

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**

総合研究報告書

研究代表者 玉腰 暁子

平成 28 (2016) 年 3 月

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…………… 1
玉腰暁子

II. 分担研究報告

- 他国・他領域におけるデータアーカイブ利活用環境 米国におけるデータアーカイブの研究利用と国内社会科学分野のデータアーカイブの現状 - …… 9
磯博康、大橋靖雄、他
- 大規模コホートデータにおける一意性の検討…………… 12
祖父江友孝
- 疫学追跡終了後コホートデータの共通利用（アーカイブ化）の際の死因データ利用に関する検討…………… 27
大橋靖雄、原田亜紀子
- アーカイブデータ提供のためのガイダンスに含める必要がある項目の整理…………… 35
玉腰暁子、祖父江友孝、他
- 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統計法解釈に関する検討…………… 38
玉腰暁子、磯博康、他
- がん登録推進法とコホート研究との関係に関する研究…………… 41
辻 一郎
- 個人情報保護法改正に伴う疫学研究への影響に関する検討…………… 44
磯博康、辻一郎、他

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ットが行われている。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。このデポジットでは、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、その活用にあたり、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ドラ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

他国・他領域におけるデータアーカイブ利活用環境 米国におけるデータアーカイブの研究利用と 国内社会科学分野のデータアーカイブの現状 -

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

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課題を整理するために、先行している米国、国内社会科学分野の現状を把握し、疫学研究、とくに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の知見を得た。

A. 研究目的

データアーカイブを二次利用するメリットは、

- 1) 既に行われている調査を繰り返さずに済み労力、資金とも無駄な投入を避けることができ、特に多くの変数を得るような調査では得られたすべての情報を調査者が解析することはないため、利用されていない変数について独自のアイデアで解析することで、新たな知見を得ることができること、
- 2) 若手研究者にとっては、自身で小規模な回収率の低い調査を行うことに比べ、質の高い調査データにアクセスできること、
- 3) 学生教育の際にも、実データを用いた教育を行うことができること、である。

しかし、疫学研究のアーカイブデータの提供・利用環境はまだ整っていない。そこで、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。

B. 研究方法

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。

社会科学系分野のデータアーカイブセンターの一つ、東京大学社会科学研究所の附属社会調査・データアーカイブ研究センターの藤原翔氏より、情報提供をいただいた。

C. 研究結果

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、コホート研究を一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ジタル化が行われている。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、倫理審査を事前に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。

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ひとつである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、データ解析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。

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、広く利用可能にすることは、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。そこで、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・単純化と必要経費の設定、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のデータアーカイブは先行しており、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。

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課題を整理するために、先行している米国、国内社会科学分野の現状を把握し、疫学研究、とくに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の知見を得た。

F. 健康機器情報

特になし

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

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。また、標本数を変化させた場合にそれぞれどのような頻度で一意性が見られるかを比較
した。100,629例全てを使用した場合と標本数を減らした場合、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
であるレコード数が増加するという分布の形状は似通っていたが、標本数が大きい場合ほど
少ない分類数でユニークセルの割合が高率に達していた。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った。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、 $S_i=K$ 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。

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 S₁ 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 S₁ を求めた。

[検討 3]

ベースとして【個人特性】と【追跡】のグループを考える。それらについて、今後の解析に支障のないと考えられる範囲で可能な限りセルの併合(まるめの処理)を行い、【個人特性】については 2 パターン、【追跡】については 4 パターンのサブグループを定義し、それらの分類数 K とユニークセル数 S₁ を求めた。

[検討 4]

21 のアンケートカテゴリに対し アンケートカテゴリのみ、【個人特性】とアンケートカテゴリをそれぞれ組み合わせた場合、【追跡】とアンケートカテゴリをそれぞれ組み合わせた場合、【個人特性】【追跡】の組み合わせに各アンケートカテゴリを組み合わせた場合、の全ての場合における分類数 K とユニークセル数 S₁ を求めた。

さらに、三府県コホートデータを使用し 100,629 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 に変化させた場合について検討を行う。分析対象となる変数は、前述の解析と同様、226 変数からなる個人レコードのうち、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 変数とした。

[検討 5]

100,629 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 に変化させた場合について、21 のアンケートカテゴリに対し アンケートカテゴリのみ、【個人特性】とアンケートカテゴリをそれぞれ組み合わせた場合、【追跡】とアンケートカテゴリをそれぞれ組み合わせた場合、【個人特性】【追跡】の組み合わせに各アンケートカテゴリを組み合わせた

場合、の全ての場合における分類数 K とユニークセル数 S₁ を求めた。

C. 結果

[検討 1]より、単体の変数で一意である個体が存在するのは、「v0502(10 年観察終了日)」「v0600(死因 ICD-9コード4桁)」「v1200(身長(cm))」「v1201(体重(kg))」「v1610(初経年齢)」「v1612(自然閉経年齢)」「v1613(手術閉経年齢)」「v1615(出産人数)」「v1616(初産年齢)」「v2101(喫煙開始年齢)」「v2102(喫煙本数/日)」「v2103(禁煙年齢)」「v2801(転入何年前か)」「v2940(最も長く就いた仕事)」「v2950(従事年数)」の 15 変数であった。(表 2)

[検討 2]より、204 の全ての変数を組み合わせた場合に一意となる個体の数は 100,605 であった。

[検討 3]より、性別 × 年齢 × 居住地の情報からなる【個人特性】グループにおいて、まるめの処理を行わない「個人特性 1」では分類数 673、ユニークセル数 19 であったのに対し、年齢を 5 歳階級とし 85 歳以上はまとめた「個人特性 2」では、分類数 120、ユニークセル数は 0 と、一意性が消失した(表 3)。

追跡に関する日付 × 転帰 × 死因からなる【追跡】グループでは、処理を行わない「追跡 1」では分類数 20,176、ユニークセル数 16,631 であったのに対し、まるめの処理として、死因 ICD-9 コードを 3 桁までとする、かつ日付を月までにする(「追跡 2」)ことによりユニークセル数は約半分、同じく死因コード 3 桁かつ日付を追跡期間(単位:月)でみる(「追跡 3」)ことによりさらに半分になり、一意性は減少した。さらに死因情報を除いて日付を追跡期間(単位:月)で見た場合(「追跡 4」)では分類数が 243、ユニークセル数が 0 になり一意性が消失した(表 3)。

[検討 4] の組み合わせから得られた 329 パターンについて、分類数、ユニークセル数、分類数に占めるユニークセル数の割合 S₁/K、を示した。また分

分類数 K を横軸、ユニークセル数 S_1 を縦軸にその分布を示した(図 1)。さらに、分類数 K を横軸、分類数に占めるユニークセル数の割合 S_1/K を縦軸にその分布を示した(図 2)。分類数が小さい時には分類数に占めるユニークセル数の割合も 80% 以下に分布するが、分類数の増加とともにユニークセルの割合が急増し、概ね分類数が 20,000 を超えると 80% 以上に分布した。すなわち、100,629 例全体に対して 16,000 例程度(16%程度)が一意性のある個体数となり、分類数の増加に比例して、一意性のある個体数が増加した。

[検討 5] 100,629 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 に変化させた場合について、の組み合わせから得られた 461 パターンについて、分類数、ユニークセル数、分類数に占めるユニークセル数の割合 S_1/K を示した(表 2)。

また 100,629 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 に変化させた場合について、分類数 K を横軸、ユニークセル数 S_1 を縦軸にその分布を示した(図 3)。さらに、分類数 K を横軸、分類数に占めるユニークセル数の割合 S_1/K を縦軸にその分布を示した(図 4)。100,629 例全てを使用した場合、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 例になるまで急増し、次に 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。100,629 例全てを使用した場合と標本数を減らした場合を比較すると、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)、プラトーに達する部分が狭くなる傾向があった。

D. 考察

各変数のユニークセル数の確認より、一意性には、変数 $v0501$ (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、変数 $v1615$ (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。

しかしながら今回すべての変数を組み合わせた場合の一意である個体の数は 100,605 であり、これは全レコード数の約 99.98% にあたる。このように大規模なコホートデータにおいては、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、本研究に利用した 10 万人規模のデータであるから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。

分類数とユニークセル数の関係から、コホートデータにおいて、変数が増えるほど分類数は増大し、概ね分類数が 20,000 を超えると一意である個体の数も分類数の 80% 以上に分布した。一意性を上げないためには、一つのファイルに含む項目数を増やさず、ファイルを分けて保管することなどが考えられるが、通常、一意性があるものとの前提で対応する必要がある。

死因に関しては、簡単分類を参考とした丸めの方法なども検討する必要がある。

10 万人規模のコホート集団の場合、分類数が全対象者数の概ね 20,000 程度で、ユニークセルの割合が 80% に達していた。対象者数を少なくするにつれて、立ち上がりが緩やかになり、100 例規模のコホート集団では、分類数が 40 程度で、ユニークセルの割合が 80% に達していた。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。

E. 結論

三府県コホートデータより、各変数、全変数あるいはいくつかの変数の組合せごとに一意性を検討した。三府県コホートデータのような10万人規模のデータの場合、分類数が概ね20,000を超えると一意性のある個体数は分類数の80%以上となり、一意性があるものとの前提で対応を考える必要がある。

10万人規模のコホート集団の場合、分類数が全対象者数の20%程度で、ユニークセルの割合が80%に達していた。100例規模のコホート集団では、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。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。

F. 健康機器情報

G. 研究発表

1. 論文発表
2. 学会発表

いずれもなし

H. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得
2. 実用新案登録
3. その他

いずれもなし

表 1 対象となる変数について（グレーの変数は解析対象外）

変数名	属性	内容	定義の相違	凡例など	備考
v0000	数値	通し ID	共通	1-100629	府県(v0400)、自治体(v0403)、地域固有 ID(v0001)のソート順
v0001	数値	地域固有 ID	府県間で異なる	(欠損なし)	
v0100	数値	観察期間=負 フラゲ	共通	1: Yes (欠損=.)	解析時除外。 追跡開始日(v0500)>観察終了日(v0521)。 愛知都市地区のみ 14 例。
v0200	数値	性別	共通	0:男, 1:女(欠損なし)	
v0300	数値	追跡開始年齢	共通	40-99(欠損なし)	
v0400	数値	府県番号	共通	1:宮城, 2:愛知, 3:大坂	
v0401	数値	地区番号	共通	1: 都市地区, 0: 対照地区	
v0402	数値	地区通し番号	共通	1: 宮城対照, 2: 宮城都市, 3: 愛知対照, 4: 愛知都市, 5: 大阪対照, 6: 大阪都市	
v0403	数値	自治体番号	府県間で異なる	宮城: 1:仙台,2:涌谷,3:田尻; 愛知: 1:名古屋,2:犬山; 大阪: 1:東成,2:能勢,3:河南,4:熊取	
v0500	SAS 日付	追跡開始日	共通	(欠損なし)	
v0501	SAS 日付	10年コホート追跡終了日(地域ごとの定義)	共通(ただし日付は地区間で異なる)	宮城: 1993/12/31, 愛知都市地区: 1995/10/31, 愛知対照地区: 1995/6/30, 大阪東成: 1994/10/31, 大阪能勢: 1993/1/31, 大阪河南: 1994/11/30, 大阪熊取: 1995/1/31	
v0502	SAS 日付	10年観察終了日(個人ごとの定義)	共通	コホート10年追跡終了日, 転出日, 死亡日のうち一番早い日(欠損なし)	
v0503	数値	10年転帰	共通	3: コホート追跡終了日において生存, 6: コホート追跡終了日までに死亡, 9: コホート追跡終了日までに転出(欠損なし)	
v0511	SAS 日付	15年コホート追跡終了日(地域ごとの定義)	共通(ただし日付は地区間で異なる)	宮城: 1993/12/31, 愛知都市地区: 2000/10/31, 愛知対照地区: 2000/6/30, 大阪東成: 1999/10/31, 大阪能勢: 1998/1/31, 大阪河南: 1999/11/30, 大阪熊取: 2000/1/31(欠損なし)	
v0512	SAS 日付	15年観察終了日(個人ごとの定義)	共通	コホート15年追跡終了日, 転出日, 死亡日のうち一番早い日(欠損なし)	
v0513	数値	15年転帰	共通	3: コホート追跡終了日において生存, 6: コホート追跡終了日までに死亡, 9: コホート追跡終了日までに転出(欠損なし)	
v0520	数値	転帰(受領データママ)	共通	3: 生存, 6: 死亡, 9: 転出(欠損なし)	各地域の最終確認時の転帰。解析時は追跡終了日定義の上調整が必要。 愛知は転出(v0520=9)で死因ありのデータが都市地区 277 例、対照地区 73 例あり 転出として扱い、死因は無視(若井先生確認済)。
v0521	SAS 日付	観察終了日(個人ごとの定義: 受領データママ)	共通	コホート追跡終了日による定義をする前の個人ごとの観察終了日。	愛知のみコホート追跡終了日より後の日付あり(2000/12/31まで)。
v0600	数値	死因 ICD-9 コード 4 桁	共通	ICD-9(例: 38.9, 150.1)	基本となる死因変数(ただし、宮城の15年死亡はv0601,v0602とOR結合が必要)。 10年死亡の死因別死亡例はこの変数のみで同定可。
v0601	数値	死因 ICD-9 コード 4 桁(宮城15年死亡)	宮城のみ	ICD-9(例: 38.9, 150.1)	宮城15年死亡の死因はv0600(ICD-9), v0601(ICD-9), v0602(ICD-10)の3変数に入っている(これら3変数間に重複なし)。 15年死亡の死因別死亡同定にはこれら3変数のOR結合が必要。
v0602	文字	死因 ICD-10 コード 4 桁(宮城15年死亡)	宮城のみ	ICD-10(例: C030, I219)	
v0610	文字	死因 ICD-9 コード	府県間で異なる	ICD-9(元データのママのため、形式が異なる)。	v0600の数値化前データ
v0611	文字	死因 ICD-9 または ICD-10 コード(宮城15年死亡)	宮城のみ	ICD-9とICD-10が混在(元データのママ)	v0601, v0602のICD-9とICD-10に分ける前のデータ
v0620	数値	10年死因不明フラゲ	共通	1: Yes(欠損=.)	v0600=. かつ v0611="" かつ 10年死亡(v0503=6)。 10年死亡例中19例あり(すべて宮城)。
v0621	数値	15年死因不明フラゲ	共通	1: Yes(欠損=.)	v0600=. かつ v0611="" かつ 15年死亡(v0513=6)。 15年死亡例中29例あり(宮城28例、大阪東成1例)。

v1000	数値	自覚症状あるか 咳	共通	1: Yes (欠損=.)	明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1001	数値	自覚症状あるか 痰	共通	1: Yes (欠損=.)	明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1002	数値	自覚症状あるか 血痰	共通	1: Yes (欠損=.)	明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1003	数値	自覚症状あるか 動悸	共通	1: Yes (欠損=.)	明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1004	数値	自覚症状あるか 息切れ	共通	1: Yes (欠損=.)	大阪東成は欠損。 明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1005	数値	自覚症状あるか 食欲不振	共通	1: Yes (欠損=.)	大阪東成は欠損。 明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1006	数値	自覚症状あるか 便秘	共通	1: Yes (欠損=.)	大阪東成は欠損。 明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1007	数値	自覚症状あるか 不眠	共通	1: Yes (欠損=.)	愛知、大阪熊取、大阪河南は欠損。 明示回答ありなら 1、それ以外は欠損。
v1008	数値	自覚症状あるか やせてきた	共通	1: Yes (欠損=.)	明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1009	数値	自覚症状あるか 疲れやすい	共通	1: Yes (欠損=.)	明示回答ありなら 1、それ以外は欠損。 宮城、大阪熊取は v1011 と矛盾ありママ。
v1010	数値	自覚症状あるか 喘息の発作	共通	1: Yes (欠損=.)	宮城、大阪東成は欠損。 明示回答ありなら 1、それ以外は欠損。
v1011	数値	自覚症状あるか なし	共通	1: Yes (欠損=.)	「なし」回答ありの場合 1、それ以外は欠損。 宮城、大阪熊取は v1000-v1009 と矛盾ありママ。
v1100	数値	既往 高血圧	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1101	数値	既往 心臓病	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1102	数値	既往 脳卒中	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1103	数値	既往 糖尿病	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1104	数値	既往 肝臓病	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1105	数値	既往 腎臓病	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は欠損。
v1106	数値	既往 胃十二指腸潰瘍	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	大阪能勢は欠損。 宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1107	数値	既往 胆石	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	大阪東成、大阪能勢、大阪河南は欠損。 宮城、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1108	数値	既往 虫垂炎	一部相違	1: Yes (欠損=.)	宮城のみ。 疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。
v1109	数値	既往 結核ろく膜炎	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 大阪能勢は結核と肋膜炎の結合 (選択肢 1,2 は 1 優先で結核と肋膜炎の OR 結合、, 選択肢 3 は 1,2 でない場合に結核と肋膜炎の AND 結合)。 宮城、大阪熊取は v1120 との矛盾ママ。
v1110	数値	既往 肺炎	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1111	数値	既往 喘息	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1112	数値	既往 慢性気管支炎	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。
v1113	数値	既往 肺気腫	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上 1 に割り振り回答なしは欠損。 宮城、大阪熊取は v1120 との矛盾ママ。

v1114	数値	既往 じん肺	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪熊取はv1120との矛盾ママ。
v1115	数値	既往 貧血	共通	1: Yes (欠損=.)	大阪東成、大阪河南のみ。 疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。
v1116	数値	既往 アレルギー性鼻炎	共通	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	大阪能勢のみ。
v1117	数値	既往 蕁麻疹	共通	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	大阪能勢のみ。
v1118	数値	既往 湿疹・皮膚炎	共通	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=.)	大阪能勢のみ。
v1119	数値	既往 痛風	共通	1: Yes (欠損=.)	宮城のみ
v1120	数値	既往 なし	共通	1: Yes (欠損=.)	「なし」回答ありの場合1、それ以外は欠損。 大阪能勢欠損。 宮城、大阪熊取はv1100-v1119との矛盾ママ。
v1200	数値	身長 (cm)	共通	(欠損=.)	外れ値ママ
v1201	数値	体重 (kg)	共通	(欠損=.)	外れ値ママ
v1202	数値	BMI 連続変数	共通	(欠損=.)	v1201/v1200/v1200*10000で計算。v1200またはv1201が0, 999.9, 欠損のいずれかの場合は欠損。
v1300	数値	胸部レントゲン検査受診有無	一部相違	1: Yes, 0: No (欠損=.)	宮城、愛知、大阪熊取は「胸部のレントゲン検査(結核検診)」。 宮城は「昨年(昭和58年)1年間に」、大阪東成、大阪河南は「この1年のあいだに」、それ以外は単に「受けたことがありますか」と聞いて下位で時期質問。
v1301	数値	胸部レントゲン検査受診有無(1年以内)	一部相違	1: Yes (欠損=.)	宮城、大阪東成、大阪河南はv1300=1、愛知はv1300=1かつ昭和60年受診、大阪能勢はv1300=1かつ昭和57年度受診、大阪熊取はv1300=1かつ昭和59年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1302	数値	胸部レントゲン検査受診有無(1年より前~2年以内)	一部相違	1: Yes (欠損=.)	宮城、大阪東成、大阪河南は欠損。愛知はv1300=1かつ昭和59年受診、大阪能勢はv1300=1かつ昭和56年度受診、大阪熊取はv1300=1かつ昭和58年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1303	数値	胸部レントゲン検査受診有無(2年より前)	一部相違	1: Yes (欠損=.)	宮城、大阪東成、大阪河南は欠損。愛知はv1300=1かつ昭和59年より前受診、大阪能勢はv1300=1かつ昭和56年度より前受診、大阪熊取はv1300=1かつ昭和58年より前受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1304	数値	胸部レントゲン検査受診場所(住民健診)	一部相違	1: Yes (欠損=.)	大阪熊取は「集団検診で(例えば成人病予防協会検診車)」、大阪東成は欠損。 v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1305	数値	胸部レントゲン検査受診場所(保健所・保健センター)	一部相違	1: Yes (欠損=.)	大阪東成は「保健所や小学校の市民検診」。 v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1306	数値	胸部レントゲン検査受診場所(職場)	一部相違	1: Yes (欠損=.)	v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1307	数値	胸部レントゲン検査受診場所(病院など)	一部相違	1: Yes (欠損=.)	v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1308	数値	胸部レントゲン検査受診場所(農協)	一部相違	1: Yes (欠損=.)	大阪熊取と大阪河南のみ。 v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1309	数値	胸部レントゲン検査受診場所(その他)	一部相違	1: Yes (欠損=.)	v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1400	数値	胃レントゲン検査受診有無	一部相違	1: Yes, 0: No (欠損=.)	宮城、愛知対照、大阪熊取は「胃のレントゲン検査(胃腸病検診)」; 大阪東成、大阪河南は「胃(ガン)検診」。 宮城は「昨年(昭和58年)1年間に」、大阪東成、大阪河南は「この1年のあいだに」、それ以外は単に「受けたことがありますか」と聞いて下位で時期質問。
v1401	数値	胃レントゲン検査受診有無(1年以内)	一部相違	1: Yes (欠損=.)	宮城、大阪東成、大阪河南はv1400=1、愛知はv1400=1かつ昭和60年受診、大阪能勢はv1400=1

					かつ昭和 57 年度受診、大阪熊取は v1400=1 かつ昭和 59 年受診。 注： ベースライン調査は、宮城は昭和 59 年 1～2 月、愛知名古屋は昭和 60 年 10～11 月、愛知犬山は昭和 60 年 7～8 月、大阪能勢は昭和 58 年 2 月、大阪熊取は昭和 60 年 2～3 月、大阪河南は昭和 59 年 11～12 月、大阪東成は昭和 59 年 10～11 月に実施。
v1402	数値	胃レントゲン検査受診有無（1 年より前～2 年以内）	一部相違	1: Yes (欠損=.)	宮城、大阪東成、大阪河南は欠損。愛知は v1400=1 かつ昭和 59 年受診、大阪能勢は v1400=1 かつ昭和 56 年度受診、大阪熊取は v1400=1 かつ昭和 58 年受診。 注： ベースライン調査は、宮城は昭和 59 年 1～2 月、愛知名古屋は昭和 60 年 10～11 月、愛知犬山は昭和 60 年 7～8 月、大阪能勢は昭和 58 年 2 月、大阪熊取は昭和 60 年 2～3 月、大阪河南は昭和 59 年 11～12 月、大阪東成は昭和 59 年 10～11 月に実施。
v1403	数値	胃レントゲン検査受診有無（2 年より前）	一部相違	1: Yes (欠損=.)	宮城、大阪東成、大阪河南は欠損。愛知は v1400=1 かつ昭和 59 年より前受診、大阪能勢は v1400=1 かつ昭和 56 年度より前受診、大阪熊取は v1400=1 かつ昭和 58 年より前受診。 注： ベースライン調査は、宮城は昭和 59 年 1～2 月、愛知名古屋は昭和 60 年 10～11 月、愛知犬山は昭和 60 年 7～8 月、大阪能勢は昭和 58 年 2 月、大阪熊取は昭和 60 年 2～3 月、大阪河南は昭和 59 年 11～12 月、大阪東成は昭和 59 年 10～11 月に実施。
v1404	数値	胃レントゲン検査受診場所（住民健診）	一部相違	1: Yes (欠損=.)	大阪熊取は「集団検診で（例えば成人病予防協会検診車）」。愛知名古屋は「市医師会健診センターまたは休日急病診療所（検診車）」。 v1400=1 かつ当該場所の受診有無・受診年いずれか回答ありなら 1。
v1405	数値	胃レントゲン検査受診場所（保健所・保健センター）	一部相違	1: Yes (欠損=.)	愛知名古屋と大阪東成のみ。 v1400=1 かつ当該場所の受診有無・受診年いずれか回答ありなら 1。
v1406	数値	胃レントゲン検査受診場所（職場）	一部相違	1: Yes (欠損=.)	v1400=1 かつ当該場所の受診有無・受診年いずれか回答ありなら 1。
v1407	数値	胃レントゲン検査受診場所（病院など）	一部相違	1: Yes (欠損=.)	v1400=1 かつ当該場所の受診有無・受診年いずれか回答ありなら 1。
v1408	数値	胃レントゲン検査受診場所（農協）	一部相違	1: Yes (欠損=.)	大阪熊取と大阪河南のみ。 v1400=1 かつ当該場所の受診有無・受診年いずれか回答ありなら 1。
v1409	数値	胃レントゲン検査受診場所（その他）	一部相違	1: Yes (欠損=.)	v1400=1 かつ当該場所の受診有無・受診年いずれか回答ありなら 1。
v1410	数値	胃レントゲン検査受診回数（5 年間）	共通	(欠損=.)	大阪東成、大阪能勢、大阪河南は欠損。
v1500	数値	健康診査または血圧検査受診有無	一部相違	1: Yes, 0: No (欠損=.)	愛知は欠損。 宮城、大阪能勢、大阪熊取は「血圧の検査」、大阪東成、大阪河南は「健康診査（血圧測定や尿検査）」。 宮城は「昨年（昭和 58 年）1 年間に」、大阪東成、大阪河南は「この 1 年のあいだに」、それ以外は単に「受けたことがありますか」と聞いて下位で時期質問。
v1501	数値	健康診査または血圧検査受診有無（1 年以内）	一部相違	1: Yes (欠損=.)	愛知は欠損。 宮城、大阪東成、大阪河南は v1500=1、大阪能勢は v1500=1 かつ昭和 57 年度受診、大阪熊取は v1500=1 かつ昭和 59 年受診。 注： ベースライン調査は、宮城は昭和 59 年 1～2 月、愛知名古屋は昭和 60 年 10～11 月、愛知犬山は昭和 60 年 7～8 月、大阪能勢は昭和 58 年 2 月、大阪熊取は昭和 60 年 2～3 月、大阪河南は昭和 59 年 11～12 月、大阪東成は昭和 59 年 10～11 月に実施。
v1502	数値	健康診査または血圧検査受診有無（1 年より前～2 年以内）	一部相違	1: Yes (欠損=.)	宮城、愛知、大阪東成、大阪河南は欠損。 大阪能勢は v1500=1 かつ昭和 56 年度受診、大阪熊取は v1500=1 かつ昭和 58 年受診。 注： ベースライン調査は、宮城は昭和 59 年 1～2 月、愛知名古屋は昭和 60 年 10～11 月、愛知犬山は昭和 60 年 7～8 月、大阪能勢は昭和 58 年 2 月、大阪熊取は昭和 60 年 2～3 月、大阪河南は昭和 59 年 11～12 月、大阪東成は昭和 59 年 10～11 月に実施。
v1503	数値	健康診査または血圧検査受診有無（2 年より前）	一部相違	1: Yes (欠損=.)	宮城、愛知、大阪東成、大阪河南は欠損。 大阪能勢は v1500=1 かつ昭和 56 年度より前受診、大阪熊取は v1500=1 かつ昭和 58 年より前受診。 注： ベースライン調査は、宮城は昭和 59 年 1～2 月、愛知名古屋は昭和 60 年 10～11 月、愛知犬山は昭和 60 年 7～8 月、大阪能勢は昭和 58 年 2 月、大阪熊取は昭和 60 年 2～3 月、大阪河南は昭和 59 年 11～12 月、大阪東成は昭和 59 年 10～11 月に実施。
v1504	数値	健康診査または血圧検査受診	一部相	1: Yes (欠損=.)	大阪東成は「地区健康相談」。

		場所（住民健診）	違		v1500=1 かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1505	数値	健康診査または血圧検査受診場所（保健所・保健センター）	一部相違	1: Yes（欠損=）	大阪東成は「保健所や小学校の市民検診」。
v1506	数値	健康診査または血圧検査受診場所（職場）	一部相違	1: Yes（欠損=）	v1500=1 かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1507	数値	健康診査または血圧検査受診場所（病院など）	一部相違	1: Yes（欠損=）	v1500=1 かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1508	数値	健康診査または血圧検査受診場所（農協）	一部相違	1: Yes（欠損=）	大阪熊取と大阪河南のみ。
v1509	数値	健康診査または血圧検査受診場所（その他）	一部相違	1: Yes（欠損=）	v1500=1 かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1600	数値	子宮がん検診受診有無	一部相違	1: Yes, 0: No（欠損=）	女性のみ定義。宮城は「昨年（昭和58年）1年間に」、愛知は「昨年（昭和59年）1年間に」、大阪東成、大阪河南は「この1年のあいだに」、それ以外は単に「受けたことがありますか」と聞いて下位で時期質問。
v1601	数値	子宮がん検診受診有無（昨年）	一部相違	1: Yes（欠損=）	女性のみ定義。宮城、愛知、大阪東成、大阪河南はv1600=1。大阪熊取はv1600=1 かつ昭和57年度受診、大阪熊取はv1600=1 かつ昭和59年受診。 注：ベースライン調査は、宮城は昭和59年1～2月、愛知名古屋は昭和60年10～11月、愛知犬山は昭和60年7～8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2～3月、大阪河南は昭和59年11～12月、大阪東成は昭和59年10～11月に実施。
v1602	数値	子宮がん検診受診有無（一昨年）	一部相違	1: Yes（欠損=）	女性のみ定義。宮城、愛知、大阪東成、大阪河南は欠損。 大阪熊取はv1600=1 かつ昭和56年度受診、大阪熊取はv1600=1 かつ昭和58年受診。 注：ベースライン調査は、宮城は昭和59年1～2月、愛知名古屋は昭和60年10～11月、愛知犬山は昭和60年7～8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2～3月、大阪河南は昭和59年11～12月、大阪東成は昭和59年10～11月に実施。
v1603	数値	子宮がん検診受診有無（一昨年より前）	一部相違	1: Yes（欠損=）	女性のみ定義。宮城、愛知、大阪東成、大阪河南はv1600と同義。大阪熊取はv1600=1 かつ昭和56年度より前受診、大阪熊取はv1600=1 かつ昭和58年より前受診。 注：ベースライン調査は、宮城は昭和59年1～2月、愛知名古屋は昭和60年10～11月、愛知犬山は昭和60年7～8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2～3月、大阪河南は昭和59年11～12月、大阪東成は昭和59年10～11月に実施。
v1604	数値	子宮がん検診受診場所（検診車・住民検診）	一部相違	1: Yes（欠損=）	女性のみ定義。大阪東成、大阪熊取、大阪熊取は欠損。大阪河南は「河南町が行った集団検診」。
v1605	数値	子宮がん検診受診場所（保健所・保健センター）	一部相違	1: Yes（欠損=）	宮城、愛知は「検診車」。
v1606	数値	子宮がん検診受診場所（職場）	一部相違	1: Yes（欠損=）	v1600=1 かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1607	数値	子宮がん検診受診場所（病院など）	一部相違	1: Yes（欠損=）	女性のみ定義。v1600=1 かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1608	数値	子宮がん検診受診場所（東成母子会自己採取）	一部相違	1: Yes（欠損=）	女性のみ定義。大阪東成のみ。
v1609	数値	子宮がん検診受診場所（その他）	一部相違	1: Yes（欠損=）	v1600=1 かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1610	数値	初経年齢	共通	（欠損=）	女性のみ定義。大阪は欠損。
v1611	数値	現在月経有り	共通	1: Yes, 0: No（欠損=）	女性のみ定義。大阪は欠損。
v1612	数値	自然閉経年齢	共通	（欠損=）	女性のみ定義。大阪は欠損。V1611に関わらず定義。宮城はv1613との重複回答あり。
v1613	数値	手術閉経年齢	共通	（欠損=）	女性のみ定義。大阪は欠損。V1611に関わらず定義。宮城はv1613との重複回答あり。
v1614	数値	出産経験あるか	共通	1: Yes, 0: No（欠損=）	女性のみ定義。
v1615	数値	出産人数	共通	（欠損=）	愛知都市地区は欠損。女性かつ「出産経験=なし以外」は定義（出産経験不明例の出産人数ママ、出産あり例の0人ママ）。女性かつ「出産経験=なし」は0人。
v1616	数値	初産年齢	共通	（欠損=）	愛知都市地区は欠損。女性かつ「出産経験=なし以外」は定義（出産経験不明例の出産人数ママ、出産あり例の0人ママ）。女性かつ「出産経験=なし」は0人。
v1700	数値	健康保険カテゴリ	一部相違	1: 国民健康保険, 2: 健康保険（政府管掌）, 3: 健康保	選択肢6は大阪対照地区のみ（2と3は欠損）。

				除(組合等), 4: 共済, 5: その他の保険, 6: 健康保険(政府管掌・組合), 9: 不明(欠損=.)	
v1701	数値	健康保険本人・家族別	共通	1: 本人・世帯主, 2: 家族(欠損=.)	
v1800	数値	ご飯1日何杯	共通	(欠損=.)	宮城と愛知は2桁。大阪は1桁。
v1801	数値	パン摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1802	数値	肉摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1803	数値	魚類摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1804	数値	卵摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1805	数値	牛乳摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1806	数値	緑黄色野菜摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1807	数値	その他の野菜摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1808	数値	果物摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1809	数値	みそ汁摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1810	数値	漬物摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	
v1811	数値	インスタント食品摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日(欠損=.)	宮城は欠損。
v1900	数値	緑茶摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上(欠損=.)	大阪能勢は「1.ほとんど飲まない」
v1901	数値	紅茶摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上(欠損=.)	大阪能勢は「1.ほとんど飲まない」
v1902	数値	コーヒー(非インスタント)摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上(欠損=.)	大阪能勢は「1.ほとんど飲まない」 宮城は欠損。
v1903	数値	インスタントコーヒー摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上(欠損=.)	大阪能勢は「1.ほとんど飲まない」 宮城は欠損。
v1904	数値	コーヒー(豆またはインスタント)摂取頻度カテゴリ	一部相違	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上(欠損=.)	宮城のみオリジナル変数。 他はv1902とv1903のOR結合(摂取多い選択肢優先で、選択肢1はAND結合)で定義。
v2000	数値	飲酒習慣カテゴリ	一部相違	1: ほとんど毎日, 2: ときどき, 3: 集まりのときだけ, 4: 以前はよく飲んだが最近飲まない, 5: 以前から飲まない(欠損=.)	大阪能勢は選択肢4はなし
v2100	数値	喫煙習慣カテゴリ	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 時々吸っている(欠損=.)	選択肢4は大阪能勢のみ。
v2101	数値	喫煙開始年齢 連続変数	共通	(欠損=.)	宮城、大阪東成、大阪熊取は生涯喫煙以外の回答ママ。 年齢との比較など論理チェックせず(10未満あり)。 99はママ(愛知)。

v2102	数値	喫煙本数/日 連続変数	共通	(欠損=.)	99 はママ(愛知、大阪)。 宮城、大阪東成、大阪熊取は生涯喫煙以外の回答ママ。
v2103	数値	禁煙年齢 連続変数	共通	(欠損=.)	年齢との比較など論理チェックせず。 宮城、大阪東成、大阪熊取は過去喫煙以外の回答ママ。 99 はママ(愛知)。
v2104	数値	吸い込むか	共通	1: Yes, 0: No (欠損=.)	宮城、大阪東成、大阪熊取は生涯喫煙以外の回答ママ。
v2200	数値	父親の既往 脳卒中	共通	1: Yes (欠損=.)	
v2201	数値	父親の既往 心臓病	共通	1: Yes (欠損=.)	
v2202	数値	父親の既往 がん	共通	1: Yes (欠損=.)	
v2203	数値	父親の既往 上記3つなし	共通	1: Yes (欠損=.)	宮城、大阪熊取は v2200, v2201, v2202 との矛盾ママ。
v2204	数値	父親の既往 知らない	共通	1: Yes (欠損=.)	大阪熊取は欠損。宮城、大阪熊取は v2200, v2201, v2202 との矛盾ママ。
v2205	数値	父親の既往 がん 食道	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2202 との矛盾ママ。
v2206	数値	父親の既往 がん 胃	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2202 との矛盾ママ。
v2207	数値	父親の既往 がん 大腸	共通	1: Yes (欠損=.)	大阪熊取は欠損。宮城、大阪東成、大阪熊取は v2202 との矛盾ママ。
v2208	数値	父親の既往 がん 肝臓	共通	1: Yes (欠損=.)	大阪東成、宮城、大阪熊取は v2202 との矛盾ママ。
v2209	数値	父親の既往 がん 肺	共通	1: Yes (欠損=.)	大阪東成、宮城、大阪熊取は v2202 との矛盾ママ。
v2210	数値	父親の既往 がん 膀胱	共通	1: Yes (欠損=.)	大阪東成、宮城、大阪熊取は v2202 との矛盾ママ。
v2211	数値	父親の既往 がん 白血病	共通	1: Yes (欠損=.)	大阪東成、宮城、大阪熊取は v2202 との矛盾ママ。
v2212	数値	父親の既往 がん その他	共通	1: Yes (欠損=.)	大阪東成、宮城、大阪熊取は v2202 との矛盾ママ。
v2300	数値	母親の既往 脳卒中	共通	1: Yes (欠損=.)	
v2301	数値	母親の既往 心臓病	共通	1: Yes (欠損=.)	
v2302	数値	母親の既往 がん	共通	1: Yes (欠損=.)	
v2303	数値	母親の既往 上記3つなし	共通	1: Yes (欠損=.)	宮城、大阪熊取は v2300, v2301, v2302 との矛盾ママ。
v2304	数値	母親の既往 知らない	共通	1: Yes (欠損=.)	大阪熊取は欠損。宮城、大阪熊取は v2300, v2301, v2302 との矛盾ママ。
v2305	数値	母親の既往 がん 食道	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2306	数値	母親の既往 がん 胃	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2307	数値	母親の既往 がん 大腸	共通	1: Yes (欠損=.)	大阪熊取は欠損。宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2308	数値	母親の既往 がん 肝臓	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2309	数値	母親の既往 がん 肺	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2310	数値	母親の既往 がん 膀胱	共通	1: Yes (欠損=.)	大阪熊取は欠損。宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2311	数値	母親の既往 がん 白血病	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2312	数値	母親の既往 がん 子宮	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2313	数値	母親の既往 がん 乳房	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2314	数値	母親の既往 がん その他	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2302 との矛盾ママ。
v2400	数値	小学生の頃の父親喫煙有無	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 知らない (欠損=.)	大阪熊取は時期を特定せず「お父さん(実父)はすっておられますか」。 選択肢 2 は熊取のみ。愛知は受領データ選択肢 1(吸う), 2(吸わない), 0 だったのでそれぞれ 1, 3, (欠損)に分配。
v2401	数値	小学生の頃の母親喫煙有無	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 知らない (欠損=.)	大阪熊取は時期を特定せず「お母さん(実母)はすっておられますか」。 選択肢 2 は熊取のみ。愛知は受領データ選択肢 1(吸う), 2(吸わない), 0 だったのでそれぞれ 1, 3, (欠損)に分配。
v2500	数値	現在同居家族喫煙あり	共通	1: Yes, 0: No (欠損=.)	
v2501	数値	現在夫喫煙あり	共通	1: Yes (欠損=.)	女性のみ定義。宮城、大阪東成、大阪熊取は v2500=1 以外でも回答あり。
v2502	数値	現在妻喫煙あり	共通	1: Yes (欠損=.)	男性のみ定義。宮城、大阪東成、大阪熊取は v2500=1 以外でも回答あり。
v2503	数値	現在父喫煙あり	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2500=1 以外でも回答あり。
v2504	数値	現在母喫煙あり	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2500=1 以外でも回答あり。
v2505	数値	現在子喫煙あり	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2500=1 以外でも回答あり。
v2506	数値	現在その他喫煙あり	共通	1: Yes (欠損=.)	宮城、大阪東成、大阪熊取は v2500=1 以外でも回答あり。
v2600	数値	暖房 なし	共通	1: Yes (欠損=.)	
v2601	数値	暖房 電気(複数回答)	共通	1: Yes (欠損=.)	宮城、大阪熊取は v2600=1 でも回答あり。
v2602	数値	暖房 エアコン(複数回答)	共通	1: Yes (欠損=.)	宮城、大阪熊取は v2600=1 でも回答あり。
v2603	数値	暖房 クリーンヒーター(複数回答)	共通	1: Yes (欠損=.)	宮城、大阪熊取は v2600=1 でも回答あり。
v2604	数値	暖房 煙突付きストーブ(複数回答)	共通	1: Yes (欠損=.)	宮城、大阪熊取は v2600=1 でも回答あり。
v2605	数値	暖房 煙突なしストーブ(複数回答)	共通	1: Yes (欠損=.)	宮城、大阪熊取は v2600=1 でも回答あり。

v2606	数値	暖房 炭（複数回答）	共通	1: Yes（欠損=.）	宮城、大阪熊取は v2600=1 でも回答あり。
v2700	数値	住居がバスの通る道路に面しているか	共通	1: Yes, 0: No（欠損=.）	
v2800	数値	生まれてからいまのところに住んでいるか	共通	1: 生まれてから住んでいる, 0: 他の市区町村から引っ越し（欠損=.）	
v2801	数値	転入何年前か	一部相違	（欠損=.）	愛知は「在住年数（計算）」として受領データセットにあり。 年齢との比較など論理チェックせず0以上ならママ。 大阪は欠損。
v2802	数値	転入年の元号	共通	1: 明治, 2: 大正, 3: 昭和（欠損=.）	宮城は欠損。v2800, v2803 との整合性チェックせずママ。
v2803	数値	転入年（和暦）	共通	（欠損=.）	宮城は欠損。v2800, v2802 との整合性チェックせず年数1以上ならママ。
v2804	数値	転入年（西暦）	共通	（欠損=.）	宮城は欠損。v2802, v2803 から単純計算（実在元号年かつ調査年以前の場合のみ）。
v2901	数値	職業 病院・診療所	一部相違	1: Yes（欠損=.）	大阪能勢は欠損。宮城は2段階質問形式の下位選択肢（場所 病院・診療所・歯科診療所・歯科技工所）。
v2902	数値	職業 美容院・理髪店	一部相違	1: Yes（欠損=.）	大阪能勢は欠損。宮城は2段階質問形式の下位選択肢（場所 美容院・理髪店）。
v2903	数値	職業 紡績・綿布・タオル工場	一部相違	1: Yes（欠損=.）	大阪能勢は「木綿・亜麻・大麻の工場」。宮城は2段階質問形式の下位選択肢（粉じん 繊維）。
v2904	数値	職業 出版・印刷工場	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（現場 印刷工場）。
v2905	数値	職業 メッキ作業所	一部相違	1: Yes（欠損=.）	大阪能勢は欠損。宮城は2段階質問形式の下位選択肢（現場 メッキ作業所）。
v2906	数値	職業 運輸業務	一部相違	1: Yes（欠損=.）	大阪能勢は欠損。宮城は2段階質問形式の第1問。
v2907	数値	職業 鉄鋼工場	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（現場 鉄鋼工場）。
v2908	数値	職業 金属製品工場	一部相違	1: Yes（欠損=.）	大阪能勢は欠損。 宮城は2段階質問形式の下位選択肢（粉じん 金属研磨）。
v2909	数値	職業 鋳物工場	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（粉じん 鋳物（サンドブラストなど））。
v2910	数値	職業 陶磁器・ガラス工場	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（粉じん 陶磁器、製造 ガラス製品）。
v2911	数値	職業 プラスチック加工	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（製造 プラスチック製品）。
v2912	数値	職業 ゴム製品製造	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（製造 ゴム製品）。
v2913	数値	職業 耐熱・保温作業	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（現場 耐熱保温作業現場）。
v2914	数値	職業 造船所	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（現場 造船所）。
v2915	数値	職業 部品解体・再生工場	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（現場 部品解体・再生現場）。
v2916	数値	職業 建設・解体現場	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（現場 建設・解体現場）。
v2917	数値	職業 石油精製・石油製品の製造	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（現場 石油精錬所、製造 石油製品）。
v2918	数値	職業 ビッチ・タール・アスファルト製造	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（製造 ビッチ・タール・アスファルト）。
v2919	数値	職業 バッテリー・電池製造	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（製造 バッテリー・電池）。
v2920	数値	職業 セメント・生コン製品製造	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（扱う仕事 セメント・生コン）。
v2921	数値	職業 石綿工場	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（扱う仕事 石綿または石綿含有物質）。
v2922	数値	職業 ランドリー・洗濯工場	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（場所 ランドリー・クリーニング）。
v2923	数値	職業 土石採取・石材加工	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の下位選択肢（粉じん 石材、採土・採石）。
v2924	数値	職業 自動車修理・整備	一部相違	1: Yes（欠損=.）	大阪能勢は欠損。宮城は2段階質問形式の下位選択肢（現場 自動車修理作業所）。
v2925	数値	職業 写真現像所	一部相違	1: Yes（欠損=.）	大阪能勢は欠損。宮城は2段階質問形式の下位選択肢（場所 写真現像所）。
v2926	数値	職業 炭坑・その他の鉱山、鉱石の精錬	一部相違	1: Yes（欠損=.）	宮城は2段階質問形式の第1問（鉱業）。
v2927	数値	職業 サービス業（公務員も含む）	共通	1: Yes（欠損=.）	宮城、大阪能勢は欠損。
v2928	数値	職業 林業	一部相違	1: Yes（欠損=.）	大阪東成、大阪能勢は欠損。宮城は2段階質問形式の第1問。
v2929	数値	職業 農業（畜産・果樹園芸も含む）	一部相違	1: Yes（欠損=.）	大阪能勢は欠損。宮城は2段階質問形式の第1問。
v2930	数値	職業 皮革製品製造	一部相違	1: Yes（欠損=.）	宮城、大阪能勢のみ。宮城は2段階質問形式の下位選択肢（製造 皮革製品）。
v2931	数値	職業 殺虫剤・枯草剤製造	一部相違	1: Yes（欠損=.）	宮城、大阪能勢のみ。宮城は2段階質問形式の下位選択肢（製造 除草剤・枯草剤・殺虫剤）。
v2932	数値	職業 家具製造	一部相違	1: Yes（欠損=.）	宮城、大阪能勢のみ。宮城は2段階質問形式の下位選択肢（製造 家具）。

v2933	数値	職業 卸売業・小売業	共通	1: Yes (欠損=.)	大阪東成のみ。
v2934	数値	職業 有機溶剤	共通	1: Yes (欠損=.)	宮城のみ。
v2935	数値	職業 放射線・放射性物質	共通	1: Yes (欠損=.)	宮城のみ。
v2936	数値	職業 ベリリウム	共通	1: Yes (欠損=.)	宮城のみ。
v2937	数値	職業 タルク(滑石)	共通	1: Yes (欠損=.)	宮城のみ。
v2938	数値	職業 染料	共通	1: Yes (欠損=.)	宮城のみ。
v2939	数値	職業 なし	共通	1: Yes (欠損=.)	宮城欠損。v2901-v2932 と矛盾ありママ(大阪熊取、愛知)。
v2940	数値	最も長く就いた仕事	一部相違	(欠損=.)	3桁(愛知のみ2桁)コードそのまま読込(0、99、999もママ)。コードの出典は不明。
v2941	数値	最も長く就いた仕事	一部相違	(欠損=.)	大阪能勢のみ。4桁コードそのまま読込。大阪東成はすべて0のため欠損に。コードの出典は不明。
v2950	数値	従事年数	共通	(欠損=.)	0以上はそのまま読込(0、99もママ)。

疫学追跡終了後コホートデータの共通利用(アーカイブ化)の際の死因データ利用に関する検討

研究分担者 大橋 靖雄 中央大学理工学部人間総合理工学科生物統計学
研究協力者 原田亜紀子 東京大学大学院医学系研究科生物統計学

研究要旨

死因情報を付与した形でのデータアーカイブ化が難しい現状を鑑みて、現制度下での運用案を提案した。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。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。

A. 目的

米国では、National Heart, Lung, and Blood Institute (NHLBI) などの公的研究資金で実施した臨床試験や疫学観察研究のデータは、Biologic Specimen and Data Repositories Information Coordinating Center (BioLINCC) に対して追跡終了後一定の期間、条件を充足した後に研究データを登録し公開するよう進められている (<https://biolincc.nhlbi.nih.gov>)。BioLINCCに収集されたデータは、使用申請を行うことにより、研究利用でき、研究データの二次利用により新たな研究や若手研究者の教育・研究機会が創出されている。翻って我が国の現状をみると、厚生

労働科学研究費補助金を受けた研究では、National Bioscience Database Center (NBDC) 等へのデータ提供が求められ、バイオサイエンス基礎研究でのデータアーカイブ化には進展はみられるが、個人の健康情報を含む疫学観察(コホート)研究領域では、未だデータの二次利用環境は整っていない。このように疫学観察研究において、データの共有化、オープンデータ化が進まない背景には、コホート研究が様々な研究費で維持されており、米国NHLBIなどに代表される大型の研究費を受けた先へのデータ提供という単純な構図になりえないこと、多くのコホート研究では主要なアウトカムとなる死因情報を人口動態統計調査の二次

利用申請によって得ており、死因を付与した状態でのデータ公開が難しい点、研究企画時点から、データアーカイブ化を見据えた倫理的諸問題の対応、調査票やデータベース構造の標準化の準備されていないなどが原因として考えられる。そこで本研究では、死因情報を付与した形でのデータアーカイブ化が難しい本邦の現状を鑑みて、現制度下で疫学追跡終了後コホートデータの共通利用を行うための運用案を提案する。

B. 方法

1. 死因情報を外したアーカイブ環境を想定し、必要時に死因を人口動態二次利用申請し、アーカイブセンター（データセンター）にて照合・集計・解析を行う運用

国内大規模循環器疫学研究（JALS）を例に、研究を統括する中央研究事務局が人口動態二次利用申請し、提案する運用例にならって死因を照合することで生じる問題点を考察し、提案例をアーカイブセンターで運用する際に想定される課題を検討した。

1) 死因データを必要時に連結させる方法の提案

JALS では、ローカルコホートから追跡調査データ（生存・死亡（死因はなし）、発症）の提供があり、死因の同定については中央事務局（データセンター）で人口動態の二次利用申請を行い一括で照合作業を行っている。この JALS で採用しているシステムを例にして、研究コンソーシアムに属する研究者が、このデータを利用した研究を計画し、人口動態の二次利用申請で死因情報を得ることによって、センターのスタッフ（必ずしも専門知識を必要としない）により簡便に照合作業が行えるような環境を検討した。

2) 人口動態統計の二次利用申請

Japan Arteriosclerosis Longitudinal Study (JALS) は、国内 58 市町村、8 職域の計 11 万人を追跡している。この JALS コホートのうち、職域コホートを除き、研究開始（2002 年 　ただし一部コホートは 1999 年）から 2012 年 12 月までの異動状況

が、住民基本台帳（住民票）情報により確認されている追跡対象者で期間中に死亡した 7,137 件を死因照合の対象とした。総務省による「統計法第 33 条の運用に関するガイドライン」を参照した上で、厚生労働省大臣官房統計情報部企画課審査解析室に人口動態調査二次利用の申請を行った。申請作業と並行し研究事務局において 2012 年 12 月までの死亡例について、性別、生年月日、死亡日、死亡時の居住地（市町村コード）のリストを作成した。

3) 提案例に従った照合作業

人口動態調査の使用許可がおりた後、性別、生年月日、死亡年月日、死亡時の居住地（市町村コード）を照合変数とし、人口動態調査データを用い原死因を確定した。

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を収集し、疫学追跡終了後コホートデータの共通利用に活用できるかどうか検討を行う

C. 結果

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

1) データ提供からアーカイブ化まで

提案するデータ利用基盤の概略を[図 1]に示した。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（**データ提供**）。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しておき、基本データベースは原則登録時から修正なしの状態、追跡データベースは、今後の追跡継続に応じて更新できる構造とする（**アーカイブ化**）。追跡データは、今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」を含むよう設計した。

2) 死因情報の申請

コンソーシアム内の研究者（あるいは一定の条件を設け、研究グループ外の研究者も可能とするか）

が、このデータベースを使用する研究を計画し、死因情報を得るために厚生労働省に対して人口動態調査二次利用申請を行うとした。(**死因情報申請**)

3) 照合作業から研究データセット作成

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し(**データ照合**)、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(**解析用データセットの作成**)。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という流れである(**利用報告、死因情報の削除**)。死因を付与したデータをアーカイブせず、研究を計画するたびに、この から の作業をアーカイブセンターで手順化し対応するという運用方法の提案である。

4) JALS を例にした上記作業の検証

2015年1月29日付で利用許可があり、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受領した。

提供を受けた人口動態データの読み込み

提供を受けるデータの構造(提供変数)に応じて変更は必要となるが、データの構造と読み込み用の SAS プログラムを作成した。

研究全体の市町村(コード)をリストアップした一覧表を読み込み(SAS プログラム)

追跡調査データベースの対象データから死因照合を行いたい死亡者を抽出する作業(SAS プログラム)

照合作業(SAS プログラム)

上記 と で作成されたデータセットについて、「死亡地(市町村コード)」、「死亡日」、「生年月日」、「性別」でマッチマージしていく。JALS の対象者で、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き、死亡が特定できていたのは7,137件であった。性別、生年月日、死亡年月日、死亡時の居住市町村名を照合変数とし、人口動態調査データと一致がみられたのは7,099件(99.5%)であった。

照合例、未照合例のリスト化(SAS プログラム Excel ファイル)

死因一致例の中に複数の候補例(一意に決まらない例)が存在する対象がないか、また未照合例の情報確認を行うために該当者のリストアップを行う必要がある。これらを SAS から Excel ファイルに出力するプログラムを作成した。複数の候補例が見つかったのは5件あり、いずれも東日本大震災の被災地域での死亡者で、死因が同じであったためその死因で照合を行った。死因が照合できなかった例は33件であり、このうち14件については、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

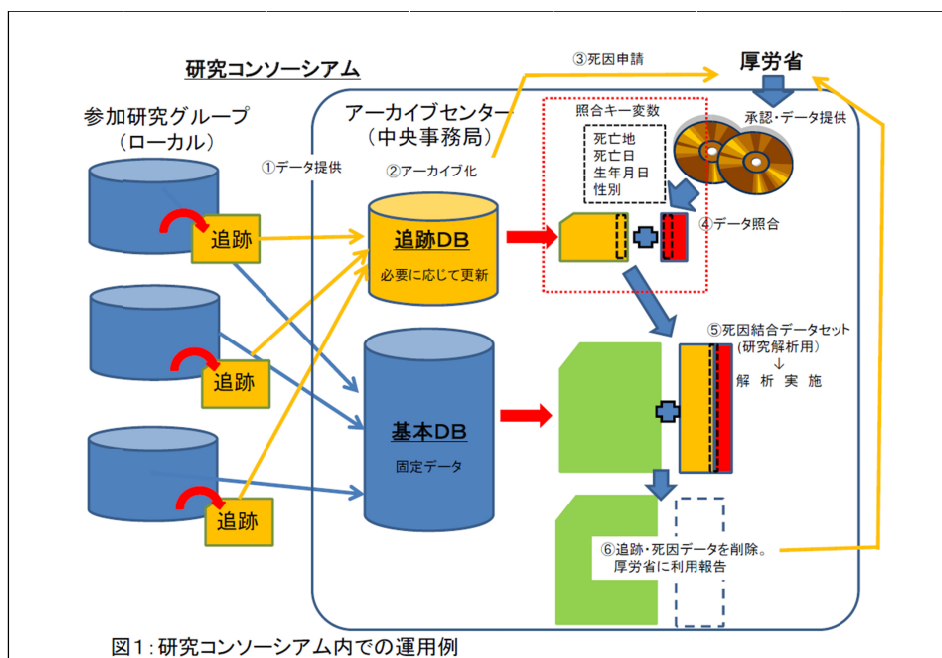


図1: 研究コンソーシアム内での運用例

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のなかった例で、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

1)国内外の分散型ネットワーク先行事例(医療・臨床データ連携(交換)の収集

国内: Standardized Structured Medical record Information eXchange (SS-MIX)

厚生労働省電子的診療情報交換推進事業(Standardized Structured Medical record Information eXchange)で策定された『電子的診療情報を他システムとの交換や地域医療連携で利用するために、診療情報を標準的な形式で蓄積・管理するデータとして保存できる領域』の仕様のことである。SS-MIXでは、これらの医療情報を「標準化ストレージ」というツールに医療情報を標準化した形式で格納・蓄積することにより、複数ベンダ間・複数システム間の相互運用性を高めることを目的としている。標準化ストレージの構造は、コンピュータの一般的なファイル格納形式と同様に階層化されたフォルダのディレクトリ構造を用いている。電子カルテシステムなどではほとんどの業務で個々の患者を軸として診療情報が格納されているため、標準化ストレージにおいても患者にひも付く各種の情報をフォルダの階層構造にルールを決めて格納している。それ以外の項目について研究間での統合をはかるために、拡張ストレージを設け統合することで、標準化ストレージを核にし、施設間連携を構築し、診療情報の研究利用、地域医療連携、災害時連携などを目指している)。

米国: Sentinel Initiative データ交換モデル

米国では2008年より、Food and Drug

Administration Amendments Act (FDA 改革法)によってFDA承認医薬品および医療製品の安全性モニタリングのため、既存のデータベースを用いたアクティブ・サーベイランスであるSentinel Initiativeが実施されることとなった。Sentinel Initiativeではデータインフラと手法の開発を目的として、Harvard Pilgrim Health Care Instituteが中心となって医療データベースを所有する数十の医療機関が協力してMini-Sentinelも実施され、各データストレージは分散したままで、コモンデータモデルに従って共通プログラムで解析するシステムが構築されている。

米国: Health Information Exchange (HIE)

CDX(Crossflo Data Exchange®)ソフトウェアは、Crossflo社が開発したデータ交換ソフトウェアで、HL7、GJXDM(Global Justice XML Data Model)、NIEM(National Information Exchange Model)、EDXL、CAP、NCPDPなど主要医療データシステムおよび国家的なデータ標準、業界標準に準拠したデータ共有方式を実装している。

本システムの医学領域の応用事例としては、モンタナ州および連邦政府の要請を受けて、国立医療情報科学センターの協力のもと、モンタナ州の4つの病院のED(救急科)と州保健福祉局(Montana Department of Public Health and Human Services)とをデータ接続するモンタナ州医療情報交換システムプロジェクト(Montana Health Information Exchange Pilot Project)があげられる。CDXによるシステムで、ほぼリアルタイムでの医療情報の交換を実現し、異種の医療データソースを迅速および効率的に接続する疾病サーベイランスシステムが構築されている。

2)国際共同疫学研究等での分散型ネットワークによるデータ連携例

ViPAR:Virtual Pooling and Analysis of Research Data(Karter KM et al. *Int.J.Epidemiol.* 2015)

ViPARは、International Collaboration for Autism Registryにおいて、国際共同研究を実施す

る6つのサイトデータを統合的に解析することを目的に開発された (Karter KM et al. *Int.J.Epidemiol.* 2015)。多くの医学研究では、倫理的、法的な理由から保有するデータの管理と保持を各サイトにおいて行う必要があるが、本研究では、解析の目的で一時的に「Virtual pooling」サイトに移動することは許可されたことから、一か所にデータを持続的に収集することなしに、物理的に離れたリモートサイトのデータを仮想的(一時的)に集約し分析を可能とするシステムが開発された。ViPAR のリモートサイトとマスターサーバの関連の関連を図4に示したが、ViPAR のマスターサーバは、リモートサイトのPCへ接続されており、研究データはそれぞれリモートサイトで管理されている。利用者はWeb上の analytical portal からアクセスし、データをリモートPCからマスターサーバのRAM上に抽出し、仮想的にデータをプールして解析を行い、終了後は保存することなく消去する。このため外部にデータを収集することなく多施設データを統合した解析が可能となっている。

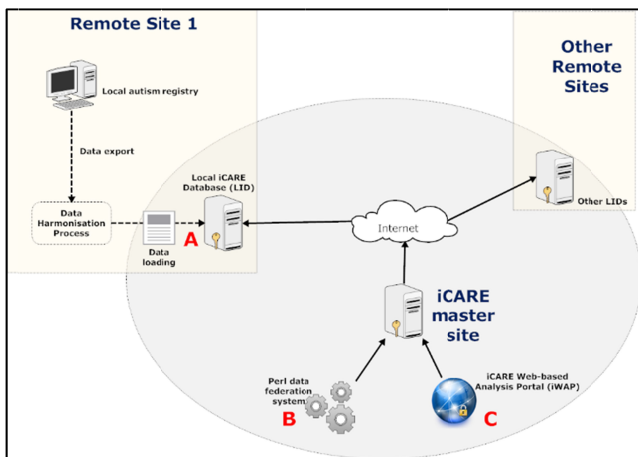


図2 ViPAR のリモートサイトとマスターサーバの関連 (Karter KM et al. *Int.J.Epidemiol.* 2015)

ViPAR の環境のうち、リモートサイトのデータベースは、データ storage として MySQL サーバを、マスターサイトとの通信のため SSH サーバが導入され構成されている(いずれもオープンソース)。実際に解析を行うマスターサーバは、ViPAR daemon と web-based のポータルで構成されている。ViPAR daemon は、ログ出力、統計パッケージのアクセスとコントロール、ローカルサーバから抽出したデータの統合作業を担う。web-based のポータルは、解析、データマネージメント

を行うインターフェイスであり、新規の解析スタート画面、解析アウトプット画面、コードのマネージメント画面の三つからなる。ViPAR を用いた具体的な解析の流れは、以下の通りである。

解析用インターフェイスを開き変数やサイトを選択
統計パッケージの種類 (R, SAS, Stata) を選択し、
テキスト入力部分に解析のための syntax を入力し
サブミットする

リモートデータベースにデータのリクエストが送られ、
マスターサーバの RAM に読み込まれ、バーチャル
プールされる(これらのデータは保存されることは
ない)。バーチャルプールデータセットは、選択
された統計パッケージに読み込まれ、解析が行わ
れる

ファイルマネージャー画面で、解析の進行状況、
完了状況が確認でき、すべての結果とログがダウ
ンロード可能になっている。

ViPAR のプログラム、マニュアル等は下記で公開されている

<http://bioinformatics.childhealthresearch.org.au/software/vipar/>

D. 考察

1 .死因情報を外したアーカイブ環境を想定し、必要時に死因を人口動態二次利用申請し、アーカイブセンター (データセンター)にて照合・集計・解析を行う運用

本検討では、死因情報を外したアーカイブ環境を想定し、必要時に死因を人口動態二次利用申請し、アーカイブセンターにて照合・集計・解析を行う運用例を提案し、国内大規模循環器疫学研究 (JALS) で、実際に運用し死因照合を行った。JALS は各コホート研究 (ローカルサイト) データを中央事務局で統合して実施している研究であることから、この研究で行う死因照合作業の実際が、本検討で想定している研究データアーカイブの環境づくりに対して参考になる点が多いように思われた。死因データの提供を受けてから実施する照合作業については、JALS の研究進捗にあわせて、統計プログラム (SAS)

を用い、データの読み込み、Excel などへの帳票出力に至るまで一連の作業の手順化をすすめてきた。多くの作業をルーチン化してきており、大部分の課題は解決してきているが、本提案例をアーカイブセンターで運用する際に予想される課題について検討する。

1)厚生労働省から受領する統計フォーム(データ形式)に大幅な変更がないことを前提とする

本検討で提案した仕組みは、人口動態二次利用申請により提供される死因データの構造に大幅な変更がないことを前提としている。構造の変更を伴わない小規模の変数追加等の変更であれば、受領後の確認作業を手順化すること(変数表の確認作業やプログラム修正箇所の明示)で、アーカイブセンタースタッフ、疫学研究に従事する研究者等で十分対応できる作業と考えられる。

2)死因照合率はコホートから提出される照合キー変数の精度に依存している

死因照合の一致率については、各コホート研究から提出される照合キー変数(死亡地、性別、死亡年月日、生年月日)の情報精度に依存している。JALSでは、自治体の住民基本台帳などの確認後の情報に基づき照合作業を実施しているが、未照合となる例は、提供情報に何らかの誤りがあり、情報の再確認後に照合可能となる例が大部分である。一方で、提供情報に誤りがないにもかかわらず照合例を検索できない例が少なからず存在しており、これらは性別、生年月日、死亡日について、人口動態統計作成の過程で発生した何らかの入力ミス等が原因として考えられる。しかし、本検討で想定しているのは追跡終了後コホート研究のデータアーカイブ化であるので、これまでの研究実施の過程(各研究での死因照合作業などの過程)でこれらの変数は十分確認されてきており、情報の精度については、進行中の研究に比べて十分高く、こうした問題の発生は少ないと考えられる。

3)死因付与データが使用できる環境が限定される

提案する方法では、死因付与後のデータセットの利用場所がアーカイブセンター(あるいは申請書に記載した研究者の所属する機関)に限られることに

なる。したがって、研究の解析も同様にこの利用場所に限られることになり、データ解析を行える環境(物理的な環境、統計家の配置等)についても検討する必要がある。また、死因データの申請者の所属機関で実施する場合には、アーカイブデータの外部利用の規約等の整備も必要があるといえる。

2.分散型ネットワーク環境を利用し、疫学追跡終了後コホートデータの共通利用が可能かどうか

分散型ネットワークが開発され、利用されている背景には、倫理的規定などからデータ使用の場が制限され物理的に施設外に出せないが、何らかの方法でデータ利用を促進したいというニーズが存在する。疫学研究でのデータ連携では、医療(臨床)データ連携で必要とされるような即時的なデータ連携、多種多様なデータをそのまま連携したいというニーズは高くはない。一方で、以下のような医療データ連携にはない、疫学研究で重視すべき項目が考えられる。

様々な統計解析を柔軟に行えること

医療(臨床)データ連携のように、定型化された集計作業(例:Mini sentinel)が中心ではないので、ViPARのような各種統計パッケージ等と連動して運用可能であるかどうかは重視すべき点であると考えられる。ただし、ViPARについては、RAM上で解析を行い、外部にデータ保存が行われないうりだけであり、施設外にデータが出るといこととの定義や解釈が利用の際の要件となってくる。

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

昨今、施設によっては外部からのアクセスに対して制限を設ける場合も多くなっている点を考慮すべきである。

人口動態統計の二次利用に関連した法令解釈

- ・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の可否
 - ・ViPARでの運用のように、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
- などが問題として考えられる。

文部科学省文化審議会著作権分科会等でも機器

利用時・通信過程における一時的なデータ蓄積については議論されており、「一時的固定（複製）」については、次のように整理されている（コンピュータに該当する部分抜粋）。

ア．瞬時的・過渡的な蓄積であり「複製」ではないもの

- ・処理装置(CPU)の読み込み
- ・ビデオ RAM への書き込み

イ．一時的固定（複製）のうち、「複製」と判断すべきものではないもの

- ・主記憶(RAM)への蓄積
- ・補助記憶のドライブキャッシュ注釈
- ・CPU における 1 次キャッシュ 2 次キャッシュ

ウ．一時的固定（複製）のうち「複製」と判断すべきもの

- ・主記憶(RAM)への蓄積（常時蓄積）

このような解釈は、技術動向を見極めて判断されるものでもあり難しい課題であるといえる。

一施設にデータを集約しない分散型ネットワークによるデータ連携（交換）は、本検討の既存研究データの連携に限ったことではなく、今後新たに疫学研究を行う際にも検討していく必要がある。このような方法を活用していくことで、共同研究の促進や研究のマネジメントの効率化・生産性の向上が期待できるほか、信頼性の高いデータ収集へつながっていくものと考えられる。

E. 結論

死因情報を付与した形でのデータアーカイブ化が難しい本邦の現状を鑑みて、現制度下での運用案を提案した。提案 1 の運用案を複数コホート研究から構成される国内の循環器疫学研究に適用したところ、死因照合率は 99.7%であり、照合作業の技術的側面、作業手順化の面では問題はなく、運用案の一つになりうると考えられた。想定するアーカイブセンター等で実際に運用を行う際には、データ利用の規約、データ利用の物理的環境、サポート体制などについては検討すべき必要があると考えられた。提案 2 については、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先事例な

ことから、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。

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

アーカイブデータ提供のためのガイダンスに含める必要がある項目の整理

研究代表者	玉腰暁子	北海道大学大学院医学研究科
研究分担者	祖父江友孝	大阪大学大学院医学系研究科
研究協力者	辻 一郎	東北大学大学院医学系研究科

研究要旨

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。その5点とは、対象者の個人情報保護、インフォームド・コンセントと倫理審査、知的財産権の帰属、寄託する項目/しない項目の判断、データ提供先の制限、である。

A. 目的

疾病の発生関連要因の検討を目的としたコホート研究では、生活習慣の詳細や既往歴等の情報をベースに、その後の疾病発生や死亡状況等を追跡し、情報を蓄積する。したがって、追跡を終了したコホート研究データのアーカイブ化を進めるにあたっては、利用を促進すると同時に倫理面や研究参加者の個人情報保護に配慮した適切なデータ提供が求められる。しかし、研究開始時点のデータ取り扱いや対象者への説明同意のあり方は、現在従うことが求められている人を対象とした医学系研究に関する倫理指針で定めるものとは異なっている。そこで、今後利活用を進めることを念頭に、データ提供の際に従うガイダンスに含める必要があると考えられる項目の整理を試みた。

B. 方法

昨年度までの研究成果、ならびに、国内のヒトデータベース(ライフサイエンス分野)の収集・公開を担うバイオサイエンスデータベースセンター(NBDC)のガイドライン、社会科学分野のデータアーカイブの運営を行っている東京大学社会

科学研究所附属社会調査・データアーカイブ研究センターSSJDAのデータ寄託手引きを参考に、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。

(倫理面への配慮)

特に必要なし。

C. 結果

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。

対象者の個人情報保護: SSJDAでは回答者のプライバシー保護に最大限の注意が払われており、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している。我々の実データを用いた過去2年間の検討でも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しており、この点に関してデータ寄託者、アーカイブ運営者

ともに十分に認識する必要がある。

インフォームド・コンセントと倫理審査：
NBDC 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を義務付けている。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。また、コホートによっては、地域との関係が現在まで維持されていないことも考えられる。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。

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。試料・情報提供者と直接提供を受ける研究者の間であれば、知的財産権は研究者（研究機関）に帰属することとなっており、最近では説明事項に含まれることも多い。一方、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。

寄託する項目/しない項目の判断：で述べたような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。

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。

D. 考察

米国に比べ日本では、未だ疫学研究データの二次料体制が整っているとはいえない。しかし、日本においても社会科学系では既に多くの調査データがアーカイブ化され、利用されていることから、その促進のための教育や環境整備が望まれる。一方で、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。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。今後そのような研究の寄託を進める際の指標になるものと考えられる。

E. 結論

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。その5点とは、対象者の個人情報保護、インフォームド・コンセントと倫理審査、知的財産権の帰属、寄託する項目/しない項目の判断、データ提供先の制限、である。

F. 健康機器情報

特になし

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

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統計法解釈に関する検討

研究代表者	玉腰暁子	北海道大学大学院医学研究科
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	中央大学理工学部人間総合理工学科
研究分担者	祖父江友孝	大阪大学大学院医学系研究科
研究協力者	辻 一郎	東北大学大学院医学系研究科
研究協力者	友岡史仁	日本大学法学部

研究要旨

現在までのところ、多くのコホート研究では統計法の規定に則り人口動態統計資料から死因情報を得ているため、死因をつけた形でのデータアーカイブの公開利用はできないと考えられる。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。

A. 目的

通常のコホート研究では、追跡情報として重要な死因情報を人口動態統計資料(死亡個票)から、定められた手続きを経て入手する。この人口動態統計資料は、統計法に基づき作成されるところ、同法40条では、同法に特別の定めがある場合を除き行政機関の長等が行った統計調査の目的以外の目的のために、当該統計調査に係る調査票情報を自ら利用したり提供してはならないとされていることから、研究目的での提供は同法第三章に規定される一定の条件の下で行われる。したがって、我々研究者が死因情報の提供を受ける多くの場合は、同法33条2号の規定が適用されるところ、入手にあたっては、公的研究費を受け続け

ていることなど、以下の様な一定の要件が課せられることになる。

平成23年6月厚生労働省「調査票情報の提供に関する利用申出手引」における

- イ 公的機関から委託を受け、又は公的機関と共同して行う調査研究に係る統計の作成等、
- ロ その実施に要する費用の全部又は一部を公的機関が公募の方法により補助して行う調査研究に係る統計の作成等、
- ハ 行政機関の長又は地方公共団体の長その他の執行機関が、その政策の企画、立案等に有用であると認める統計の作成等。

しかし、このような要件は、追跡が終わったコホート

研究にとって決して低いハードルではない。総じて研究者は、複数の研究を実施しているものであること、競争的資金の場合は確実に研究費が得られるとは限らないことなどが、この理由として挙げられる。

さらに、統計法43条2項の規定によれば、同法に基づき提供を受けた情報(コホート研究では死因、死亡の事実と死因年月日は通常別の情報源から入手される)について、「当該調査票情報又は当該匿名データをその提供を受けた目的以外の目的のために自ら利用し、又は提供してはならない」とされているため、対象者の個人情報秘匿した上でもデータの二次利用への提供は認められていないと解される。

以上であって、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死因説明に具体的に貢献しうることを考え合わせれば、上記に掲げた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。このため、人口動態統計資料(死因個票)の利用を可能にし得る現実的な考え方について、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、意見交換を行った。

B. 方法

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、検討を行った。

(倫理面への配慮)

特になし

C. 結果

統計法に基づく調査は機関統計調査として義務的に行われ、その目的は公的統計の作成に限定されているため、当該調査を「研究目的」といった同法の目的外利用に供する場合は厳格に制限しておこうというのが趣旨と解すべきことになる。そ

の場合、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も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ずる。しかし、確かに、統計法の限界がある一方、同法以外に個別に死因情報を取得することの困難さ、今までに行われているほとんど全てのコホート研究で同法に則り死因情報が取得されてきた現状に照らし、当該情報を長期的かつ利用者の範囲を広げてコホート研究に有効活用する道を何らかの形で法制度上担保する必要性を論じておく意義は、十分ににある。

この場合の論点として、統計法の解釈運用を変更することで足りるのか、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。

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。対しての場合は、

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分、個人情報保護に照らした保護に係る必要なスタンダードを充足する詳細な議論を要する。または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視点からすれば、現時点ではの選択肢が方向性として現実的と考えられる。

D. 考察

今回の論点は、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しようとするものである。統計法の従来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、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。

E. 結論

現在までのところ、多くのコホート研究では統計法の規定に則り人口動態統計資料から死因情報を得ているため、死因をつけた形でのデータアーカイブの公開利用はできないと考えられる。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。

F. 健康機器情報

特になし

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

がん登録推進法とコホート研究との関係に関する研究

研究分担者 辻 一郎 (東北大学大学院医学系研究科公衆衛生学・教授)

研究要旨

がん登録等の推進に関する法律の条文や政令・省令をもとに、追跡終了後コホート研究を用いたデータアーカイブが構築された場合における、がん登録情報などの活用をめぐる諸問題について検討を行った。本研究班に含まれるコホート研究は、附則第2条(経過措置)により、「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」を講じたうえで、従来通り、研究対象者におけるがん罹患情報を得ることが可能であると思われる。

一方、がん登録推進法と政令・省令により定められた要件・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については、法令上のさらなる検討が必要であると思われる。

A. 目的

平成28年1月から「がん登録等の推進に関する法律(以下、「がん登録推進法」)」が施行された。がん登録推進法は、がん登録情報の調査研究への利活用とがん登録等に係る個人情報の保護の2つを共に重視している。

本研究では、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、がん登録情報の利用の可能性について3年かけて検討を行った。

B. 方法

がん登録推進法の条文及びがん登録推進法の実施に関する政令・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。

(倫理面への配慮)

特になし

C. 結果

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、法施行日(平成28年1月1日)の前に開始されたがんに係る調査研究については、附則第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、政令第三百二十三号(平成27年9月9日制定)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、厚生労働省告示第四百七十一号(平成27年12月15日公表)で定められた。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

がん登録推進法の条文、厚生科学審議会がん登録部会での議論と政令案などについて検討した。

本研究班に含まれるコホート研究は、いずれも研究対象者が五千人以上であり、ベースライン調査から相当期間が経過しているために全ての対象者と連絡を取るとは困難な状況にある。また、同意代替措置についても第1項は満たしているため、第2項と第3項を適切に講ずることにより、従来通り、研究対象者におけるがん罹患情報を取得することが可能であると思われる。

一方、いくつかのコホート研究をもとにデータアーカイブを作成し、研究班以外の者にも公開して利用を

認める場合、データベース情報を匿名化する必要が生じる。その場合、データアーカイブが作成されて以降の、がん登録情報との照合は不可能となる。そのため、データアーカイブが作成された時点で追跡は終了とすることについての合意が必要となる。

一方、がん登録推進法と政令・省令により定められた要件・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、がん登録推進法と政令・省令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

がん登録等の推進に関する法律の条文や政令・省令をもとに、追跡終了後コホート研究を用いたデータアーカイブが構築された場合における、がん登録情報などの活用をめぐる諸問題について検討を行った。本研究班に含まれるコホート研究は、附則第2条(経過措置)により、「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」を講じたうえで、従来通り、研究対象者におけるがん罹患情報を得ることが可能であると思われる。

一方、がん登録推進法と政令・省令により定められた要件・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については、法令上のさらなる検討が必要であると思われる。

F. 健康機器情報

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

個人情報保護法改正に伴う疫学研究への影響に関する検討

研究代表者	玉腰暁子	北海道大学大学院医学研究科
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	祖父江友孝	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
研究協力者	三浦克之	滋賀医科大学大学院医学系研究科

研究要旨

改正された個人情報保護法において、病歴が要配慮個人情報に位置づけられたことから、疫学研究において生じうる影響を検討した。その結果、追跡情報の入手を医療機関等から受けている場合に、過剰反応により情報が提供されないことが懸念された。

A. 目的

個人情報保護法が2015年9月に改正され、2年以内に施行されることとなった。この改正では、個人情報の定義を明確化することによりグレーゾーンを解決し、また誰の情報かわからないように加工された匿名加工情報について自由な利活用を可能とし、経済を活性化することが図られる一方、不正な提供には罰則が設けられ不正な個人情報流通を抑止する措置がとられた。さらに、その取り扱いによって差別や偏見、その他の不利益が生じるおそれがあるため、特に慎重な取り扱いが求められる個人情報として「要配慮個人情報」が定められたが、この中に本人の人種、信条、社会的身分、犯罪の経歴、犯罪により害を被った事実とともに病歴が加えられた。要配慮個人情報では、取得にあたっては原則として本人の同意を得ること、オプトアウト手続きによる第三者提供が認められないこととされている。ただし、関連性を有する範囲内で利用目的を変更したり、匿名加工情報に加工して第三者に提供したりすることは可能である。

喫煙や食事、運動などに代表される生活習慣や病歴等と人々の健康との関連を明らかにし、国民の健康的な生活を守るための対策樹立に資する(実例:健康日本21、喫煙対策、日本人のためのがん予防法など)ことを目的に、多くの疫学研究、特にコホート研究が実施されてきた。研究開始時に収集される情報に病歴が含まれることが少なくなく、また追跡情報では死因のみならず、病院等の協力により疾病罹患情報が用いられる。これから開始される研究はもちろん、2000年代に入って指針や個人情報保護法の施行以降に開始された研究では、対象者への説明・同意時に取得・提供項目に関し配慮が行き届いていると考えられるが、それ以前の(現在の施策に活かされる成果を輩出してきた)コホート研究では、必ずしも十分な対応がなされていたとは言い難い面がある。そこで、今回、病歴が要配慮個人情報に位置づけられたことにより、以前から行われているコホート研究において生じうる影響、さらに倫理指針の例外規定が正しく認知されていないために疫学研究全般に対し起こりうる事

例につき検討した。

B. 方法

個人情報保護法の条文、特に要配慮個人情報に関する規定を確認するとともに、疫学会の関係者から各々の研究の現在の運用等に関する情報を得るとともに議論を行った。

(倫理面への配慮)

特になし

C. 結果

【以前から行われているコホート研究】

指針等の整備前から開始されたコホート研究であっても、開始以降に倫理審査を受けるなど、研究者は可能な限り対象者の個人情報保護に配慮して研究を行っていた。また、研究開始時に得ている病歴に関しては、本人の自己申告によるものが多く、また医療機関や健診機関から得ている場合でも、その利用に関しては何らかの説明が行われていた。しかし、対象者が多いこと、開始時期が古く対象者が研究に参加していることを記憶していない場合もあること、遡っての同意手続きによって対象者の偏りが懸念されることなどにより、開始時点の説明が現在の基準に照らした場合に不十分であっても改めての同意手続き等を経ることなく、病院から追跡情報として疾病情報を受けている研究も存在した。したがって、今回、病歴が第三者提供に際して対象者からの同意が必須である要配慮個人情報と位置づけられたことから、医療機関等の過剰反応によりコホート研究に重要な追跡情報を得る際に支障を来すことが懸念された。また、仮に、改めて再同意手続きをとり同意が得られた者のみを観察対象とした場合には、対象者の偏りを生じることから研究の価値を著しく損ね、ひいては公衆衛生上の不利益につながることを危惧された。

一方、本研究班で検討している追跡終了後のコホート研究をアーカイブ化し、二次利用に用いる際には、いずれせよ対象者を遡れないような匿名化措置が必要と考えられることから、既に収集している病歴をあわ

せて提供することは差し支えないと考えられた(死因に関しては、統計法の規定によるため、現状では提供不可)。ただし、非常に稀な病気の場合等は、詳細な病名ではなく、ある程度まとめた疾患概念として示すなどの配慮を行うことが必要であろう。

【疾病登録による罹患率等の把握のための疫学研究】

学会や研究者が主体となって地域の医療機関等から医療情報を収集・整理し、循環器疾患(脳血管疾患、心疾患)や腎透析、難治性疾患、新興感染症など、今後予防法や治療法の解明を進めなければならない多くの病気の罹患率、予後、地域差等の実態を明らかにするための疾病登録研究がある(がんはがん登録法が平成28年より施行されるため除く)。このような研究は予防法の解明や医療資源の投入量の判断などに必要である。しかし、同意した患者のみを対象とする研究では正確な罹患率や予後等の把握ができずに誤った解釈につながり、公衆衛生上のリスクを生じることになる。したがって例外規定に該当するものとして研究を実施する必要があるが、医療機関等が要配慮個人情報の「病歴」として本人同意のない情報提供を拒否した場合、このような研究を実施することができなくなる。また、この研究においては、患者の予後の実態を明らかにするため住民基本台帳や死亡小票、人口動態統計データを行政機関で閲覧・照合する必要があるが、この情報提供が拒否されると研究の実施が困難になる。また、新興感染症の流行、化学物質や放射線の曝露を受けるような事故等があった場合、地域別の疾病動向は同意の有無の関わらず全数調査されるべきであり、同意を前提とすると誤判断につながり、社会的な危機につながる。

【特定の疾患の患者の予後及び予後要因を明らかにする疫学研究(臨床研究)】

特定の疾患の患者の予後および予後要因を明らかにするための疫学研究(臨床研究)が、多数の医療機関が共同して実施されることがある。この場合、地域での罹患率を明らかにすることは必ずしも目的にしているが、予後の実態と予後規定因子や治療効果

を明らかにする上で必要不可欠な研究である。多くの場合、連結可能匿名化により医療機関から情報提供を受けており、医療機関でのオプトアウトで対応されているが、本人同意が必要として提供を拒否された場合、研究を実施することができなくなる。特に、本人の同意が困難な難治性疾患や認知症、精神疾患などにおいて予後や予後要因を明らかにする研究ができなくなる。

【複数の既存資料を個人情報キーとして突合する疫学研究】

疫学研究では本人に新たな侵襲を伴うことなく、保健事業等で収集された既存資料を突合して行う研究が実施される。例えば、検診の効果を検証して、がん検診や循環器健診の質を向上させることを目的とした研究で、検診受診者と非受診者の死亡率等を比較する研究は、検診の改良や効果判定を行うために必須であり、公衆衛生上の意義が大きい。この場合、受診者からは同意が取れるが、非受診者からは同意が取れないので、同意を前提とすると研究の価値が大きく毀損され、検診そのものの評価が歪められる。同様の例としては、母子健診事業として行われる乳幼児健診とその後の学校健診の結果、あるいは職域健診と退職後に国保で把握されるレセプトなど、異なる機関で収集される健康情報や病歴を結び付ける研究の場合、やはり同意した人に限定するとあやまった情勢判断をしてしまう危険性がある。

【構成員全員の健康増進対策を検証・評価する研究】

公衆衛生では集団の構成員全員の健康増進対策をどのように行うかを検証・評価する研究が必要である。そしてその対策が「全員」を対象としていて「同意の取れた者」を対象としていない場合、同意取得者のみで検証を行うと施策全体を誤る可能性がある。これは健康日本 21、タバコ対策、飲酒問題への対策など国レベルのものから、都道府県健康増進計画、特定健診等実施計画(市町村)、職場での健康管理・作業管理など様々なレベルで公衆衛生の向上のために必要である。このような研究は将来の制度改善のための提案として必要であり、公的機関が主体の場合だけ

でなく、研究者側の提案としての研究も必要である。

【定期健康診断のデータを用いた疫学研究】

地域や職域における定期健康診断(特定健診等)や人間ドックから得られたデータは、健康診断自体の評価の他、疫学研究に活用されてきた。本データは疾病治療状況の実態把握のほか、コホート研究であればベースライン時点での当該疾病の除外、追跡中の疾病罹患の把握、あるいは曝露要因や交絡要因としての考慮など疫学研究の根幹に関わる重要な項目である。行政的あるいは保健医療サービスとして蓄積されている、あるいは将来蓄積されるデータで、研究としての個別同意は通常は取られないケースがあり、本データの使用に本人同意が必要となると研究が大幅に制限されるばかりでなく、疫学研究としての科学性を保つことが困難になる。例外規定に該当するものとして研究を実施する必要があるが、健康診断実施機関、保険者等が要配慮個人情報の「病歴」として本人同意のない情報提供を拒否した場合、本研究を実施することができなくなる。

【「病歴」の範囲が明確でないために起こりうる事例】

個人の健康状態や疾病に関する情報には下に示すように様々な種類がある。現状では「病歴」の範囲が明確でないために、どのような場合に要配慮個人情報にあたるのか混乱が生じ、疫学研究者が情報提供を得られなくなることが懸念された。

- ・ 医師が記載した診断書病名
- ・ レセプト(診療報酬請求)に記載されたレセプト病名、疾患別医療費等
- ・ 診断のために行われた画像データ
- ・ 疾患特異的なマーカーの検査値
- ・ 死亡小票・死亡診断書の死因情報
- ・ 市町村への住民票請求による生死の確認

D. 考察

改正された個人情報保護法において、病歴が要配慮個人情報に位置づけられたことから、以前から行われているコホート研究において生じうる影響を検討した。その結果、追跡情報の入手を医療機関等から受

けている場合に、過剰反応により情報が提供されないことが懸念された。またその他にもいくつかの研究事例で、病歴が要配慮個人情報となったことにより影響を受けることが懸念された。そこで、研究者が適切な個人情報保護体制を整えることはもちろんであるが、その下で、これまでと同様に利用が可能になるようガイダンスの整備等を進めることが重要と考えられる。今までも追跡のために対象者の異動情報を自治体から得る際、ごく一部の自治体からではあるが提供されなかったという事例が生じていることから、個人情報保護法改正を受けて行われる指針やガイダンス改正においてはこの点に十分配慮したものとなるよう働きかけることが重要である。

E. 結論

改正された個人情報保護法において、病歴が要配慮個人情報に位置づけられたことから、以前から行われているコホート研究において生じうる影響を検討した。その結果、追跡情報の入手を医療機関等から受けている場合に、過剰反応により情報が提供されないことが懸念された。

F. 健康機器情報

特になし

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

研究成果の刊行に関する一覧表

なし