

201315042A

厚生労働科学研究費補助金

循環器疾患・糖尿病等生活習慣病対策総合研究事業

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

平成 25 年度 総括・分担研究報告書

平成 26 (2014) 年 3 月

研究代表者 玉腰 暁子

目 次

I.	総括研究報告	
	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：進捗報告.....	1
	玉腰暁子	
II.	分担研究報告	
	米国におけるデータアーカイブの研究利用：現状と課題.....	5
	大橋靖雄、祖父江友孝、他	
	統計行政におけるデータ利用の動向.....	8
	大橋靖雄、祖父江友孝、他	
	データアーカイブ利用に際して必要と考えられる研究倫理教育.....	11
	辻一郎、磯博康、他	
	ヒトに由来するデータ活用と知的財産・法的課題.....	14
	磯博康、辻一郎、他	
	大規模コホートデータにおける一意性の検討.....	17
	祖父江友孝	
	Japan Arteriosclerosis Longitudinal Study (JALS)データにおける一意性の検討.....	39
	大橋靖雄、原田亜紀子、他	
III.	資料.....	53

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：進捗報告

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

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。そのため、初年度は、米国のデータ利用環境、統計法に基づき実施された国内統計資料の利用のあり方、データ二次利用にまつわる倫理的課題と知財、研究倫理教育のあり方について、現状を確認し、課題を整理するとともに、各コホートが持つデータを公開した場合の一意性の問題について、実データを基に検討した。その結果、統計法の規定から、人口動態統計資料から得られた死因情報をコホート研究のアウトカムとして公開することはできないことが明らかとなった。そのため、コホート研究データをアーカイブ化し二次利用を進めるためには、死因情報のソースとなる人口動態統計調査の統計法内における位置づけを変更することが必要と考えられた。また、二次利用に際しては、対象者の同意の範囲への考慮とそれに伴う匿名化情報としての取り扱いはもちろん、二次利用者の責務を意識することが重要である。さらに、特に試料の利用にあたっては、知財に関し、関係者で事前に十分に検討し、ルールを明確にしておくことが必要と考えられた。なお、多くの項目を組み合わせれば一意性があるものとの前提で、二次利用対応を考えていくことが必要である。

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（東京大学大学院医学系研究科・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

本研究では、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。初年度は、米国のデータ利用環境、統計法に基づき実施された国内統計資料の利用のあり方、データ二次利用にまつわる倫理的課題と知財、研究倫理教育のあり方について、現状を確認し、課題を整理した。また、各コホートが持つデータを公

開した場合の一意性の問題について、実データを基に検討した。

B. 研究方法

1. 各分野の専門家から現状を伺い、追跡を終了したコホート研究データを二次利用する際の課題を整理した。
2. 三府県コホートデータと JALS 死亡者データを用い、それぞれどのような頻度で一意性が見られるかを検討した。

C. 研究結果

1. 資料・データベースの二次利用の現状と課題
 - 1-1. 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、個人同意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。

1-2. 日本における公的統計の利用

公的統計は、その体系的かつ効率的な整備及びその有用性の確保を図ることを目的とする統計法で規定されている。2007年の全部改正にあたっては、社会の情報基盤としての統計というコンセプト、そしてそれを実践するための統計データの利用促進、特に二次利用に積極的に取り組むことが盛り込まれた。そのため現在では、オーダーメイド集計や匿名データの利用が推奨されている。なお、ほぼ全てのコホート研究で死因情報は、人口動態統計調査の利用手続きを経て入手されていることから、この情報を更に別の二次利用（提供）に供することは現行の統計法の下では不可能である。

1-3. 試料・情報の二次利用に関する説明と同意

新しく開始される研究では、二次利用まで見越した同意を取得することが望ましい。当初の研究会市時点では具体的な二次利用の詳細は決まっていなことも多いが、広範囲同意で二次利用に関する同意も得、実際に二次利用を進める際には倫理審査委員会など第三者の意見も聞いたうえで実施にあたってはオプト・アウト方式(研究に関する情報を公開し、拒否権を保障する)を採用することが推奨され

る。それに対し、既に進んでいるコホート研究では、そのような二次利用を想定していないことが多く、一次同意の範囲では二次利用をカバーできていない。そのような試料・情報であっても有効に研究に活用するためには、対象地域や対象者に対する情報公開を多面的かつ十分に行ったうえで拒否権を保障することが重要であると考えられる。また、二次利用者も研究にあたっては情報公開をしっかりと行うとともに、試料・情報提供者に対し、研究成果に関する報告や有用な個別結果の返却など、できるだけ利益を還元するよう務めることも必要であろう。

1-4. 試料・情報二次利用に必要な研究倫理教育

二次利用に供される試料・情報の特徴を考えると、データアーカイブを用いて研究を行う二次利用者に対しては、以下の点を目指した研究倫理教育が必要と考えられた。

- 一次収集者の苦勞を知る
- 試料・情報の背後には一人ひとりの生身の人間がいることを知る
- ヒト試料研究一般に求められる倫理性を理解する
 - ヒト試料の地位
 - 同意取得の原則
 - 同意原則の緩和が許容されるための条件と発生責務
 - 個人情報の安全管理措置
- 各ステークホルダーに課せられる倫理的責務を知る
- 知財・オーナーシップ等についての適切な知識を持つ
- データ不正等の罪深さを知る
- 情報倫理について理解する

1-5. 試料・情報二次利用に関連する知財

ヒトに由来する試料は、細胞や血液といった形でそのヒトの中にあるときは、当然そのヒトに帰属する。しかし、研究利用に関する説明が行われ、同意を得た上で採取された細胞や血液は、インフォームド・コンセントで示された条件で使用するという約束の下でドナーが譲渡したものと考えることができる。また、その試料・情報を用いて研究が行われ

た場合、研究によって試料・情報に付加価値がつく。したがって、いったん研究に提供された試料・情報を用いて行われた研究の成果として知的財産権が発生した場合、その権利は試料・情報提供者ではなく、研究者(研究機関)に帰属すると考えるのが適切である(対象者への説明時点でもちろん述べておくべき内容である)。バンクや共同研究の形で提供される試料・情報を用いた二次利用で開発された研究成果に関する特許権の分配方法については、事前に取り決めを交わしておくのが望ましい。一次情報収集者の努力、バンク運用者の作業を考えると、そこにどう報いるかを二次利用者は常に意識することも重要であろう。このことは特許権の配分のみならず、論文の共著者に加えるかどうか、謝辞に述べるかどうかも含めて求められる。

2. 一意性の検討

2-1. 三府県コホート対象者に関する検討

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。

2-2. JALS 対象者中の死亡者に関する検討

複数のコホートが共同で実施しているJALSのデータを用いて、死亡者約5000人の性、ベースライン時年齢、死亡時年齢、地域、死亡年データによる個人特定可能性につき検討した。性、ベースライン時年齢単独での特定性は低かったが、死亡時年齢が40歳代もしくは90歳代後半である場合や小規模コホートの対象者で特定可能性が高くなった。さらに死亡年月日を月までで丸めた場合でも個人が一意に特定される場合が観察された。また、これら変数を2つ以上組み合わせた場合には、一意に特定されるものの数が上昇した。

D. 考察

1. 試料・情報二次利用の現状と課題

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ス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。

国内の統計行政に関しては、統計法で規定されている。改正統計法は、社会基盤としての統計をコンセプトに、公的統計の体系的かつ効率的な整備及びその有用性の確保を図ることが目的とされている。公的統計の利用に関しても範囲が拡大され、秘匿しない調査票情報の提供、秘匿処理した調査票情報の利用(匿名データの利用)、行政機関に特別集計を依頼(オーダーメイド集計)が可能である。なお、提供された調査票情報や匿名データなどを統計の作成に関連する目的以外に利用・提供した者や、守秘義務規定に違反した者に対して、罰則が定められている。したがって、コホート研究データをアーカイブ化し二次利用を進めるためには、アウトカムに用いている死因情報のソースとなる人口動態統計調査の統計法内における位置づけを変更することが必要と考えられた。

研究倫理の観点からは、試料・情報の二次利用に際しては、対象者の同意の範囲への考慮とそれに伴う匿名化情報としての取り扱いはもちろん、二次利用者の責務を意識することが重要である。したがって、アーカイブやバンクによるデータ二次利用を行う研究者に対する研究倫理教育では、通常の研究倫理

修内容に加え、二次利用に関する同意のあり方、オープンシップや知財に関する内容を含めることが必要と考えられた。また、アーカイブやバンクによるデータ二次利用を進めるためには、その後が生じ得る知的財産検討に関し、事前に関係者間で十分に検討することが必要と考えられた。

2. 一意性の検討

一変数で見た場合、一意性には、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、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パターンが考えられた。前者に対しては例えば日付データを月までにするなどにより分類数を減らすことで一意性を減少させることが可能であり、後者に対しては一定値以上（以下）については直接表示せず、無限までの片側区間で表示するといった方法により一意性の減少が図られる。しかし、すべての変数を組み合わせた場合には全レコード数の約99.98%が一意であったが、変数が多くなる（質問項目が多い）ことによる一意性は容易に避けられるものではない。また、10万人規模のデータではまるめ処理などによりある程度の一意性の減少がみられるが、規模が小さくなると一意性が上がる可能性も高い。したがって、多くの項目の組み合わせを行う場合には、一意性があるものとの前提で対応を考える必要がある。

また、複数の比較的小規模なコホートをあわせて実施されているコホート研究における死亡者データで検討したところ、地域の情報が特に個人の特定に繋がりやすいこと、死亡時年齢が若年もしくは高齢である場合に一意性が高じることが明らかとなった。また、死亡年月日については、月までに丸めた場合でも個人が一意に特定される場合があることから、可能な限り死亡年までの情報とすることが求められる。さらに死亡者につき用いる情報（性、ベースライン時年齢、死亡時年齢、地域、死亡年）を一変数ではなく二変数以上とした場合には、組み合わせによって一意性が上昇した。したがって、分類を粗くするなどの対応が必要と考えられた。

E. 結論

米国のデータ利用環境、統計法に基づき実施された国内統計資料の利用のあり方、データ二次利用にまつわる倫理的課題と知財、研究倫理教育のあり方について、現状を確認し、課題を整理するとともに、各コホートが持つデータを公開した場合の一意性の問題について、実データを基に検討した。その結果、統計法の規定から、人口動態統計資料から得られた死因情報をコホート研究のアウトカムとして公開することはできないことが明らかとなった。そのため、コホート研究データをアーカイブ化し二次利用を進めるためには、死因情報のソースとなる人口動態統計調査の統計法内における位置づけを変更することが必要と考えられた。また、二次利用に際しては、対象者の同意の範囲への考慮とそれに伴う匿名化情報としての取り扱いはもちろん、二次利用者の責務を意識することが重要である。さらに、特に試料の利用にあたっては、知財に関し、関係者で事前に十分に検討し、ルールを明確にしておくことが必要と考えられた。なお、多くの項目を組み合わせれば一意性があるものとの前提で、二次利用対応を考えていくことが必要である。

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

分担報告書

米国におけるデータアーカイブの研究利用:現状と課題

研究分担者	大橋靖雄	東京大学大学院医学系研究科
研究分担者	祖父江友孝	大阪大学大学院医学系研究科
研究代表者	玉腰暁子	北海道大学大学院医学研究科
研究協力者	瀬戸口聡子	Duke 大学医学部

研究要旨

米国で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。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。

A. 目的

国内ではまだ十分に活用が進んでいないデータベースの研究利用に関し、米国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。

B. 方法

実際に米国でデータベースを用いた研究に従事されている Duke 大学瀬戸口聡子准教授より、情報を得た。

C. 結果

[米国で比較的高頻度に研究活用されているデータベース]
電子カルテデータ: 病院毎に異なるシステムを導入しており、それぞれ研究用 DB 化を試みているが、統一した全国的なデータベースは存在しない

- EPIC—近年、このシステムを導入する病院が増加
クレームデータ(レセプトデータ): 保険システム毎に存在
- Medicaid—貧困層
- Medicare—高齢者・一部の疾患
患者レジストリデータベース: 主に学会が作成登録
- National Cardiovascular Data Registry—American College of Cardiology、主に手技(ステント、PCI など)毎に患者を登録
- STS National Database—Society of Thoracic Surgeons、胸部外科の各領域(先天性心臓外科、成人心臓病手術、呼吸器手術)手術の登録
- SEER (Surveillance Epidemiology and End Results) Program—National Cancer Act に基づく州単位のがん登録のうち精度のよいもの

コホート研究データベース: 二次利用のために公開されているものもあるが、まだそれほど広がっていない。現在、National Heart, Lung, and Blood Institute (NHLBI) が研究費を受けて作ったデータに関しては 3 年経過後には提出するよう求めている

- ARIC
- Framingham Study
- Women's Health Initiative

[データベースの結合利用事例]

米国では、データベース同士の結合も進められており、例えば ARIC、Framingham Study は Medicare Data とリンクされている。

また、データベース内にある ID、あるいは生年月日や病院 ID、入院日を利用して、患者や医師情報を結合することが可能であり、複数のデータベースを用いた研究を行うことができる。治療法による予後を比較、(埋め込み型除細動器と頸動脈ステント術) する研究では、患者レジストリデータベース (Society for Vascular Surgery's Vascular Registry (SVS-VR) と Carotid Artery Revascularization and Endarterectomy (CARE)) に Medicare Data、さらに有料の病院情報、医師情報 (American Hospital Association) などを結合した。このような方法により、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が、データ入手の事務作業等で手元にデータが来るまでに 2 年かかることもある。なお、患者個人の ID を用いなくても、生年月日、医療手技、入院日とプロバイダー ID を利用すれば、95% 以上が正しく突合された。また、米国では保険に加入する際サインする同意書に、データの研究利用の項目があるため、保険者がデータを研究利用するために改めて同意を得ることはされず、また ID を削除したデータが販売も行われているのが現状である(ただし、州により対応が異なる場合がある)。

[データ二次利用システム事例]

<https://biolincc.nhlbi.nih.gov/home/>

米国では、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。

データ提供のタイミングは、RCT であれば最後の患者登録から 3 年後、または主目的の論文が公表されてから 2 年後、観察型の疫学研究では追跡終了の 3 年後、または付随的な研究から 2 年後である。その場合、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。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。

2014 年 2 月時点で確認したところ、134 の研究(臨床研究 91 件、疫学研究 42 件、両者 1 件)が登録されていた。なお、2000~10 年(2010 年時点では登録は 76 研究)の間に、498 のリクエストが認められ、うち 74 件(15%)が他国の研究所からであり、224 論文が公表されたとのことである。

[データベース標準化の方向性]

米国では、各所に散らばるデータを 1 箇所に集めるのではなく、それぞれの場所に置いたまま共同研究ができるようデータフォーマットを統一する distributed network model という試みが始まっている。例えば、OMOP (Observational Medical Outcomes Partnership) は政府と民間の共同によるクレームデータや電子カルテをつなぐデータベースネットワーク構築とリサーチ手法の研究の試みである。その際、項目名、変数の与え方等も全て統一化する(コモンデータモデル)ことにより、複数のデータベースで同時に同じ解析を行うことを目的として、現在米国内の 10 データソース、1 億 3000 万人をカバーする規模で進められている。また、FDA の mini-sentinel は特に薬の市販後調査を年頭に、同様の試みをしている。ここでは、

18の保険会社データベースを用いており、1億3000万人がカバーされている。

D. 考察

米国では、NHLBIから研究費を受けて実施された臨床研究、疫学研究は研究終了後一定期間が経過した後は、公開される仕組みが整えられている。しかし、日本ではまだその議論が始まったばかりであり、バイオサイエンスデータベースセンター(NBDC)が立ち上がってはいるものの、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。また既存データベース同士を個人単位で連結することで、データ価値が倍増することも考えられるが、そのための仕組みは整っていない。

今後、観察が終了したコホート研究の二次利用を適切に進めるためには、以下の点に関し、議論を重ね、体制を構築していくことが重要と考えられる。

- 個人情報保護と研究活用のためのバランス
- 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使用簡易性のバランス
- 共通化プロセスの透明化と公正なシェアの仕組み

E. 結論

米国でのデータベース活用とその成果、留意点等について確認した。それらを参考に、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、そのために検討すべき課題を列記した。

F. 研究発表

1. 論文発表
 2. 学会発表
- いずれもなし

G. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得
 2. 実用新案登録
 3. その他
- いずれもなし

統計行政におけるデータ利用の動向

研究分担者	大橋靖雄	東京大学大学院医学系研究科
研究分担者	祖父江友孝	大阪大学大学院医学系研究科
研究代表者	玉腰暁子	北海道大学大学院医学研究科

研究要旨

国内の統計行政に関し、特に改正統計法を中心に、情報を得た。改正統計法では、社会基盤としての統計をコンセプトに、公的統計の体系的かつ効率的な整備及びその有用性の確保を図ることが目的とされている。公的統計の利用に関しても範囲が拡大され、秘匿しない調査票情報の提供、秘匿処理した調査票情報の利用(匿名データの利用)、行政機関に特別集計を依頼(オーダーメイド集計)が可能である。なお、提供された調査票情報や匿名データなどを統計の作成に関連する目的以外に利用・提供した者や、守秘義務規定に違反した者に対しては、罰則が定められている。したがって、コホート研究データをアーカイブ化し利用を進めるためには、アウトカムに用いている死因情報のソースとなる人口動態統計調査の統計法内における位置づけを変更することが必要と考えられた。

A. 目的

追跡が終了したコホート研究をアーカイブ化することを念頭に、現在の統計行政におけるデータ利用の動向について確認する。

有している。政府統計は約 500 あり、実際に毎年実施されている統計は 300 弱である。独立行政法人統計センターは、各府省が実施する重要な統計を作成、あるいは必要な統計技術の研究等を行う役割を担っている。

B. 方法

総務省統計局で統計情報作成に携わっている上田聖氏より、情報を得た。

[匿名データの作成と利用]

元来、統計は行政のため、という位置づけであった。そのため、2007 年の改正統計法が公布されるまでの統計法では、第 15 条「何人も、指定統計を作成するために集められた調査票を、統計上の目的以外に使用してはならない」とされ、データ利用を強く制限する方針であった。第 2 項に「前項の規定は、総務大臣の承認を得て使用の目的を公示したものについては、これを適用しない」との例外規定が置かれていた

C. 結果

[統計行政]

日本では、各府省がそれぞれ必要な統計を作成する分散型のシステムを採用している。その中で総務省は、国勢調査や経済センサスという全数調査等を所管しているほか、各府省の調査に対する許認可権を

ものの、統計利用の運用上は第1項が中心となっていた。

1980年代に入ると、欧米では個人情報秘匿した匿名データを作成するようになり、その利用が活発になる。さらにパブリックユースファイルといわれるネット上でダウンロードできる仕組みも作られるようになってきた。それに対し、日本は調査票利用のハードルを上げてきたが、十分な対応はできなかった。そのため、いくつかの調査について、匿名データを作成し、提供することを開始した。

[改正統計法]

統計法は、公的統計の体系的かつ効率的な整備及びその有用性の確保を図ることを目的としている。2007年の全部改正にあたっては、社会の情報基盤としての統計というコンセプト、そしてそれを実践するための統計データの利用促進、特に二次利用に積極的に取り組むことが盛り込まれた。

[統計データの利用促進]

改正統計法では、従前の運用を引き継ぎ以下の1.が規定されるとともに、調査票情報の利用範囲が以下の2.及び3.として拡大された。

1. 秘匿しない調査票情報の提供(第33条):①行政機関、地方公共団体、独立行政法人、②①と共同研究を行う者、もしくは委託研究を行う者、③公的な競争研究資金を受けている者(省令第9条)
2. 秘匿処理した調査票情報の利用(匿名データの利用)(手数料あり)(第35条・36条):①学術研究を行う者、②高等教育を行う者
3. 行政機関に特別集計を依頼(オーダーメイド集計)(手数料あり)(第34条):①学術研究を行う者、②高等教育を行う者

提供された調査票情報や匿名データなどを統計の作成に関連する目的以外に利用・提供した者や、守秘義務規定に違反した者に対して、罰則が定められている(第42条・43条、罰則規定は第7章)。また現

在までに、6つの匿名データ化された統計調査(例:全国消費実態調査、住宅・土地統計調査等)が利用可能となっており、積極的な研究利用が推奨されている。なお、平成24年7月の資料(調査票情報の二次的利用の状況について)によれば、平成21~23年度中のオーダーメイド集計の利用件数は、10調査26件である。

これらとは別に、調査票情報を高次元クロス集計した統計(擬似マイクロデータ、1セル内の客対数は3以上)が作成され、提供されている。これは、演習用として多数ダウンロードされているほか、オーダーメイド集計を依頼する前段階のデータチェック用、SAS社のデータ分析コンテストの対象データとしても用いられている。

[次期統計行政の動向]

現在、平成26~30年度の計画の策定作業中である。その中で、調査票情報の提供に関して、セキュリティの観点からオンサイト利用やプログラム送付型集計・分析に段階的に移行することが検討されている。また、オーダーメイド集計については、利用条件が緩和される方向である。

[統計データの利用促進]

ビッグデータ時代に入っていることもあり、統計センターでは、統計API(application programming interface)、統計GIS、データキューブ、モバイル空間統計に関する取り組みが始められている。前2者は、世界最先端IT国家創造宣言(平成25年6月14日閣議決定)で取り上げられたものである。

D. 考察

統計行政の歴史的背景と現状、今後の見通しについて、概観した。社会基盤としての統計というコンセプトから統計法が全部改正され、さらにビッグデータの時代に入っていることから、調査票情報の利用範囲は広がりつつある。

一方で、観察が終了したコホート研究データには、

重要なアウトカムとして、死亡に関する情報が含まれている。公的研究費により実施されるライフサイエンス分野の研究では、現在、論文発表等で公表された成果に関わる生データの複製物、又は構築した公開用データベースの複製物を、バイオサイエンスデータベースセンターに提供することが求められている。また、多額の費用と長期の追跡、そして多くの人々の協力の下で蓄積されてきたコホート研究データをアーカイブ化し、追跡調査/研究終了後も適切に利用することは重要である。しかし、ほぼ全てのコホート研究では、特に死因情報は人口動態統計調査の利用手続きを経て入手されていることから、この情報を更に別の二次利用(提供)に供することは現行の統計法の下では不可能である。したがって、追跡が終了したコホート研究のデータアーカイブ化を進めるためには、戸籍法に基づき全国民対象に行われている死亡、出生等の届けから作成される人口動態統計調査の統計法制上の位置づけを変更しない限り対応できないことが判明した。

E. 結論

統計行政に関し、特に改正統計法を中心に、情報を得た。コホート研究データをアーカイブ化し利用を進めるためには、アウトカムに用いている死因情報のソースとなる人口動態統計調査の統計法制上の位置づけを変更することが必要と考えられた。

F. 研究発表

1. 論文発表
 2. 学会発表
- いずれもなし

G. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得
 2. 実用新案登録
 3. その他
- いずれもなし

データアーカイブ利用に際して必要と考えられる研究倫理教育

研究分担者	辻一郎	東北大学大学院医学系研究科
研究分担者	磯博康	大阪大学大学院医学系研究科
研究分担者	大橋靖雄	東京大学大学院医学系研究科
研究代表者	玉腰暁子	北海道大学大学院医学研究科
研究協力者	松井健志	国立循環器病研究センター

研究要旨

データ二次利用に際しての研究倫理の課題、ならびに研究倫理教育の現状に関し、情報を得た。試料・情報の二次利用に際しては、対象者の同意の範囲への考慮とそれに伴う匿名化情報としての取り扱いはもちろん、二次利用者の責務を意識することが重要である。したがって、アーカイブやバンクによるデータ二次利用を行う研究者に対する研究倫理教育では、通常教育研修内容に加え、二次利用に関する同意のあり方、オーナーシップや知財に関する内容を含めることが必要と考えられた。

A. 目的

疫学研究で収集された個人単位のデータをアーカイブ化し、研究者の二次利用に供する場合に必要と考えられる研究倫理教育について検討し、今後の体制構築の一助とする。

B. 方法

研究倫理の第一人者であり、その教育にも通じている松井健志氏より情報を得た。

C. 結果

[データ二次利用に伴う倫理課題]

■同意の範囲と匿名化

疫学研究で収集される試料・情報は対象者個人から生じる。研究参加にあたって対象者は、一次

利用者（データ収集者）と同意文書を交わし、通常は、この一次同意の内容を元に研究が実施される。そして、アーカイブ化やバンクはこれら一次的に集められた情報の二次利用と位置づけることができる。その際、当初から二次利用まで見越した同意を取得でいていけばよいが、多くは具体的な研究の詳細は決まっておらず、そこまでの対応は行われていない。

プライバシーを保護する方法として、同意と匿名化を挙げることができる。上述したように多くの場合、二次利用に関する同意まで一次利用時に得ておくことは困難である。そこで、EU 指令 95/46 号「個人データ処理にかかわる個人の保護及び当該データの自由な移動に関する欧州議会及び理事会の指令」では、個人情報守秘について

適切な安全管理措置をしていること、本人の同意を得ることが合理的に実現困難で、正当かつ十分な公共の利益の見込みがある、という3条件全てを満たす場合は、提供者の同意を得ることなく情報にアクセスすることが許容された。その際にはリスクの低減が求められることから、プライバシー侵害が起こりにくいよう、連結可能匿名化であれば対応表を第三者管理にする、連結不可能匿名化する、といった対応が生じてきた。しかし、このような匿名化による対応が進むと、そもそも提供者の同意そのものの存在意義が揺らぐ、という大きな問題が生じてくる。

そこで、第三の方策として、当初の同意から、二次利用に関して、広範同意を得ておくことが考えられる。広範同意とは、例えば「がん・脳卒中などの生活習慣病に関する医学研究に使ってよい」など、不確定ではあるもの用途・範囲が幾分限定されイメージしやすくなっているものである（これに対し、「将来の医学研究に使ってよい」など研究の範囲がまったく特定されていないものは包括同意と呼ばれている）。

したがって、新しく開始されるコホート研究では、当初より広範同意で二次利用に関する同意も得、実際に二次利用を進める際には倫理審査委員会など第三者の意見も聞いたうえで、実施にあたってはオプト・アウト方式（研究に関する情報を公開し、拒否権を保障する）を採用することが推奨される。それに対し、既に進んでいるコホート研究では、当初二次利用を想定していないことが多く、一次同意では二次利用をカバーできていない。そのような試料・情報であっても有効に活用するためには、対象地域や対象者に対する情報公開を多面的かつ十分に行ったうえで拒否権を保障することが重要であると考えられる。この際、

「同意」の意義を考えると、可能な限り、当初包括同意であっても、将来的な利用に関し何らかの同意が得られている対象者に限定するなどの対応が求められる。また、二次利用者も情報公開をし

っかり行うとともに、試料・情報提供者に対し、研究成果に関する報告や有用な個別結果の返却など、できるだけ利益を還元するよう務めることも必要であろう。

■二次利用者の責務

二次利用の際は、一次利用と異なり、研究者と対象者との関係性が必然的に薄くなる。したがって、利用している試料・情報は個人に由来する尊厳あるものであることに対し、敬意を払う気持ちを常に持つよう心がけることが重要である。また、一次利用者あるいはバンク等に対しては、得られた研究成果をフィードバックする、出所を明確化する、場合によっては論文著者に加えるなどの配慮も求められる。

[データアーカイブ利用に際して必要な研究倫理教育]

そこで、データアーカイブを用いて研究を行う二次利用者に対しては、以下の点を目指した研究倫理教育が必要と考えられる。

- 一次収集者の苦勞を知る
- 試料・情報の背後には一人ひとりの生身の人間がいることを知る
- ヒト試料研究一般に求められる倫理性を理解する
 - ヒト試料の地位
 - 同意取得の原則
 - 同意原則の緩和が許容されるための条件と発生責務
 - 個人情報の安全管理措置
- 各ステークホルダーに課せられる倫理的責務を知る
- 知財・オーサーシップ等についての適切な知識を持つ
- データ不正等の罪深さを知る
- 情報倫理について理解する

今までのところ、研究倫理教育は主に臨床試験を対象として行われてきている。E-Learning 教

必要となる研究倫理教育カリキュラムは？

- 一次収集サイトでのリクルート実地研修
- 講義・演習
 - 研究倫理原論
 - 研究不正（データ不正等）
 - 研究倫理の原則
 - ヒト資料研究・バイオバンクの倫理（概論）
 - ヒト資料研究における同意（各論）
 - ヒト資料研究と参加者への利益還元（各論）
 - アーカイブ化資料の二次利用に係る倫理（各論）
 - ヒト資料研究に係る規制・法令
 - 医学研究における個人情報保護（各論）
 - Authorship・知財とその配分：制度・契約等
 - 情報倫理（cf. 「倫理倫理の規範(FNE)」プロジェクト1999-2003; 高木、平塚
<http://www.fnc.f.u-tokyo.ac.jp/> <http://www.fnc.or.jp/~fncproject/1999-2003/>)

材では、国外では、NIH e-learning、OHRP e-learning for Assurance、FHI e-learning、国内では、ICR 臨床研究入門、CITI Japan、CRT-webなどをあげることができる。また、研修・セミナー型のももいくつか行われている。しかし、機会はそれほど多くないのが現状である。それらのカリキュラム内容を参考に考えると、今後、データアーカイブ化を進めていくにあたり、二次利用者に対し新たに研究倫理教育を行う場合には、図に示したような内容を盛り込むことが有用と考えられる。

D. 考察

疫学研究で得られたデータをアーカイブ化し二次利用体制を整備する場合に、二次利用者に必要な研究倫理教育内容につき、その課題を踏まえ検討した。

一次利用と異なり、二次利用に関しては研究の全体像を示しての同意は得られていないことが多い。そのような試料・情報であっても、時間と労力を費やし集められたデータであることから、有効に活用することが望まれるが、どのような対応方法が適切かについては、十分な議論が必要である。また、二次利用者は、データ提供者、一次利用者等に対し敬意を払って、研究を実施することが求められる。したがって、研究倫理教育内容には、通常研究者に求められるものに加え、データアーカイブやバンクに特徴的な同意、オーサiershipや知財に関する内容を含めることが必要と考

えられた。

E. 結論

データ二次利用に際しての研究倫理の課題、ならびに研究倫理教育の現状に関し、情報を得た。アーカイブやバンクによるデータ二次利用を行う研究者に対する研究倫理教育では、通常の研究教育研修内容に加え、二次利用に関する同意のあり方、オーサiershipや知財に関する内容を含めることが必要と考えられた。

F. 研究発表

1. 論文発表
 2. 学会発表
- いずれもなし

G. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得
 2. 実用新案登録
 3. その他
- いずれもなし

平成 25 年度厚生労働科学研究費補助金(循環器疾患・糖尿病等生活習慣病対策総合研究事業)
「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」
分担研究報告書

ヒトに由来するデータ活用と知的財産・法的課題

研究分担者	磯博康	大阪大学大学院医学系研究科
研究分担者	辻一郎	東北大学大学院医学系研究科
研究分担者	大橋靖雄	東京大学大学院医学系研究科
研究代表者	玉腰暁子	北海道大学大学院医学研究科
研究協力者	隅藏康一	政策研究大学院大学

研究要旨

試料・情報二次利用に際しての知的財産権・法的な課題、ならびに遺伝子特許の現状に関し、情報を得た。アーカイブやバンクによるデータ二次利用を進めるためには、その後に生じ得る知的財産検討に関し、事前に検討しておくことが必要と考えられた。

A. 目的

疫学研究で収集された試料・情報を二次的に利用する場合の知的財産権・法的課題について整理し、今後の体制構築の一助とする。

B. 方法

生命科学分野の知的財産権等に関する調査研究を実施されている政策研究大学院大学の隅藏康一氏より情報を得た。

C. 結果

[知的財産権(特許権)の意義]

研究成果を実用化し社会に貢献することはアカデミアの使命の一つと認識されているが、近年特にその傾向が強まっている。研究成果を実用化するためには、アカデミアと企業が協力し合うことが必要であり、研究・開発のうちの開発段階は、最終製品を作り出す機能を持っていないアカデミアが単独で行うより、企

業が市場動向を鑑みながら資金を投入して実施する方が効率的である。そのように開発に企業が関与する場合にアカデミアの研究成果が特許化されていれば、その特許の使用許諾を受けて開発を行う企業を一社に絞ることができ、企業にとっては製品化後に市場を独占することが期待できることから、資金投入をしやすくなる。

また、特許権を取得することにより、目的別に許諾条件を変えることが可能となる他、同一分野の研究推進に必要な特許権を押さえている機関と交渉する際に自身の保有する特許権とのクロスライセンスを持ちかけて交渉を有利に進めることもできる。したがって、特許制度は、アカデミアにおける基礎研究を円滑に進めることにも貢献している。このように、特許権は単に保有者に収入をもたらすのみならず、社会的な見地からも必要とされている。ただし、健康や医療に関する発明の場合、独占排他権である特許権と公共性のバランスが重要であることは言うまでもない。

[遺伝子特許]

ヒトゲノムと人権に関する世界宣言(UNESCO1997年)では、第1条で「象徴的な意味において、ヒトゲノムは、人類の遺産である」と記されている。ヒトの遺伝子の解析により、遺伝子診断方法や医薬品の開発が進められ、人々の健康維持や新たな医療技術につながられる。特許は発明に対して付与されるもので、その対象は、「物」と「方法」であり、自然界に天然に存在する物質はその対象とならない。ヒトの遺伝子解析に関しては、当初は分離・抽出された状態であれば天然物ではないというロジックで特許権が認められていた。しかし、ミリアッド社のBRCA遺伝子に対する特許をめぐる米国の裁判において、その特許性について議論が生じた。これまでは遺伝子特許により結果的にポリヌクレオチドにコードされた「情報」が保護されてきたが、今後は「物」としてcDNA(RNAから人工的に作成されたもののため、天然に存在するとはいえない)が特許取得対象となる。

[データ二次利用の意義]

通常、研究プロジェクトは一定期間が経過すると終了する。そのため、研究終了後は研究主体自体がなくなることになるが、それまでに集められた試料・情報を活用する方策を検討することは重要であり、現時点では、バイオバンクがその役を担うことが多いと考えられる。ただし、当初のインフォームド・コンセントの範囲の問題があるため、今後収集される試料・情報に関しては、バンクへの寄贈や二次利用を念頭に置いたインフォームド・コンセントを得ておくことが望ましいと考えられる。

[ヒト由来試料の帰属と特許権]

ヒトに由来する試料は、細胞や血液といった形でそのヒトの中にあるときは、当然そのヒトに帰属する。しかし、研究利用に関する説明が行われ、同意を得た上で採取された細胞や血液は、インフォームド・コンセントで示された条件で使用するという約束の下でドナーが譲渡したものと考えることができる。また、その試料・

情報を用いて研究が行われた場合、研究によって試料・情報に付加価値がつく。したがって、いったん研究に提供された試料・情報を用いて行われた研究の成果として知的財産権が発生した場合、その権利はサンプルの提供者ではなく、研究者(研究機関)に帰属すると考えるのが適切である(対象者への説明時点でももちろん述べておくべき内容である)。

バンクから提供される試料・情報を用いた二次利用で開発された研究成果に関する特許権の分配方法については、事前に取り決めを交わしておくのが望ましい。一次情報収集者の努力、バンク運用者の作業を考えると、そこにどう報いるかを二次利用者は常に意識することも重要であろう。このことは特許権の配分のみならず、論文の共著者に加えるかどうか、謝辞に述べるかどうかも含めて求められる。ちなみに、バイオバンクジャパンでは、収集機関、バイオバンク、研究機関の特許権持分比率は、4:3:3と覚書で定められている(ただし、変更は可能)。

D. 考察

今後のデータアーカイブ・バンク体制構築の一助とするため、ヒトに由来する試料・情報を用いて研究を実施する際の知的財産権のあり方、さらにそれを二次利用する場合の配分に関する現在の考え方を整理した。

ヒト由来試料から生じる特許権は、更なる研究開発のためにも重要である。その権利は通常、研究者(研究機関)に帰属し、試料提供者には認められていない。また、それらの情報がバンク等に集められたものを二次利用する場合には、事前に特許の配分や論文執筆につき契約書を交わしておくことが重要である。

E. 結論

試料・情報二次利用に際しての知的財産権・法的な課題、ならびに遺伝子特許の現状に関し、情報を得た。アーカイブやバンクによるデータ二次利用を進めるためには、その後に生じ得る知的財産権等に関し、事前に検討しておくことが必要と考

えられた。

F.研究発表

1. 論文発表
 2. 学会発表
- いずれもなし

G. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得
 2. 実用新案登録
 - 3.その他
- いずれもなし

大規模コホートデータにおける一意性の検討

研究分担者 祖父江友孝 大阪大学大学院医学系研究科

研究要旨

個票データの開示を行う際には、一意性のあるデータは個人が同定される可能性があるの
で、一意性のあるデータがどの程度存在するかを検討しておく必要がある。今回、三府県コ
ホートデータにおいて、どのような頻度で一意性が見られるかを確認した。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。複数の変数をそれぞれ組み合わせた場合の分類数 K
とユニークセル数 S_1 のパターンから、分類数の増加に伴い一意であるレコード数は急
増した。一意性は容易に避けられるものではなく、利用の際には一意性があるものと考えて
対応することが必要と考えられた。

A. 目的

三府県コホートデータについて、どのような頻
度で一意性がみられるか検討する。

B. 方法

三府県コホートデータを使用し 100,629 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 (表 1)。

検討に際し変数と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
(表 2)。

(1) 定義

対象の個体 (本研究の場合は 100,629 例) が数

種類の変数の組み合わせに基づいて K 個のセルに
分けられたとき、1 つのセルに含まれる個体数が i
のセル数を $S_i(i = 1, 2, \dots, N)$ とする。つまり、
 $\sum S_i = N$ となる。今回注目するのは個体数が 1 のセ
ルの数であるユニークセル数 S_1 である。なお、個
体自体を呼ぶときには一意という単語を用いるが、
セルに対してはユニークセルという単語を用いる。

(2) 検討内容

[検討 1]

204 変数それぞれ単変数についての、分類数 K と
ユニークセル数 S_1 を求めた。

[検討 2]

全体 (204 変数すべてを組み合わせた場合) の
分類数 K とユニークセル数 S_1 を求めた。

[検討 3]