

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 26 年度 総括・分担研究報告書

研究代表者 玉腰 暁子

平成 27 (2015) 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：進捗報告..... | 1 |
| 玉腰暁子 | |

II. 分担研究報告

| | |
|---|----|
| 社会科学系データアーカイブの現状 SSJDA から | 7 |
| 磯博康、大橋靖雄、他 | |
| NBDC ヒトデータベースにおけるヒト関連研究のデータ共有 | 10 |
| 玉腰暁子、祖父江友孝、他 | |
| 大規模コホートデータにおける一意性の検討 | 14 |
| 祖父江友孝 | |
| がん登録推進法における死亡情報活用に関する研究 | 23 |
| 辻一郎 | |
| 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 | |

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、データを広く共有化し二次利用を促進するためのシステムであるデータアーカイブ化が進められている社会科学系ならびにライフサイエンス系のデータアーカイブセンターの現状を把握した。疫学研究データには機微情報を含むのみならず、収集する項目数も多いことから、80%程度のレコードは一意性があるものとして対応することが必要であり、研究開始時点での対象者への説明のあり方、完全に連結不可能匿名化するタイミングやその方法などの検討とともに、研究内容によっては共同研究を締結して行うなどの対応が必要と考えられた。また、今後、疫学研究のデータアーカイブを構築し利用を進めていくためには、教育も含めた利用環境整備も必要である。一方、統計法の規定から、現状では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利用することはできない。そのため、がん登録推進法による死亡者情報票の活用、ならびに現制度化で運用するために必要に応じて死因情報を入手・利用する方法を提案した。

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。今年度は、国内における社会学分野ならびにライフサイエンス分野のデータアーカイブ

の現状、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検討した。さらに、人口動態統計情報を利用して把握されている死因情報は、統計法の規定により、データアーカイブ時には制限つきであれ公開情報とできないことから、その対応として代替案として、がん登録推進法において示された死亡者情報票の活用の提案、ならびに現制度化での利用方法の提案を行った。

B. 研究方法

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アーカイブを公開する場合の課題を検討した。

三府県コホートデータを用い、分類数、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。

がん登録推進法について、条文や政令をもとに、
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、死亡者情報活用を
めぐる諸問題について検討を行った。

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。

C. 研究結果

-1 社会学分野のデータアーカイブの現状

データアーカイブセンターの意義は、統計調査、
社会調査の個票データを収集・保管し、その散逸を
防ぐとともに、学術目的での二次的な利用のために
提供することにある。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。データアーカイブを二次利用
するメリットは、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
できること、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
イディ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、若手研究者にとっては、自身で小規
模な回収率の高くない調査を行うことに比べ、質の
よい調査データにアクセスできること、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。

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、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。

-2 ライフサイエンス分野のデータアーカイブ の現状

バイオサイエンスデータベースセンター (NBDC)
では特にヒトに関する情報に特化した NBDC ヒトデ
ータベースを構築し、2013 年 10 月から運用が開始
された。ヒトを対象とするデータであるため、特に
個人につながる情報の保護対策が重要となる。そこ
で NBDC では、欧米のデータベースを参考に受け皿
づくりが進められている。取り扱われるデータは匿
名化されたもののみで、レベルに応じたアクセス制
限が行われている。データ提供と利用に関する審査
は、NBDC で行われるが、原則として試料提供者か
らデータ共有に関する同意を事前に受けておくこ
とが求められている。ただし、過去に収集された既
存試料・情報で同意を取り直すことが困難な場合に
は、データ共有について倫理委員会で承認されてい
ることが要件である。データ共有に関するガイドラ
イン、セキュリティレベルに関するガイドライン等
が定められ、HP 上で公開されている。2015 年 2 月
現在、研究データ 15 件が HP に公開され、そのうち
制限公開 10 件、オープン 5 件であるが、すべてゲ
ノムに関連するもので、いわゆる疫学研究のデー
タは今までのところ、寄託はされていない。

一意性の検討

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。

がん登録推進法における死亡者情報票の活用

死亡者情報票を利用したコホートデータ追跡情報入手法を検討するため、癌登録推進法における死亡者情報票の取り扱いについて確認した。がん登録推進法第11条(死亡者情報票の作成及び提出)に死亡者情報票の作成に関する事項は明記されている。それによると、死亡者情報票は、死亡の届書その他の関係書類に基づいて、市町村長が作成するもので、死亡した者に関する氏名、性別、生年月日、死亡の時にける住所、死亡の日、死亡の原因、死亡診断書の作成に係る病院又は診療所の名称及び所在地その他の厚生労働省令で定める情報が含まれる。全死亡者に関する死亡者情報票は、電磁的記録又は書類により作成され、保健所、都道府県を経て、国(国立がん研究センター)に提出される。それを受けて、国立がん研究センターは、死亡者情報票と全国がん登録情報とを照合する。その照合期間は、厚生科学審議会がん登録部会の政令案では100年とされたことから、がん患者の生命予後をほぼ完璧に追跡することが可能になったと思われる。

データアーカイブ化における死因情報の利用：コホートデータに死因情報を付与する方法

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

開利用することは不可能な状態にある。そこで、代替案の一つとして、死因を連結した形でのデータセット構築・アーカイブ化ではなく、必要時に中央(アーカイブデータを保持するセンター等)で死因照合作業を行って解析用データセットを作成する方法とその妥当性をJALSの実データを用いて検討した。

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。

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。実際、社会科学系分野では、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。ライフサイエンス分野では、NBDCがデータアーカイブセンターの役割を担い始めた。しかし、いまだ緒についたところで、特にヒトを対象としたデータに関して実績があがるのはまだこれからと考えられた。加えて、疫学研究は単に生体情報のみならず、生活習慣や心情等に関する情報も収集されることが多く、追跡結果も死亡・疾病罹患など機微情報を含む。さらに、収集する項目数も多いため、その組み合わせにより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。したがって、研究開始時点での対象者への説明のあり方、完全に連結不可能匿名化にするタイミングやその方法など、今後の検討課題である。また、単に塩基配列などの公開と異なり、疫学研究で構築されたデータセットの公開内容は一部に制限し、機微情報を取り扱う場合には共同研究を締結するなどの対応が必要と思われる。

統計法の規定から、現状では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利用することはできない。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。がん登録推進法により提供される死亡者情報票データを活用して、米国のNDIと同様のシステムを作るには法制度の改革が必要であるが、それが実現すれば、わが国の疫学研究・臨床研究や医薬品・医療機器開発は発展すると思われ、それは政府「健康・医療戦略」の目指すところと合致するものであろう。

一方で、現制度化での運用方法を検討するため、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た。JALSで実際に死因照合を行ったところ99.5%で照合が可能であり、照合作業の技術的側面、作業手順化の面で問題はなかった。今後検討すべき課題の一つとしては、人口動態統計二次利用申請に基づくことから、データの保持期間が公的研究費の継続期間に限定されることがあげられる。研究の質や結果の再現性を保証するという点では、解析に使用した死因付きのデータセットが長期に保持できることが望ましい。また、今回提案する方法では、死因付きデータセットの利用場所、すなわち解析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。そのため、アーカイブデータの利用規定も合わせて、アーカイブセンターで対応する場合は、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。

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、社会科学系ならびにライフサイエンス系のデータアーカイブセンターの現状を把握した。疫学研究データには機微情報を含むのみならず、収集する項目数も多いことから、80%程度のレコードは一意性があるものとして対応することが必要である。研究開始時点での対象者への説明のあり方、完全に連結不可能匿名化にするタイミングやその方法などの検討とともに、共同研究を締結するなどの対応が必要と思われる。また、疫学研究のデータアーカイブを構築していくためには、先行する社会科学系データアーカイブの運営システムから学ぶと同時に、利用のための環境整備も必要と考えられた。一方、統計法の規定から、現状では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利用することはできない。そのため、がん登録推進法による死亡者情報票の活用、ならびに現制度化で運用するために必要に応じて死因情報を入手・利用する方法を提案

した。

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

社会科学系データアーカイブの現状 SSJDA から

| | | |
|-------|-------|------------------|
| 研究分担者 | 磯 博康 | 大阪大学大学院医学系研究科 |
| | 大橋康雄 | 中央大学理工学部人間総合理工学科 |
| | 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 | 藤原 翔 | 東京大学社会科学研究所 |

研究要旨

データアーカイブ化に関し、社会科学系分野の現状と疫学研究に応用する場合の課題を把握・整理した。社会科学系分野では、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。その促進のために二次利用教育も行われており、疫学研究のデータアーカイブを構築していくためには、社会科学系データアーカイブの運営システムから学ぶと同時に、利用のための環境整備も必要である。さらに、多くの社会科学系の調査データは、調査時点で個人情報を収集しないことも多く、対象者を追跡することが前提であるコホート研究とは、そもそもの配慮事項が異なる部分もある。特に、研究開始時点での対象者への説明のあり方、完全に連結不可能匿名化にするタイミングなど、今後の検討課題と考えられる。

A. 目的

社会科学系分野で進んでいるデータアーカイブ化の現状を把握し、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参考とする。

B. 方法

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. 結果

[社会科学系データ]

社会調査とは、社会現象について知るために、一定

の科学的方法を用いて、現地においてデータの収集・分析をする過程と定義される。政治学、経済学、経営学、社会学、教育学、法律、法学調査等がその範囲となり、調査目的により、家庭、職業、教育、収入、疾病等、様々な変数が含まれる。

社会科学系調査では、調査時点で個人情報を収集しないことも多い。そのため、厳密には個人情報には当たらないと考えられるが、特殊な状況下では地域・属性などが分かる事で対象者が絞り込まれる(例えば、ある高校のある年時の卒業生で身長が 190cm、ある地域で子どもが 6 人、など)こともあり、情報の取り扱いには注意が求められる。

[社会科学系データアーカイブの意義]

社会科学系では量的、質的データのいずれもデー

データアーカイブに寄託されており、二次利用者が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ということがより一般的になってきている。その背景には、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。したがって、データアーカイブセンターの意義は、統計調査、社会調査の個票データを収集・保管し、その散逸を防ぐとともに、学術目的での二次的な利用のために提供することにある。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により、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。

一方で、データアーカイブを二次利用するメリットはいくつか挙げられる。第一に、既に行われている調査を繰り返さずに済むため、労力、資金とも無駄な投入を避けることができる。第二に、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。第三に、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、質のよい調査データにアクセスできる。さらに、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。

[データアーカイブセンターの沿革と運営]

1996年5月、東京大学社会科学研究所の附属施設として、日本社会研究情報センターが設立された。1998年4月から、センターではSSJデータアーカイブ(Social Science Japan Data Archive)を運営し、統計調査、社会調査の調査個票データと調査方法等に関する情報を収集・保管し、学術目的での二次分析のために、学内外の教員、大学院生等に提供を開始した。同時に、二次分析普及のために研究会や計量的研究法に関するセミナーも実施された。2009年4月に社会調査・データアーカイブ研究センターと改組

され、現在にいたっている。

データアーカイブ二次利用促進と円滑化のために、さらに次のような活動も順次行われている。二次分析優秀論文表彰(2005年6月開始)、リモート集計システムの稼働(2005年10月開始)、SSJDA Direct(データダウンロードシステム)の稼働(2009年4月開始)、寄託者表彰(2010年2月開始)、NESSTARシステム(メタデータ閲覧・オンライン分析システム)の稼働(2014年1月開始)、SSJDA Direct(データダウンロードシステム)への完全移行(2014年2月)。

現在、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る。

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。

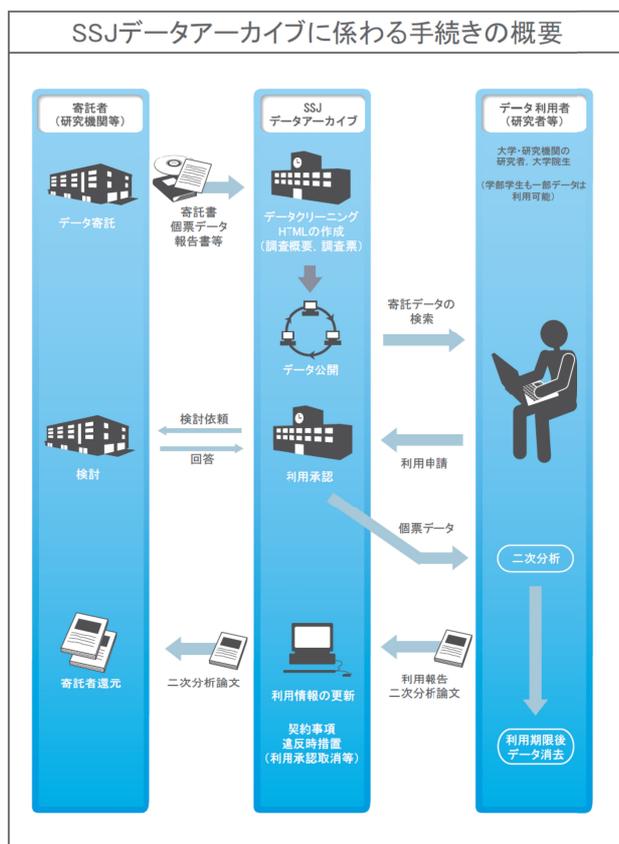
[他の社会科学系データアーカイブ]

国際的には、Roper Center(アメリカ)、ICPSR(アメリカミシガン大学)、GESIS(ドイツ)、UKDA(イギリス)等があげられる。国内では、RUDA(Rikkyo University Data Archive)、SORD(Social and Opinion Research Database、札幌学院大学、社会調査)、KUMA(神戸大学マイクロデータアーカイブ、公的統計)、Hi-Stat Social Science Database Network(一橋大学経済研究所、公的統計)、レヴァイアサン・データバンク(政治学系データ)等がある。

[データアーカイブ利活用の現状]

現在、SSJDAでは次の図のような仕組みで手続きが行われている

(<http://ssjda.iss.u-tokyo.ac.jp/2008image.pdf>)。



SSJDAには現在、約1600のデータセットが公開されており、2013年度は2700件の利用があった。

提供にあたっては、調査対象者個人が特定できるような情報は公開されないようセンターが秘匿処理を行っている。また、利用は学術目的のみで、利用者も研究者、大学院生、教員の指導を受けた大学生に限定される。そのうえで、利用者には、個々の回答者等が識別できる形式では発表しないことの誓約が義務付けられている。多くのデータの利用期間は1年(延長申請可能)であり、利用期限後は、個票データ消去が求められている。

D. 考察

社会科学系分野では、既に調査データのアーカイブ化とその利用の仕組みが整っており、多くの人々の協力により、うまく活用されているようである。しかし、データ寄託がある一定数に達し、利用のメリットが十分に浸透しなくては、労力のみが大きくなる懸念がある。したがって、疫学研究のデータアーカイブを構築していくためには、社会科学系データアーカイブの運

営システムから学ぶと同時に、利用のための環境整備も必要である。さらに、多くの社会科学系の調査データは、調査時点で個人情報収集しないことも多く、対象者を追跡することが前提であるコホート研究とは、そもそもの配慮事項が異なる部分もある。特に、研究開始時点で対象者への説明のあり方、完全に連結不可能匿名化するタイミングなど、今後の検討課題と考えられる。

E. 結論

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により運営されているSSJDAを例に、社会科学系分野のデータアーカイブ化の現状を把握し、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

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

NBDC ヒトデータベースにおけるヒト関連研究のデータ共有

| | | |
|-------|-------|--------------------------------|
| 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 | 大橋康雄 | 中央大学理工学部人間総合理工学科 |
| | 祖父江友孝 | 大阪大学大学院医学系研究科 |
| 研究協力者 | 箕輪真理 | 科学技術振興機構 バイオサイエンスデータベースセンター |

研究要旨

国内で進められているライフサイエンス分野のデータベース統合化とヒトデータベースの現状について把握した。データ共有にかかる労力は大きい、その意義もまた大きいものと思われる。しかし、いまだ緒についたばかりであり、特にヒトを対象としたデータに関して実績があがるのはまだこれからと考えられた。疫学研究データについても受け入れは可能であるが、機微情報に関しては共同研究などを締結する必要があると考えられた。

A. 目的

ライフサイエンス分野の公的科学研究費の公募要項には、現在、論文発表等で公表された成果に関する生データの複製物、また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 (NBDC) に提供するように求められている。そこで、このようにライフサイエンス分野のデータ統合を目的に進められている NBDC の現状を把握し、疫学研究のデータアーカイブ化を進めるための参考とする。

B. 方法

NBDC の箕輪真理氏より、NBDC が収集しているデータベースの現状につき情報提供をいただいた。

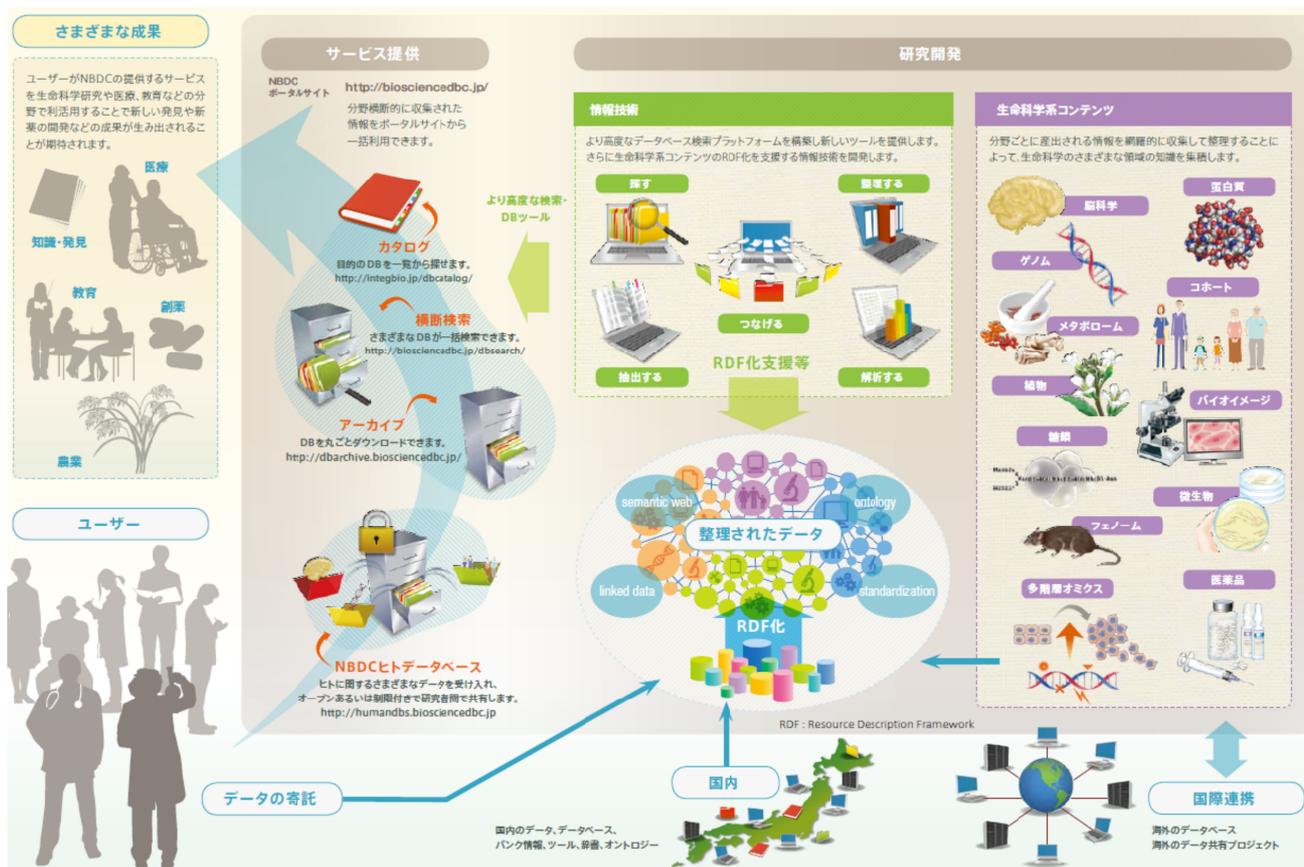
C. 結果

[NBDC の沿革・役割]

平成 18 年度から文部科学省「ライフサイエンス分

野の統合データベース事業」により、サイエンス分野のデータベースの統合作業が開始された。これは 5 年の時限付プロジェクトであったが、終了後、その成果を継続発展させることを目的として、平成 23 年 4 月に科学技術振興機構 (JST) に新たに「バイオサイエンスデータベースセンター」(NBDC: National Bioscience Database Center) が設置された。

NBDC ではライフサイエンス研究分野のデータの統合を進めるために、四省(文部科学省、厚生労働省: 医薬基盤研究所、農林水産省: 農業生物資源研究所、経済産業省: 産業技術総合研究所創薬分子プロファイリング研究センター)連携を進めるほか、国際連携に向けた取り組みも展開されている。一方で、データベース整備・充実、統合化、公開のみならず、その役割にはデータベース研究開発戦略の立案、データベースの基盤的技術開発も含まれる。



NBDCでは各省で行われた生命科学系プロジェクトを一覧できるデータベースカタログを整備し、個別にデータ寄託を受け付けるデータベースアーカイブも運用している。また、多様なデータベースを横断的に検索できるシステムを提供している。取り扱われているデータベースの範囲は広く、ゲノム、分子のレベルからイメージまで多岐に渡り、登録されている生物種も様々である。

[NBDC ヒトデータベース]

NBDCでは特にヒトに関する情報に特化したNBDCヒトデータベースを構築し、2013年10月から運用が開始された。その背景要因として、公的資金が投入されて実施された研究データに対する共有化要請、解析技術の進歩により大量に創出されるデータの有効活用、解析資源の効率的な利用が求められていることを挙げることができる。さらに、日本人のゲノム情報を用いた研究の必要性、日本人を対象とした個別化医療・臨床研究の推進、将来的な日本人・人類全体に関わる特に医学分野への貢献が重要な要素といえ

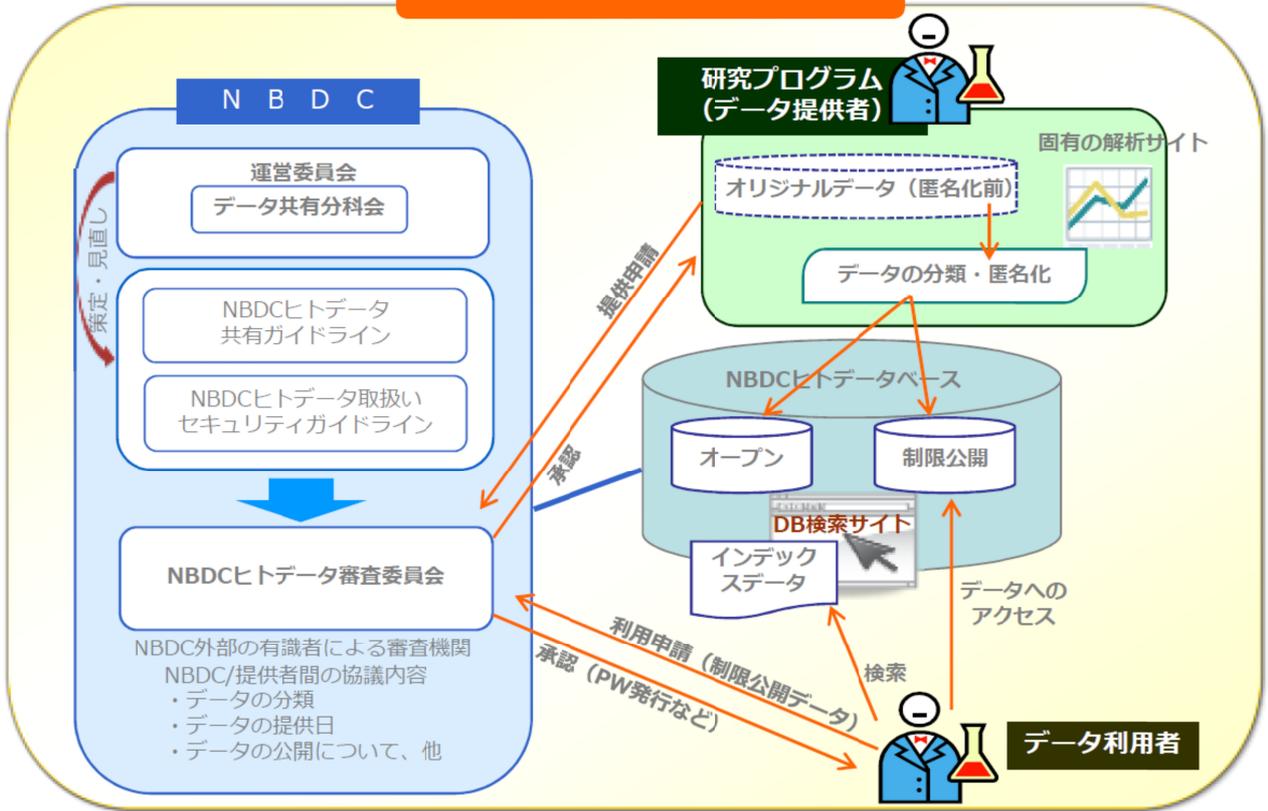
る。

ヒトを対象とするデータであるため、特に個人につながる情報の保護対策も重要である。NBDCでは、欧米のデータベースを参考に受け皿づくりが進められ、匿名化されたデータのみを扱い、さらにレベルに応じたアクセス制限が行われている。データ提供と利用に関する審査は、NBDCで行われるが、原則として試料提供者からデータ共有に関する同意を事前に受けておくことが求められている(ただし、過去に収集された既存試料・情報で同意を取り直すことが困難な場合は、共有されることが倫理委員会で承認されていることを要件とする)。データ共有に関するガイドライン、セキュリティレベルに関するガイドライン等が定められ、HP上で公開されている

(<http://humandbs.biosciencedbc.jp/guidelines>)

コホート研究に関しても受け入れは可能とのことである。ただし、匿名化されたデータのみが受け付けられるので、追跡データなどが必要な場合には、別に共同研究契約を結んで研究を実施することが必要となるであろう。

データ共有の仕組み



研究データは15件がHPに公開(2015年2月現在)され、そのうち制限公開10件、オープン5件であるが、すべてゲノムに関連するもので、いわゆる疫学研究のデータに関しては今までのところ、寄託はされていない。

たデータに関して実績があがるのはまだこれからと考えられた。

科研費で実施された疫学研究の情報も今後は寄託される可能性がある。実際、東北メディカル・目ババンクやオーダー迷路医療の実現プログラムはその方向で動いている。一方、疫学研究は単に生体情報の

D. 考察

国内で進められているライフサイエンス分野のデータベース統合化とヒトデータベースの現状について把握した。データ共有にかかる労力は大きい、その意義もまた大きいものと思われる。しかし、いまだ緒についたばかりであり、特にヒトを対象とし

| データの種類 | データ提供者 | 受入 | データベースセンター | 公開 | データ利用者 |
|------------------------------|---|---------|------------|----|---------|
| NBDCヒトデータベース | 1. オープン | 提供申請が必要 | | | 自由に利用可能 |
| | 2-1. 制限公開 (標準レベル[Type I] セキュリティ) | 提供申請が必要 | | | 利用申請が必要 |
| | 保管・利用に際してTypeIセキュリティレベルを要する | | | | |
| | 2-2. 制限公開 (ハイレベル[Type II] セキュリティ) | 提供申請が必要 | | | 利用申請が必要 |
| 保管・利用に際してTypeIIセキュリティレベルを要する | | | | | |
| 3. 公開待機 | 提供申請が必要 | | | | 利用できない |
| TypeIIと同レベルのセキュリティを適用 | | | | | |
| 4. 匿名化前・公開留保 | NBDCヒトデータベースでの共有対象外セキュリティレベルを各自で設定 | | | | 利用できない |

データ利用時に必要なセキュリティ環境

利用したいデータの制限公開レベルによって求められるセキュリティ環境レベルが異なります。

◆標準レベル [Type I]セキュリティ

- ・データは、所属組織LANに接続する制限公開データサーバ（ファイアウォール機能で所属組織LANの他の機器との間の通信が適切に管理されていること）、またはネットワークに接続しない制限公開データサーバに保存し、当該制限公開データサーバ外に移動しないこと。
- ・所属組織LAN内で、やむを得ず一時的に制限公開データサーバ外にデータを移動しなければならない場合は、利用後速やかに消去すること。
- ・データのコピーは作成しないこと。ただし、以下の場合は例外とする。
 - ①データをバックアップする場合。
 - ②データ移動時に一時的に作成する場合。
 - ③ソフトウェアによって一時的に作成される場合。
- ・データへのアクセスはデータ利用者に限定し、端末からのみ行うこと。

◆ハイレベル [Type II]セキュリティ

- ・標準レベル [Type I]セキュリティにおいて必要な対策に加え、制限公開データサーバに関して以下の対策を講じること。
- ・以下の条件を全て満たすサーバ室に制限公開データサーバを設置すること。
 - ①生体認証を用いて入室者を限定していること。
 - ②入室記録を自動取得し、後日監査可能であること。
 - ③申請した用途専用のサーバ室であること。専用サーバ室を確保できない場合は、常時施錠された専用のサーバラックに制限公開データサーバを格納すること。

詳しくは下記サイト内の「ガイドライン」をご覧ください。

<http://humandbs.biosciencedbc.jp/>



みならず、生活習慣や心情等に関する情報も収集されることが多い。また追跡結果も死亡・疾病罹患など機微情報を含む。そのため、単に塩基配列などの公開と異なり、疫学研究で構築されたデータセットの公開は一部に制限し、機微情報を取り扱う場合には共同研究を締結して行うなどの対応が必要と思われる。

E. 結論

NBDCにおけるライフサイエンス分野のデータベース、特にヒトを対象とするデータベースの運営状況、ガイドラインなど被験者保護対策について、確認した。疫学研究データについても受け入れは可能であるが、機微情報に関しては共同研究などを別途締結して実施する必要があると考えられた。

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

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
ホートデータを用いて、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るものとして対応する
必要がある。

A. 目的

三府県コホートデータについて標本数を変化さ
せた場合に、どのような頻度で一意性がみられる
かを比較し検討する。

B. 方法

三府県コホートデータを使用し 100,629 例全て
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昨年度と同じく、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。

グループ化についても昨年度と同様に、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。

(1) 定義

対象（本研究の場合は三府県コホート 100,629
例と、それより標本抽出された 1 万例、1,000 例、
100 例）の個体が、数種類の変数の組み合わせに
基づいていくつかのセルに分類されたとき、この
とり得た分類数を K とする。さらに 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。このとき、 $\sum S_i = K$ 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。

(2) 検討内容

[検討 1]

昨年同様、ベースとして【個人特性】と【追跡】のグループを考え、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)を行う。今回は【個人特性】については昨年と同じ2パターンで変更なし、【追跡】については昨年の4パターンに新たに2パターンを追加した6パターンのサブグループを定義した。それらの分類数 K とユニークセル数 S_1 を求めた。

[検討 2]

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 S_1 を求めた。

C. 結果

[検討 1]より、日付×転帰×死因からなる【追跡】グループでは、今回新たに検討した「追跡4」(まるめの処理としてICD-9コードを17の疾病大分類とする、かつ日付を月までにする)では分類数5,229、ユニークセル数2,083であった。「追跡5」(まるめの処理としてICD-9コードを17の疾病大分類とする、かつ日付を追跡期間(単位:月)でみる)では分類数1,593、ユニークセル数439であった。

昨年度の4パターンにこれらを追加したことで、最も大きいまるめの処理である「追跡6」(昨年度の「追跡4」に当たるもの:死因情報を除いて日付を追跡期間(単位:月)でみる)で一意性が消失するに至るまで、分類数とユニークセル数は漸減傾向を示した。(表1)

[検討 2] 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)。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

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を用いて、いくつかの変数の組合せごとに一意性を検討した。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

該当なし

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

がん登録推進法における死亡情報活用に関する研究

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 平成 28 年 1 月より施行される「がん登録等の推進移換する法律」について、条文や政令をもとに、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、死亡情報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については、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。また、がん登録推進法により提供される死亡者情報票データを活用して、米国の National Death Index と同様のシステムを作ることの可能性について検討した。

A. 目的

平成 25 年 12 月 6 日に国会成立した「がん登録等の推進に関する法律(以下、「がん登録推進法」)」により、平成 28 年 1 月から全国がん登録が始まる。

全国がん登録では、2 つの情報が国(国立がん研究センター)に集められる。第 1 に、がん罹患情報であり、がん患者を診断した医療機関(全ての病院と希望する診療所)からの情報が都道府県を經由して集められる。第 2 に、死亡者情報であり、市町村から保健所・都道府県を經由して集められる。

がん登録推進法第 3 条(基本理念)第 4 項では、「全国がん登録及びがん診療情報の収集により得られた情報については、これらががん患者の診療等を通じて得られる貴重な情報であることに鑑み、民間によるものを含めがんに係る調査研究のために十分に活用されるとともに、その成果ががん患者及びその家族をはじめとする国民に還元されなければならない」として、がん登録情報の研究活用が推奨されている。

本研究の目的は、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、死亡情報活用をめぐる諸問題について考察することである。

B. 方法

がん登録推進法の条文を読解するとともに、がん登録推進法の実施に関する政令・省令などを審議する場である厚生科学審議会がん登録部会での議論を聴取し、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、死亡情報活用をめぐる諸問題について考察した。

C. 結果

(1) 死亡者情報票の作成と照合

死亡者情報票の作成に関する事項は、がん登録推進法第 11 条(死亡者情報票の作成及び提出)で明記されている。それによると、死亡者情報票は、死亡の届書その他の関係書類に基づいて、市町村長が作

成するものとされている。死亡者情報票に含まれる情報は、死亡した者に関する氏名、性別、生年月日、死亡の時の住所、死亡の日、死亡の原因、死亡診断書の作成に係る病院又は診療所の名称及び所在地その他の厚生労働省令で定める情報とされている。

死亡者情報票は、電磁的記録又は書類により作成され、保健所、都道府県を経て、国(国立がん研究センター)に提出される。すなわち、市町村は全死亡者に関する死亡者情報票を提出するのである。

それを受けて、国立がん研究センターは、死亡者情報票と全国がん登録情報とを照合する。これにより、前者のなかから、後者に含まれる者だけが抽出されることになる。

(2) 生存確認の期間

全国がん登録データベースに記載された者と死亡者情報票との照合(生存と死亡の別に関する調査)を行う期間については、がん登録推進法の第12条第2項で「前項の規程による照合は、がんに係る調査研究のためにがん罹患した者が生存しているか死亡したかの別を調査する必要があると認められる期間として政令で定める期間が経過した全国がん登録情報等については、(略)」とあるように、法律ではなく政令により別途定めるとされている。

そこで、厚生科学審議会がん登録部会は「がんに係る調査研究のためにがん罹患した者が生存しているか死亡したかの別を調査する必要があると認められる期間は、100年とする」という政令案を取りまとめた。

これにより、がん患者の生命予後をほぼ完璧に追跡することが可能になったと思われる。

(3) 調査研究目的での利用手続き

これについては、がん登録推進法の第21条(その他の提供)に規定されている。該当項は、厚生労働大臣ががん登録情報の提供を行う場合(二以上の都道府県に係る都道府県がん情報の提供の求めを受けた場合)は第3・4項であり、都道府県知事ががん登録情報の提供を行う場合(当該都道府県に係る都道府県がん情報の提供の求めを受けた場合)は第7・8項で

ある。

第3項と第7項は匿名化されない情報の提供に係る規程であり、第4項と第8項は匿名化された情報の提供に係る規程である。第3項と第7項は、調査研究に対して情報提供を行うための4要件を規定している。第1に、その調査研究が、がん医療の質の向上等に資するものであること。第2に、その調査研究を行う者が、がん医療の質の向上等に資する研究の実績を相当程度有すること。第3に、その調査研究を行う者が、がん罹患等の秘密の漏洩防止その他の当該全国がん登録情報の適切な管理のために必要な措置を講じていること。第4に、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。

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として政令で定めるものについては、附則第2条(経過措置)により「施行日後に、その対象とされている者(略)の第21条第3項第4号又は第8項第4号の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)全国がん登録情報又は都道府県がん登録情報の提供については、同条第3項第4号又は第8項第4号の規程は、適用しない」とされている。

そして、法施行後に本人の「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」として政令が定める要件について、厚生科学審議会がん登録部会は、「法施行前に調査研究の対象者となっていた場合又は法施行前に承認された研究計画に従って対象者となっていたが、法の施行までに集めきらず、法施行後に対象者となった者」かつ「すでに集められた対象者の数が多い場合又はすでに集められた対象者の数は少

ないが、募集後長期間が経過し連絡することが困難な場合、若しくはその連絡先を把握していないデザイナーの場合」とした。

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については、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。

がん登録推進法の法案が国会に提出されたときの理由書には「がん登録により得られた情報の活用について定めることにより、がんの罹患、診療、転帰等の状況の把握及び分析その他のがんに係る調査研究を推進し、もってがん対策の一層の充実に資する必要がある」と述べられている通り、がん登録推進法では、適切な個人情報保護と漏洩防止のもとでの研究活用が推奨されている。

現在、多くのがん関連学会で臓器がん登録(学会登録)が行われている。それは、各学会が標的とする臓器がんに関する詳細な臨床情報(がんの組織型や治療法など)を登録するものであり、がん登録推進法による全国がん登録よりも詳細な情報を収集している。一方、学会登録では、追跡精度の低さが問題となっている。医療機関を受診しなくなった患者について、その後の情報が得られないからである。そこで、学会登録に参加する各医療機関ががん登録推進法第20条(病院等への提供)により当該がん患者に係る生存・死亡の確認情報(死亡時には、死亡年月日や死因など)の提供を求め、入手した情報を匿名化したうえで学会の臓器がん登録に集約できれば、(生存死亡の確認期間は100年と規定されるため)100%に近い追跡精度で生命予後を把握することが可能となる。これにより、わが国のがん臨床研究は飛躍的に発展するであろう。

一方、そのような恩恵が他の疾患に及ばないことは

残念なことである。市町村が国に提出する死亡者調査票は、がん罹患患者だけでなく、すべての死亡者を対象とするものである。しかし、全国がん登録情報に記載された者以外の情報は、現行法制では利用されない。

コホート研究は研究対象者の生存・死亡の確認を長期にわたって行うものであるが、追跡期間中に転居した場合、その後の追跡が困難になることが多い。それに対して、死亡者情報票を追跡に利用することがもし可能となれば、追跡の精度は飛躍的に向上することが期待される。

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おいて国際的に有利な地位を確保することができた。

がん登録推進法により提供される死亡者情報票データを活用して、米国のNDIと同様のシステムを作るには法制度の改革が必要であることは言うまでもないが、それが実現すれば、わが国の疫学研究・臨床研究や医薬品・医療機器開発は発展すると思われ、それは政府「健康・医療戦略」の目指すところと合致するものであろう。

E. 結論

平成28年1月より施行される「がん登録等の推進移換する法律」について、条文や政令をもとに、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、死亡情報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については、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。また、がん登録推進法により提供される死亡者情報票データを活用して、米国の National Death Index と同様のシステムを作ることの可能性について検討した。

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

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
研究協力者 原田亜紀子 東京大学大学院医学系研究科公共健康医学専攻

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。提案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。想定するアーカイブセンター等で実際に運用を行う上では、データ利用の規約、データ利用の物理的環境についてあらかじめ検討すべき事項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>)。2010年時点で、BioLINCCには、56臨床試験、20疫学観察研究のデータが収集されており、使用申請を行うことにより、研究利用できるとともに、生物統計学系の学生などに対してはテストデータの提供なども実施されている(BioLINC HP: FAQs <https://biolincc.nhlbi.nih.gov/faqs/#q2>)。2000～2010年の間で、約500件の利用申請に対し、出版物が224件、68本の論文が出版されており、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といえる。

翻って我が国の現状をみると、厚生労働省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、いくつかの要因が考えられる。一点目は、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である。二点目は、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、この情報を付与した状態でのデータ公開が難しい点である。三点目としては、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点にあると考えられる。

そこで本報告では、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、現制度下で研究基盤づくりを進めていく上で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・解析を行う運用例を提案する。

2. 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、上記1で提案する運用例になって死因を照合する。この際に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する。

1) 人口動態統計の二次利用申請

現在、我々が行っている国内の循環器疫学研究 Japan Arteriosclerosis Longitudinal Study (JALS) では、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、人口動態調査二次利用の申請書類を作成し、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、研究開始から 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。

2) 提案例に従った照合作業の検証

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と照合し原死因を確定した。

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研究基盤の提案

1) 統計法 33 条下での個票データの提供

本邦の多くのコホート研究では、死因情報は人口動態統計調査の二次利用申請を行い入手している。コホート研究を行うにあたっては、個別 (研究対象者個人) の死因が必要となることから、公的統計の二次的な利用方法のうち、「統計法 33 条に基づく個票データの利用を申請する方法」がとられている。この利用申請は、公的な研究費の補助を受けた研究などに限定され、研

究者は利用したい統計の担当省庁の事前相談窓口を利用の申し出を行い、事前審査と本審査を経て許可される仕組みとなっている。利用申請、実際の利用にあたっては、統計法 33 条により、データを利用する者、利用する場所、利用する目的の範囲が厳しく制限されており、本報告において議論の対象としている死因情報を付加したデータのアーカイブ化 (オープンデータ化、一定の制限を設けた下での二次利用) は、不可能な状態にある。

2) 死因データを必要時に連結させる方法の提案

上述の通り現行の統計法下では、ローカル (各研究グループ) が行った二次利用申請で得た死因を付与したデータをアーカイブすること (中央に集めること) はできないので、死因を連結した形でのデータセット構築はあきらめて、必要時に中央 (アーカイブデータを保持するセンター等) で死因照合作業を行って解析用データセットを作成する方法を提案する。

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、JALS を国内疫学研究のコンソーシアム (アーカイブ事業に参加する研究グループ) に、中央事務局をアーカイブセンターに置き換え考えることで、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) により容易に照合が行えるような環境を検討する。

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」を含むように設計する。

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する。この際に、データの利用場所およびその利用者に、研究申請者とアーカイブセンターを含む形で申請を行う（**死因情報申請**）。

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に、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

2. 必要時に死因を照合する方法の実際（JALS 事例）

JALS で現在実施している方法をもとに、照合手順について具体的に説明する。

1) 追跡情報の収集

コホート（各参加研究）から中央事務局（アーカイブセンター）に、死因照合の際にキーとなる「死亡地（市町村）」、「生年月日」、「死亡年月日」、「性別」を含めた追跡情報を提出する。死亡地（市町村）については、市町村コードに加え、都道府県コード、保健所コードなどを併記した対応表を作成した。

2) 二次利用申請

必要とする統計情報の変数、上記1)でリストアップされた市町村、必要とする調査対象年などをもとに、死因情報の申請範囲をとりまとめる。その他の申請書類を合わせて作成し申請した。

3) 承認後、データの受領

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。提供を受けたデータ（ファイル）の形式は、CSVファイル形式で、1年1ファイルで提供された。

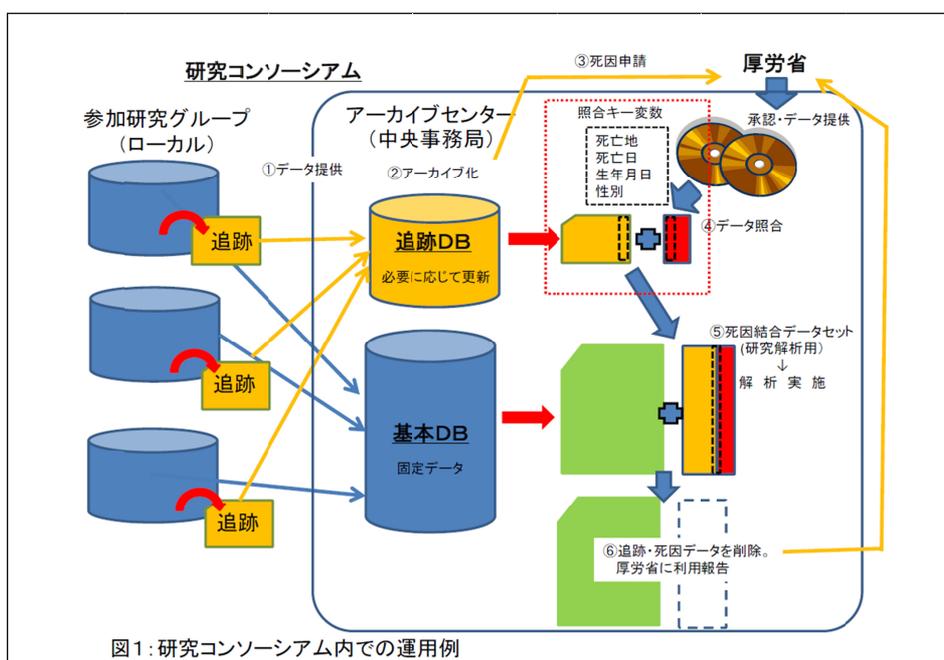


図1: 研究コンソーシアム内での運用例

ある。

4) 照合作業

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プログラム例を付録に示した(SAS プログラム#1)。

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#2)

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#3)

照合作業(SAS プログラム#4)

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#5)

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(帳票例を付録に示した)。

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。

D. 考察

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た。この提案法を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 は国内コホート研究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(例:性別が男女逆、1日ずれている、生年月日と死亡日が同じ日付となっているなど)。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. 市町村合併の問題

JALSの例では、研究実施期間内に平成の市町村大合併の時期を含んでおり、実際、市町村リストの作成作業が煩雑で、データ受領後の照合作業において工夫が必要となっている。合併年月日のリスト化、合併前後の自治体の包含関係(新規設置であるか、吸収合併であるか等)について注意を払ってリストを作成しておかねばならない。しかし、これらの対応については一度その対応手順を検討してしまえば、以後は無理なく対応できるものと考えられる。

4. 研究成果報告に関連した対応

1) 死因つきデータセットの保持可能な期間

昨今の臨床研究実施に関連した不祥事に関連し、研究の質保証の観点から(結果の再現性など)、研究成果報告後に、データ再検証・再解析対応が必要なことも考慮しておく必要があるといえる。また、このような事例に限らず、論文化手続き(査読課程)などが長期化して再解析実施の可能性が考えられる場合に、死因付きデータをいつまで保持できるかは課題であるといえる。公的研究費を継続できるような場合には、死因情報を保持することは可能であるとはいえ、研究の質や結果の再現性を保証するという点では、解析に使用した死因付きのデータセットが長期に保持できることが望ましいといえる。

2) 死因つきデータセットが使用できる環境が限定され

る

提案する方法では、死因つき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、研究の解析もこの利用場所に限られることになる。アーカイブデータの利用規定も合わせて、アーカイブセンターで対応する場合は、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。

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。提案した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上では、データ利用の規約、データ利用の物理的環境についてあらかじめ検討すべき事項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

付録

作業1.人口動態統計データの読み込み (SAS プログラム#1)

<データセット構造>

以下のようなデータセットが1年1ファイル提供されているとする(ただし変数は申請内容により異なる)。

例)

| | 調査年 | 都道府県 | 保健所 | 市町村 | 性別 | 生年月日 | | | | 死亡年月日 | | | | 原死因 |
|----|------|------|-----|-----|----|------|---|---|---|-------|---|---|---|-----|
| | | | | | | 元号 | 年 | 月 | 日 | 元号 | 年 | 月 | 日 | |
| 1 | 1999 | | | | | | | | | | | | | |
| 2 | 1999 | | | | | | | | | | | | | |
| 3 | 1999 | | | | | | | | | | | | | |
| 4 | 1999 | | | | | | | | | | | | | |
| 5 | 1999 | | | | | | | | | | | | | |
| 6 | 1999 | | | | | | | | | | | | | |
| 7 | 1999 | | | | | | | | | | | | | |
| 8 | 1999 | | | | | | | | | | | | | |
| 9 | 1999 | | | | | | | | | | | | | |
| 10 | 1999 | | | | | | | | | | | | | |
| 11 | 1999 | | | | | | | | | | | | | |
| 12 | 1999 | | | | | | | | | | | | | |
| 13 | 1999 | | | | | | | | | | | | | |
| 14 | 1999 | | | | | | | | | | | | | |
| 15 | 1999 | | | | | | | | | | | | | |
| 16 | 1999 | | | | | | | | | | | | | |
| 17 | 1999 | | | | | | | | | | | | | |
| 18 | 1999 | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |

H11(1999)年から H24 年までのデータを読み込む

Cドライブの「C:¥死因照合¥照合_hxx」の階層に、shibo_hxx.csv ファイルが保存されている状態 (xx は年数)。

<SAS プログラム例 #1>

```
%MACRO CSVREAD (YEAR)
```

```
data JINKO.H&YEAR ;
```

```
infile "C:¥死因照合¥照合_h&year.¥sibo_h&year.csv" dlm=',' MISSOVER DSD lrecl=32767
```

```
firstobs=1 ;
```

```
input
```

```
COL1 $2.
```

```
COL3 $1.
```

```
COL14 $2.
```

```
COL15 $2.
```

```
COL16 $3.
```

```
COL17 $1.
```

```
·
```

```
·
```

```
·
```

```
;
```

CSV ファイル形式で提供された下記の人口動態データを読み込み、調査年毎に SAS データセットとして格納する。

申請内容に応じて変数は変更

調査年,都道府県,保健所,市区町村,
性別,生年月日__元号,生年月日__年,
生年月日__月,生年月日__日,死亡日__元号,
死亡日__年,死亡日__月,死亡日__日,原死因,

```
%MEND ;
%CSVREAD(11);%CSVREAD(12);%CSVREAD(13);%CSVREAD(14);
%CSVREAD(15);%CSVREAD(16);%CSVREAD(17);%CSVREAD(18);
%CSVREAD(19);%CSVREAD(20);%CSVREAD(21);%CSVREAD(22);
%CSVREAD(23); %CSVREAD(24);
```

DATA JINKO ;

```
SET JINKO.H11(in=A) JINKO.H12(in=B) JINKO.H13(in=C) JINKO.H14(in=D) JINKO.H15(in=E)
```

```
JINKO.H16(in=F) JINKO.H17(in=G) JINKO.H18(in=H) JINKO.H19(in=I) JINKO.H20(in=J)
```

```
JINKO.H21(in=K) JINKO.H22(in=L) JINKO.H23(in=M) JINKO.H24(in=N) ;
```

run;

調査年ごとにセットして 1 つの SAS データセットとして格納する。

 データセット“JINKO”

作業 2. 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む

追跡データに市町村コードを事前に付与しておくか、各研究と該当する市町村の対応リストを作成し読み込む。

《データイメージ》

| 研究識別コード | 都道府県 | 都道府県コード | コホート名 | 保健所 | 保健所コード | 市町村 |
|---------|------|---------|-------|-----|--------|-----|
| A01 | 北海道 | 1 | a | aa | 11 | 〇〇町 |
| A01 | 北海道 | 1 | a | bb | 22 | 〇△町 |
| A03 | 秋田 | 5 | b | cc | 33 | 〇×町 |
| A05 | 秋田 | 5 | c | dd | 44 | △×町 |
| | . | | | | | |
| | . | | | | | |
| | . | | | | | |
| A30 | 沖縄 | | | | | |

作業 3. 追跡調査データベースの対象データから照合を行いたい死亡者のみを抽出

市町村コードが事前に入力されていない場合は、作業 2 のリストから付与し読み込む

| 研究識別コード | 市町村 | ID | 性別 | 生年月日 | 死亡日 | 生死(死亡=2) | 市町村コード |
|---------|-----|----------|----|------------|------------|----------|--------|
| A01 | 〇〇町 | 1000222 | 1 | 1906-02-25 | 2009-11-07 | 2 | 01548 |
| A01 | 〇〇町 | 1000200 | 1 | 1906-11-28 | 2001-11-21 | 2 | 01548 |
| A01 | 〇〇町 | 1000030 | 1 | 1907-04-02 | 2008-12-31 | 2 | 01548 |
| A01 | 〇〇町 | 1004000 | 2 | 1911-01-14 | 2006-04-29 | 2 | 01548 |
| A01 | 〇〇町 | 1004020 | 1 | 1911-06-21 | 2008-07-13 | 2 | 01548 |
| A01 | 〇△町 | 1005030 | 2 | 1911-06-29 | 2001-05-13 | 2 | 01575 |
| A01 | 〇△町 | 1105030 | 1 | 1909-01-01 | 2000-02-22 | 2 | 01575 |
| A01 | 〇△町 | 1105044 | 1 | 1909-12-01 | 2000-04-30 | 2 | 01575 |
| A01 | 〇△町 | 1106048 | 1 | 1939-03-09 | 2000-10-25 | 2 | 01575 |
| A01 | 〇△町 | 1126048 | 1 | 1924-09-19 | 2011-12-25 | 2 | 01575 |
| A03 | 〇×町 | 9987 | 2 | 1934-09-20 | 2003-05-11 | 2 | 05324 |
| A03 | 〇×町 | 6989 | 2 | 1952-07-30 | 2006-03-17 | 2 | 05324 |
| A03 | 〇×町 | 7878 | 1 | 1922-11-19 | 2006-06-17 | 2 | 05324 |
| A03 | 〇×町 | 9900 | 1 | 1953-01-04 | 2011-03-11 | 2 | 05324 |
| A03 | 〇×町 | 6677 | 1 | 1910-01-07 | 2005-09-30 | 2 | 05324 |
| A03 | 〇×町 | 6789 | 1 | 1928-07-18 | 2008-05-03 | 2 | 05324 |
| A05 | △×町 | 99993737 | 2 | 1944-10-24 | 2005-06-13 | 2 | 05367 |
| A05 | △×町 | 98883799 | 1 | 1955-11-28 | 2010-06-23 | 2 | 05367 |
| A05 | △×町 | 99994444 | 2 | 1910-10-24 | 2012-08-09 | 2 | 05367 |
| A05 | △×町 | 98896677 | 1 | 1966-01-02 | 2008-07-13 | 2 | 05367 |

 読み込んだ照合用死亡者データセットを “S_SICHO” とする。

作業4 照合対象データと人口動態データをキー変数で結合

データセット JINKO と S_SICHO とを性別、生年月日、市町村コード、死亡日の情報でマージする

<SAS プログラム例 #4>

```
(1) data S_SICHO;  
    set S_SICHO ;
```

```
mkey=compress(put(sex,best.)||"_"||put(birth,yymmdd10.)||"_"||shicyou_c||"_"||put(shibou,yymm  
dd10.));
```

```
run ;
```

(1) 研究対象データで、照合時のキー変数を作成、付与を行う。
照合時のキー変数とは、
性別、生年月日、市町村コード、死亡日の情報を結合し、文字列としたものである。
(プログラム中"mkey"として記載)

```
(2) data JINKO ;  
    set JINKO ;
```

```
mkey=compress(put(sex,best.)||"_"||put(birth,yymmdd10.)||"_"||shicyou_c||"_"||put(shibou,yymm  
dd10.));
```

```
run ;
```

(2) 人口動態データで、照合時のキー変数を作成、付与を行う。

```
(3) data S_SHOGO S_ERR JINKO_ERR ;  
    merge S_SICHO(in=a) JINKO(in=b) ;  
    by MKEY ;  
    if a and b then output S_SHOGO ;  
    else if a and b^=1 then output S_ERR ;  
    else if a^=1 and b=1 then output JINKO_ERR ;  
run ;
```

(3) 作成したキー変数を使用し、
研究の死亡データと、人口動態データを結合する。
結合した対象者は死因照合例としてデータセットに格納する。
結合しなかった対象者は、照合不能例として別のデータセットに格納する。

```
(4) data DUPTAI ;  
    set S_SHOGO ;  
    by mkey ; if first.mkey then count=0 ; COUNT+1 ;  
        IF COUNT=2 ;  
        KEEP MKEY;  
        run ;
```

```
data DUPLIST ;  
merge S_SHOGO duptai(in=a) ;  
    by mkey ;  
    if a ;  
run ;
```

(4) 照合候補例が複数発生した場合のデータセットを格納する。

作業5 死因照合

上記作業4で作成したデータセットをもとにして、(1)照合例、(2)未照合例、(3)重複例(候補例が複数存在したものを)をリストアップする

- (1) 照合例 データセット … “S_SHOGO”
 (2) 未照合例 … “S_ERR”
 (3) 重複例(候補例が複数存在) … “DUPLIST”

JALS では、これらのデータセットから EXCEL へ出力するプログラムを作成済みである。

a) 人口動態統計 照合状況集計表

| 受領人口動態データ件数(～2012)= XXXXXX | | 重複削除後人口動態データ件数(～2012)= XXXXXX | | | | | | | |
|----------------------------|------------|-------------------------------|-------|-------------|------------|--------------|----------------------------|--------------------------------------|--|
| 集計 | | | | | | | 率 | | |
| 統合研究死亡者数 | 人口動態統計対象者数 | 結合件数 | 結合者数 | 結合中結合キー重複者数 | 統合研究対象者のみ数 | 人口動態統計対象者のみ数 | 結合率 【結合者数÷統合研究死亡者数×100】 | 結合中結合キー重複率 【結合中結合キー重複者数÷結合者数×100】 | |
| 7,137 | XXXXX | 7,104 | 7,099 | 5 | XX | XXXXXXX | 99.47 | 0.07 | |

b) コホート別照合・未照合数表

| コホート毎不一致対象者数 | | |
|--------------|----------|---------|
| コホート | JALS死亡者数 | 不一致対象者数 |
| 1 | 126 | 1 |
| 2 | 181 | 4 |
| 3 | 2120 | 4 |
| 4 | 76 | 0 |
| 5 | 127 | 0 |
| 6 | 252 | 4 |
| 7 | 520 | 6 |
| 8 | 118 | 0 |
| 9 | 0 | 0 |
| 10 | 112 | 0 |
| 11 | 295 | 3 |
| 12 | 0 | 0 |
| 13 | 899 | 2 |
| 14 | 93 | 1 |
| 15 | 0 | 0 |
| 16 | 0 | 0 |
| 17 | 356 | 1 |

・
・
・
・

c) 未照合 (不一致) 対象者一覧

| 受領人口動態データ件数 (~2012) = XXXX 重複削除後人口動態データ件数 (~2012) = XXXX 統合研究対象者のみ数 = XXXX | | | | 統合研究死亡者数 = XXXX 結合者数 = XXXXX | | | | | |
|--|---------|-----------|---------------|---------------------------------|----------------|-----------|---------|----|---------|
| コホート番号 | ローカルコード | ID | 性別 男=1,女=2 | 生年月日 | 都道府県・保健所・市町村結合 | 死亡日 | 生存状態コード | 死因 | 市町村名 |
| 1 | | aaaaaaa | 1 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | 〇〇市 |
| 2 | | bbbbbbb | 1 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | 〇△町 |
| 3 | | ccccccc | 1 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | △△町 |
| 5 | | ddddddd | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | △×町 |
| 6 | | eeeeeee | 1 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | ××村 |
| 6 | | fffffffff | 1 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | 〇〇町 |
| 7 | | sssssss | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | △△市 |
| 8 | | hhhhhhh | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | 〇〇町 |
| 9 | | iiiiiiii | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | ××町 |
| 9 | | jjjjjjjj | 1 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | △△〇市××区 |
| 11 | | kkkkkkk | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | △△〇市××区 |
| 13 | | lllllllll | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | △△〇市××区 |
| 13 | | mmmmmm | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | △△〇市××区 |
| 20 | | nnnnnnn | 1 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | 〇△町 |
| 21 | | oooooooo | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | ××市 |
| 31 | | ppppppp | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | ××市 |
| 32 | | qqqqqqq | 2 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | ××市 |
| 33 | | rrrrrrrr | 1 | 19XX-XX-X | 0xxxx | 20XX-XX-X | 2 | | ××市 |

d) 重複対象者一覧

| 受領人口動態データ件数 (~2012) = XXXX 重複削除後人口動態データ件数 (~2012) = XXXX 重複者数 = 3 | | | | 統合研究死亡者数 = XXXXX 結合者数 = XXXXX | | | | | | | | | | |
|---|---------|---------|---------------|----------------------------------|----------------|-----------|---------|----------|------|---------------|------------|----------------|-----------|------|
| コホート番号 | ローカルコード | ID | JALS情報 | | | | | 人口動態統計情報 | | | | | | |
| | | | 性別 男=1,女=2 | 生年月日 | 都道府県・保健所・市町村結合 | 死亡日 | 生存状態コード | 死因 | 市町村名 | 性別 男=1,女=2 | 生年月日 | 都道府県・保健所・市町村結合 | 死亡日 | 原死因 |
| 3 | | AAAAA | 1 | 1935-XX-XX | 0XXXX | 20XX-XX-X | 2 | | 〇△町 | 1 | 1935-XX-XX | 0XXXX | 20XX-XX-X | aaa |
| 3 | | AAAAA | 1 | 1935-XX-XX | 0XXXX | 20XX-XX-X | 2 | | 〇△町 | 1 | 1935-XX-XX | 0XXXX | 20XX-XX-X | aaa |
| 3 | | ZZZZZZZ | 1 | 1940-XX-XX | 0XXXX | 20XX-XX-X | 2 | | ××市 | 1 | 1940-XX-XX | 0XXXX | 20XX-XX-X | bbbb |
| 3 | | ZZZZZZZ | 1 | 1940-XX-XX | 0XXXX | 20XX-XX-X | 2 | | ××市 | 1 | 1940-XX-XX | 0XXXX | 20XX-XX-X | bbbb |
| 7 | | WWWWW | 2 | 1930-XX-XX | 0XXXX | 20XX-XX-X | 2 | | 〇〇市 | 2 | 1930-XX-XX | 0XXXX | 20XX-XX-X | cccc |
| 7 | | WWWWW | 2 | 1930-XX-XX | 0XXXX | 20XX-XX-X | 2 | | 〇〇市 | 2 | 1930-XX-XX | 0XXXX | 20XX-XX-X | dddd |

研究成果の刊行に関する一覧表

2014 年度：なし