、がん登録から得たがん罹患情報の研究目的での二次利用についても、今後同様の検討が必要と考えられた。

### 分担研究者

磯 博康(大阪大学大学院医学系研究科・教授)  
大橋靖雄(中央大学理工学部・教授)  
祖父江友孝(大阪大学大学院医学系研究科・教授)  
辻 一郎(東北大学大学院医学系研究科・教授)

### A. 研究目的

本研究は、国内で実施され追跡を終了した複数のコホート研究情報を共通化し、その利用環境を

整え、将来にわたって向後終了するコホート研究も組み入れ可能な体制を構築するために必要な事項を検討することを目的とした。特に配慮すべき事項の確認のため、先行している国内外の事例収集、実データを用いた一意性の検討、共通化可能項目の検討、アーカイブ化を進めるに際して必要なガイダンス項目の整理を進めた。一方で、統計法の規定により、人口動態統計情報を利用して把握している死因情報は、データアーカイブに付加

して二次利用できないことから、現制度化での運用方法の提案を行うとともに、今後、死因を付加して提供することを可能にするための統計法解釈の論点を提示することを目指した。

## B. 研究方法

米国ならびに国内で先行している社会科学分野の事例と留意点を確認し、今後の方向性を考える一助とするため、各分野の専門家より情報を得た。

三府県コホートデータ約10万例全てについて検討を行う。各個人レコードは226変数からなるが、そのうちIDや数値化前データの変数、他と内容の重複する変数など22変数を除いた204変数を分析対象とした。検討に際し変数をその内容の近いもの同士で組み合わせてカテゴリ化し27のカテゴリを作成した。また、それらのカテゴリを内容から【個人特性】【追跡】【アンケート】の3グループに分けた。

現制度下で二次的に死因情報を利用するため、研究の都度、コホートデータに死因情報を付与する方法について、JALSデータを用い検討した。また、国内外の医療（臨床）データ連携、疫学共同研究などにおける分散型ネットワークの先行事例を収集した。

これまでの研究成果や先行する分野のデータアーカイブセンターのガイドライン等を参考に、特に研究開始時期のデータ取り扱いが現在のものと異なり、かつ多数を取り扱う疫学研究データのアーカイブ化の場合にガイダンスに含めておくべき項目を検討した。

統計法の位置づけと人口動態統計資料の利用価値の観点から、情報関係法制に詳しい法律学の専門家である友岡史仁日本大学法学部教授を交え、検討を行った。

がん登録推進法の規定及び同法に関する政省令を読解し、関連する研究者と議論することにより、追跡終了後コホート研究を用いた共通化データベースが構築された場合における、がん登録情報(が

ん罹患情報と死亡者情報)の利用をめぐる諸問題について考察した。

## C. 研究結果

### -1 米国における二次データ利用の現状

米国では、電子カルテデータ、レセプトデータ、患者レジストリデータのみならず、コホート研究データについても公開が始まっている。これらのデータを用いることで、コホート研究を1から始めるのに比べ少ない労力でデータセットを作ることが可能である。また、NHLBIから研究費を受けた行われたRCTとコホート研究を含めた全てのデータを最大利用することを意図して、Biologic Specimen and Data Repository Information Coordinating Centerを事務局としたデータデポジトリが行われている。データ提供にあたっては、インフォームド・コンセントに沿った形にすること、個人同定情報は削除すること、地域情報は外すことなどがガイドラインで規定されている。このデポジトリでは、データあるいは生体試料二次利用希望者はHP上で、各研究の詳細を確認し、自身の用いたい研究に対しリクエストをすることができるが、その際、倫理審査を受けておくなどの規則も定められている。このシステムにより、すでに他国からの申請も含め多くの研究が実施されている。

### -2 社会学分野のデータアーカイブの現状

社会科学系では、特に若手の研究者がデータアーカイブを利用して、オリジナルな枠組みで分析を行い、新たな知見を出していくことがより一般的になってきている。そのためのセンターの1つであるSSJDA(Social Science Japan Data Archive)には現在、約1600件のデータが寄託されており、2013年度は2700件の利用があった。このように活用が進んでいる背景には、データアーカイブセンターが設立されたこと、ならびに二次分析のメリットが広く研究者に認識されたことがある。このようにデータが収集・公開され第三者が分析する

ことは、データの再現性を確認することにつながる。また、特に公的資金が投入され実施された調査データに関しては、調査者個人のものではないという認識も広まりつつある。

現在、SSJDAの運営費用は文部科学省（2010年度より国立大学法人共同利用・共同研究拠点）東京大学社会科学研究所から運営費、データアーカイブに関わる科学研究費で賄われている。データアーカイブセンターの活動として行われている業務の主なものは、データ寄託の依頼・受付、データ整理、データ秘匿処理、メタデータの作成、データ利用の受付・提供、リモート集計の提供、二次利用成果の公開、データ寄託者の表彰、二次利用促進と適切な解析のための研究会・セミナーの開催等多岐にわたっている。

#### 三府県コホートデータによる一意性の検討

三府県コホート対象者10万人強について204変数を分析対象として検討したところ、変数を1つずつ個別に見た場合の一意性は小さかったが、全変数を組み合わせた場合、一意であるレコード数は対象者の約99.98%であった。一変数で見た場合、一意性には、10年観察終了日のように、分類数が大きいことでそれぞれに振り分けられる個体数が少なくなるため生じる一意と、出産人数において出産人数が20人というように、疫学的にまれな属性の個体が存在したために生じる一意の大きく2パターンが存在した。

さらに三府県コホート対象者約100,000例全てを使用した場合と、無作為抽出により標本数を1万、1,000、100に変化させた場合各々で、分類数 $K$ とユニークセル数 $S_1$ 、分類数に占めるユニークセル数の割合 $S_1/K$ を検討した。全例を使用した場合、分類数 $K$ が増加するとともに、ユニークセル数 $S_1$ およびユニークセル数の割合 $S_1/K$ は増加するが、ユニークセル数の割合 $S_1/K$ については、分類数 $K$ が約20,000例になるまで急増後80%程度でプラトーに達し、分類数 $K$ が80,000例あたりからさらに

増加する、というパターンを示した。この傾向を全例を使用した場合と標本数を減らした場合で比較すると、分布の形状は似通っていたが、急増する部分の勾配が緩やかになり（100例使用の場合は40例程度まで）、プラトー部分が狭くなる傾向があった。

#### -1 死因情報を外したアーカイブデータを利用する運用案

提案するデータ利用基盤の概略は次のとおりである。研究コンソーシアムに参加する各研究が、基本データ（生活習慣、検査データなど）と死因を除いた追跡データをアーカイブセンターに提供する。アーカイブセンターでは、基本データベースと追跡データベースを分けて構築する。その際、基本データベースは原則登録時から修正なしの状態、追跡データベース（その後の死因照合作業で必要となる「死亡地（市町村）」、「死亡日」、「生年月日」、「性別」を含む）は、今後の追跡継続に応じて更新できる構造とする。このデータベースを利用した研究を行いたい研究者は、死因情報を得るために厚生労働省に対し人口動態調査二次利用申請を行う。承認後に提供を受けた死因情報をアーカイブセンター内で、「死亡地（市町村）」、「死亡日」、「生年月日」、「性別」をキー変数として、保有する追跡情報と照合する。死因を付与した一時的な解析データセットを作成し、研究計画に基づいた解析に使用する。研究終了後は死因情報を削除（抹消）し、厚生労働省に利用後報告を行う。

この方法の妥当性を確認するため、JALS対象地域の市町村で1999年1月1日から2012年12月31日までに発生した死亡の調査票情報を厚生労働省に申請、入手した。JALSの対象者で、当該期間中に死亡が特定できていた7,137件（職域コホートと死亡調査データが確定していないコホートを除く）のうち、99.5%が、性別、生年月日、死亡年月日、死亡時の居住市町村名をキー変数として

人口動態調査データと一致した。なお、不一致のうち 14 件は以前に JALS が行った死因照合作業において既に未照合が判明しており、各コホートに対して死亡時情報を確認したがいずれも情報に誤りがなく、人口動態統計作成の過程で入力間違い等が発生した事例と判断した。このため、今回照合出来なかった例は、実質として 19 件 (0.27%) であった。

#### -2 分散型ネットワーク国内外の医療（臨床）データ連携、疫学共同研究利用などの先行事例収集と、疫学追跡終了後コホートデータの共通利用についての検討

近年整備されつつある分散型ネットワークについて、医療（臨床）データ連携、国際疫学共同研究での 5 つの先行事例を収集し検討を行ったところ、医療（臨床）と疫学研究の事例では、データ連携（交換）の目的（必要とする背景）、運用方法などが異なっていることが明らかとなった。

アーカイブデータ提供のためのガイダンス項目含める必要があると考えられたのは、以下の項目であった。

A 対象者の個人情報保護：実データを用いた我々の過去 2 年間の検討では調査項目が多岐に渡り、かつ対象者数が多いコホート研究では、データの一意性はあるものと考えた方がよいと結論した。回答者のプライバシー保護の観点から、調査地域を粗く束ねるなどの手段を用いて、個々の回答者の識別を不可能にすることが望まれる。

B インフォームド・コンセントと倫理審査：新たに開始される研究では、対象者からデータ寄託に関する同意が得られていることを原則とした上で、寄託者の所属機関で倫理審査を受けることが可能である。しかし、既に追跡が終わったコホート研究の場合、その開始は今から 20 年以上前であると考えられ、研究の説明同意プロセスは現在と異なる。そのため、対象者の同意あるいは再同意の取得を前提とすることは困難であり、所属機関での

倫理審査に加え何らかの形で二次利用に関する広報・情報公開を行うことが適切と考えられた。

C 知的財産権の帰属：生体試料を伴わない情報のみのアーカイブでは、知的財産権の発生は稀と思われるが、生体試料を収集・保管するコホート研究では、研究終了時にも試料が残ることが想定される。二次利用者とデータ寄託者の間の分配方法については、一定のルールが定まっていないことから、利用開始時に取り決めを交わしておくことが望ましい。

D 寄託する項目/しない項目の判断：個人同定の可能性が特に高い項目は、丸め処理をするか、提供しないことが望まれる。また、調査方法や設定が特殊である等何らかの理由により、研究内容を十分に理解しない二次利用者の解析・解釈によって誤った結論を引き出すおそれがある項目についても、提供しないことが推奨される。

E データ提供先の制限：二次利用者の資格、利用上の制約等については、各研究関係者の意向を反映するような区分けの周知とそれに沿った対応が求められると考えられた。

#### 統計法解釈に関する検討

コホート研究に不可欠とされる死因情報を人口動態統計資料から得ることは、統計法上の目的外利用を意味するため、その活用に当たり、統計法上設けられた諸手続といった一定の制約が生じている。しかし、データの二次利用により、公費を投入し多くの人々の協力を得て収集された試料・情報の有効利用、若手研究者の育成、データの検証等幅広い研究利用とその成果に直結し、ひいてはより正確な健康関連要因の解明に具体的に貢献しうることを考え合わせれば、統計法の規定による制約を受けない形で、死因も含めたより広い形による該当データの利用可能性が切望される。このことから、当該情報を長期的かつ利用者の範囲を広げ、コホート研究に有効活用する道を法制度について論ずる意義は、十分にあると思われる。

この場合の論点として、A 統計法の解釈運用を変更することで足りるのか、B 統計法とは異なる別立法を行う必要があるのか、という二点が考えられる。A の場合は、上記の制度趣旨と過去の運用の経緯に照らして乗り越えるべき解釈論を構築する必要があると考えられ、その意味で、死因情報以外の件についても考慮に入れた制度全体の観点から議論を要するものと思われる。対して B の場合は、A とは異なり立法的措置により解決するため、特別法として位置付けられ、法解釈上の齟齬はきたさないとされる一方、既存の仕組みとは異なる新たな制度設計を求められるぶん、個人情報保護に照らした保護に係る必要なスタンスを充足する詳細な議論を要する。A または B のいずれを選択するにせよ、新たな立法の煩雑さや既存の仕組みの有効活用という効率性の観点からすれば、現時点では A の選択肢が方向性として現実的と考えられた。

#### がん登録推進法における死亡者情報票の活用

病院・診療所と市町村から届けられた情報をもとに全国がん登録データベースが構築される。同データベースには、がん罹患情報と罹患者における生存死亡に関する情報の 2 つが含まれる。そして、生存死亡の確認の期間は、政令により百年間と定められた。

がん登録情報をコホート研究に活用する際は、匿名化されない情報の提供を受ける必要がある。そのための要件の 1 つが調査対象者の同意であり、「提供の求めを受けた情報に係るがん罹患者が生存している場合、その調査研究を行う者が、当該がん罹患者から調査研究目的で当該がん登録情報が提供されることについて同意を得ていること」とされている。

ただし、がん登録推進法施行日（平成 28 年 1 月 1 日）以前に開始されたがんに係る調査研究については、同法附則 2 条（経過措置）により「同意を得ることが当該がんに係る調査研究の円滑な遂行

に支障を及ぼすものと認められる場合として政令で定める場合に該当するものである場合において、（略）これらの同意に代わる措置として厚生労働大臣が定める指針に従った措置が講じられているときは、（略）適用しない」とされている。

「同意を得ることが当該がんに係る調査研究の円滑な遂行に支障を及ぼすものと認められる場合」については、がん登録等の推進に関する法律施行令（平成 27 年 9 月 9 日政令 323 号）により以下の各号のいずれかに該当する場合と定められた。

- 一 調査研究の対象者が五千人以上の場合
- 二 調査対象者と連絡を取ることが困難であること、または対象者の同意を得ることが調査研究の結果に影響を与えること。

そして「同意に代わる措置として厚生労働大臣が定める指針」が、厚生労働省告示第 471 号（平成 27 年 12 月 15 日）によって定められた。それによると、(1) 「人を対象とする医学系研究に関する倫理指針」第 5 章に即して研究対象者等からインフォームド・コンセントの取得等を実施していること、(2) 全国がん登録情報等の提供を受けることについて情報公開等の措置を講ずること、(3) 研究対象者が提供について拒否できる機会を保障すること、という 3 点が同意代替措置とされた。

#### D. 考察

国民の税金を投入し、多くの人手と長期の追跡を経て構築されたコホート研究データをアーカイブ化し、広く二次利用を可能とすることは、研究の透明性確保、第三者による研究結果の検証、若手の育成に寄与するのみならず、研究の無駄・重複を減らし、必要な公費・労力を新しい有意義な研究に向けるという意義もある。

日本よりデータベースの二次利用が進んでいる米国において、研究に用いられているデータベースの種類、活用事例とその成果、二次利用を進めるにあたっての留意点等について確認した。日本では、二次データの公開システムとしてバイオサ

イエンデータベースセンター（NBDC）が立ち上がったばかりであり、仕組みそのものに関する議論が十分に行われてはいない。今後、日本においても公的研究費を受けた研究を適切に二次利用することが求められているが、そのためには、個人情報保護と研究活用とのバランス、事務手続きの標準化・単純化と必要経費、データの適正使用と質保障のためのサポート、共通化によるデータマニピュレーション、情報のロスとデータ容易使用のバランス、共通化プロセスの透明化と公正なシェアの仕組み等につき、議論を重ねていくことが必要と考えられた。また、国内でも社会科学系分野は先行しており、そのためのセンターが設立され、現在では多くのデータが二次利用されている。しかし、今の形になるまでに、10～20年の年月を要しており、データ寄託がある一定数に達するまで、利用のメリットが十分に浸透するよう働きかけるとともに、利用のための環境整備も必要と考えられた。

前述したようにアーカイブ化により二次利用環境を整える意義は大きいですが、疫学研究では生活習慣や生活環境のみならず、個人の心身面の健康状態情報を収集することから、その取り扱いには慎重であることが望まれる。特に三府県コホートをを用いた一意性の検討からも明らかなように、多数を対象に多項目のデータを長期にわたって収集するコホート研究では、一意性があるものとの前提で対応を考える必要がある。今回、特に追跡の終わったコホート研究の利活用を進めることを念頭に、データ提供の際に従うべきガイダンスに含める必要がある項目を5点に整理した。今後、追跡が終了したコホート研究データの寄託を進める際の指標になるものと考えられる。

多くのコホート研究では、死因情報を人口動態統計資料から得ており、統計法の規定上、現状では非常に重要な情報であるにもかかわらずアーカイブデータに死因情報を付加して二次利用に供することはできない。そこで、現制度下での運用方

法につき、2つの提案を行った。提案1の運用案を複数のコホート研究から構成される国内循環器疫学研究（JALS）に適用したところ、死因照合率は99.7%であり、照合作業の技術的側面、作業手順化の面で問題はなく、運用案の一つになりうると考えられた。提案2については、医療（臨床）、疫学研究で先行している分散型ネットワークの事例を収集し、本検討事例である疫学追跡終了後コホートデータの共通利用の場面に応用可能な事例があるか検討を行った。疫学共同研究における分散型ネットワークを用いたデータ連携の例などから、リモート環境で様々な統計解析を柔軟に行えること、ローカルサイト（大学、研究所、医療機関を想定）へアクセスを行う上でのセキュリティ、さらには、人口動態統計の二次利用をローカルサイト単位で各々申請し、死因を付与したデータを統合して用いることや、施設外部のPC上のRAMへの一時的な書き出し（保存はされない、電子データの「一時的蓄積」）の可否など、法令解釈の問題にも留意する必要があると考えられた。疫学研究追跡終了後のコホートデータの共通利用に関して、提案した二つの方法で運用していくことは可能と考えられるが、実際に運用する場合には、前者においてはデータ利用の規約、データ利用の物理的環境、サポート体制、後者は、人口動態二次利用申請による死因データ利用の法令解釈などの課題があると考えられた。

一方でコホート研究にとって死因情報は非常に重要なものであり、利用できないことによりアーカイブデータの二次利用の価値は低下すると考えられる。また、米国ではNational Death Index（NDI）という、厚生省（U.S. Department of Health and Human Services）の下部機関が、研究目的での生存・死亡確認情報（死亡時には、死亡年月日や死因などを含む）の提供を行っている。研究者は、調査対象者リスト（氏名、性、生年月日、住所、社会保障番号など）を提出し、審査にパスすると、有料（基本料350ドル+対象者1人1年あたり15セント）で、上記情報が提供される。これにより、米国の疫学研究・臨床研究のレベルと即時性は飛躍的に向上し、医学研究や医薬品開発に

において国際的に有利な地位を確保することができたといえ、今のままでは日本の疫学研究は後塵を拝する。そこで、法の解釈、または立法により、コホート研究において重要な追跡情報となる死因を公的調査情報(人口動態統計資料)から得た上で二次的利用を可能にできないかを検討した。しかし、これには統計法の従来の解釈経緯などといった従前のコホート研究とは異なる方向性が必要とされており、今後も継続して、法学分野からの知見に照らし、より視野を広げた説得力ある議論を展開する必要があると考えられた。また、がん登録推進法の施行により、追跡情報としてがん罹患情報の利用も広がるものと考えられる。しかし、がん登録推進法と同法に関する政省令の規定において定められた要件・方法などに従って、がん登録情報の研究利用の承認を受けた者が、データアーカイブにより情報を第三者に研究目的で提供することの可否については、それらの規定の中では一切言及されていない。この問題については、コホート研究で人口動態統計資料から得た死因情報をデータアーカイブに付加して提供するための法解釈と同様の検討が必要になると思われる。

## E. 結論

疫学研究により得られたデータを広く共有化するためのシステムであるデータアーカイブ化に向けた検討を行った。統計法の規定上、死因情報を付加してのデータ公開・二次利用は認められていないことなどを踏まえて、現制度下での運用方法を2つ提案した。一つ目はアーカイブセンターなどに死因情報以外のデータを集約管理し、必要に応じて従来通りの死因照合作業を実施し、死因を付加したデータセットを作成し解析を行う方法である。二つ目は、データを保持者のもとに置いたまま、必要と判断した情報だけを選択的に共有させる分散型ネットワークによる方法である。また、実データを用いて、一意性に関する検討を行い、10万人規模のコホート集団はもちろん、100例規

模のコホート集団でも分類数が全対象者数の40%程度で、ユニークセルの割合が80%に達していたことから、コホートの規模にかかわらず、80%程度のレコードは一意性があるものとして対応する必要があることを示した。これら研究成果や他国・他領域での実例を参考に、データアーカイブの活用を進めるため、データ提供の際に従うべきガイダンスに含める必要がある項目をA対象者の個人情報保護、Bインフォームド・コンセントと倫理審査、C知的財産権の帰属、D寄託する項目/しない項目の判断、Eデータ提供先の制限、の5点に整理した。

一方で、エンドポイントとしての死因情報は非常に価値が高いものであることから、追跡が終了したコホート研究の二次利用を進めるために、死因情報のソースとなる人口動態統計調査の有効活用の方策を模索することが望まれ、その際の論点を2つ提示した。統計法の従来の解釈経緯などといったコホート研究とは異なる議論の方向性が必要と考えられ、医学分野の知見だけではない法学分野からの知見に照らし、今後、より視野を広げた説得力ある議論の展開が不可欠である。なお、がん登録から得たがん罹患情報の研究目的での二次利用についても、今後同様の検討が必要と考えられた。

## F. 健康危機情報

なし

## G. 研究発表

1. 論文発表  
なし
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

なし