

201412032A

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

(循環器疾患・糖尿病等生活習慣病対策政策研究事業)

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

平成 27 (2015) 年3月

研究代表者 玉腰 暁子

目 次

I.	総括研究報告	
	追跡終了後コホート研究を用いた共通化データベース基盤整備とその活用に関する研究：進捗報告.....	1
	玉腰暁子	
II.	分担研究報告	
	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に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、①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、②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、③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④現制度化での利用方法の提案を行った。

B. 研究方法

①各分野の専門家から現状を伺い、追跡を終了したコホート研究データアーカイブを公開する場合の課題を検討した。

②三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。

③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。

④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

①-1 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できないため、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、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。

① -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。

② 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。

③ 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、死亡の時ににおける住所、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。

④ 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。①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。②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。③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。④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。⑤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。⑥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

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

分担報告書

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。

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。

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。

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。

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

[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。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。

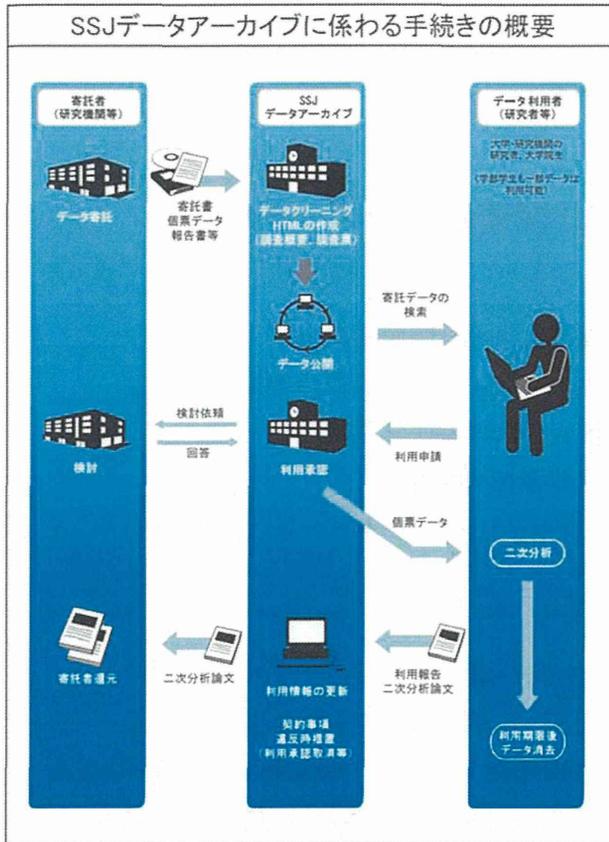
運営費用は文部科学省(2010年度より国立大学法人共同利用・共同研究拠点)、東京大学社会科学研究所から運営費、データアーカイブに関わる科学研究費で賄われている。

[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 at 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。

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わ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。

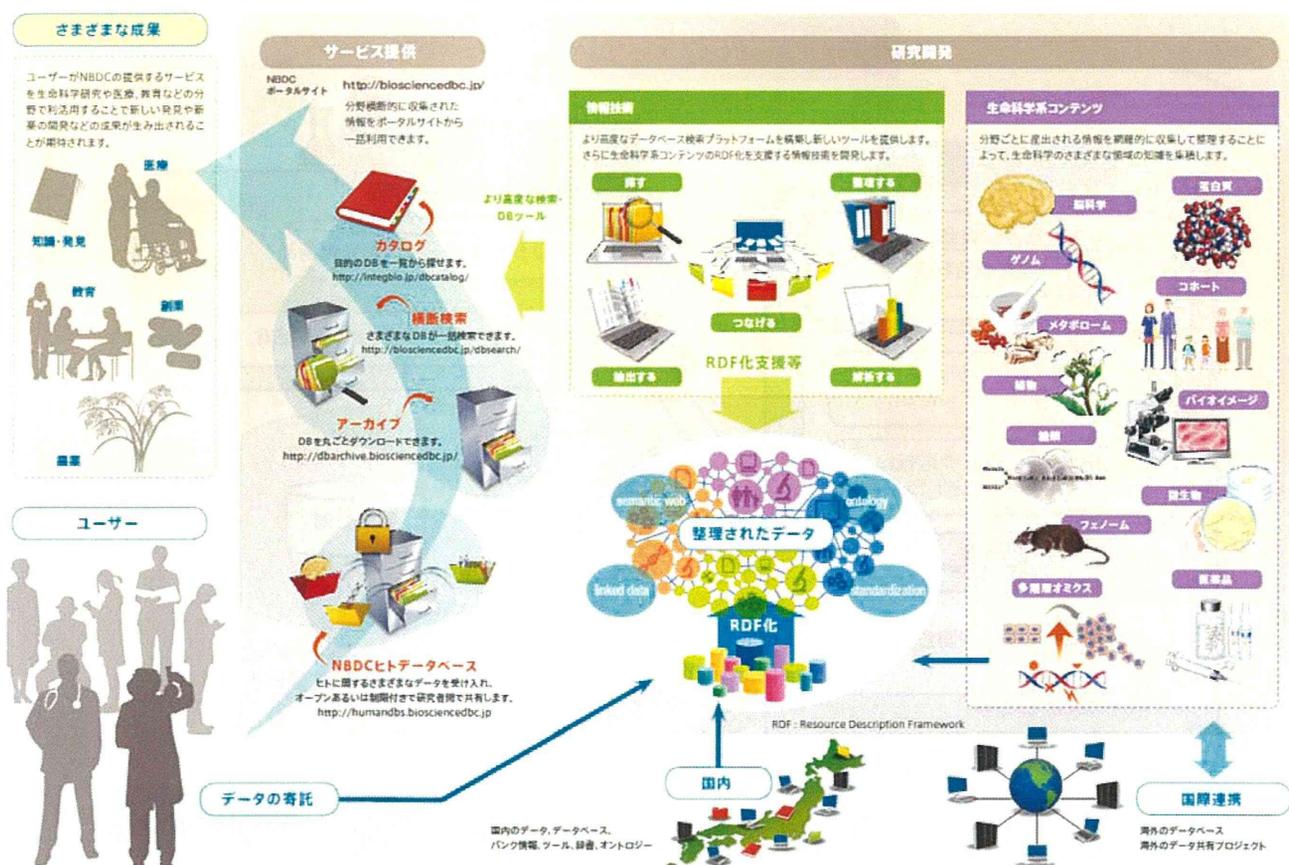
C. 結果

[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

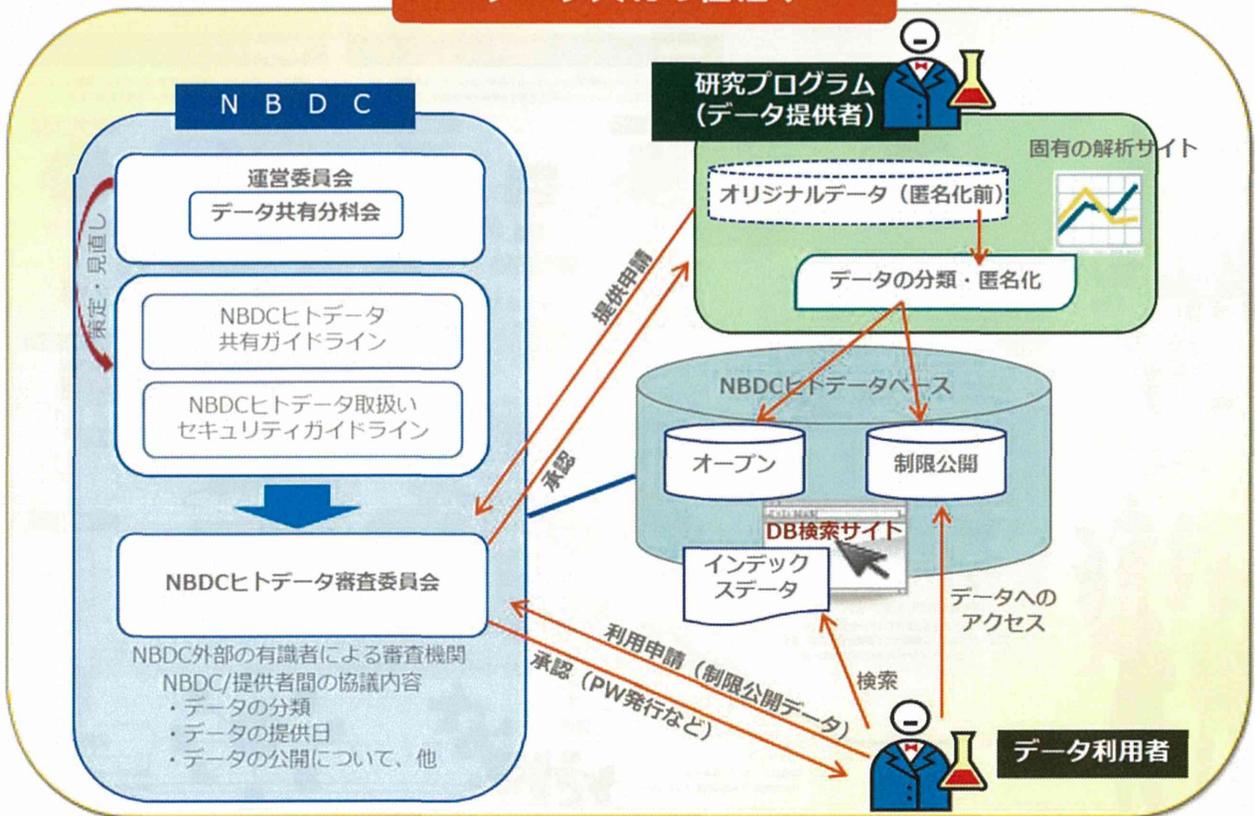
る。

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。

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。

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・目が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

データの種別	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Type I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Type II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
Type II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ることでカテゴリ化
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①アンケートカテゴリのみ、②【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、③【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、④【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、①~④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

表1 ベースライングループ(個人特性【追跡】)に対するまるめの処理について

個人特性	処理内容	分類数	ユニークセル数
1	処理なし(性別×年齢×居住地)	673	19
2	年齢5歳階級かつ85歳以上まるめ	120	0

追跡	処理内容	分類数	ユニークセル数
1	処理なし(日付:日まで×転帰×死因ICD-9コード4桁)	20176	16631
2	日付:月まで×転帰×死因ICD-9コード3桁	11398	7936
3	追跡期間(月)×転帰×死因ICD-9コード3桁	6225	3654
4	日付:月まで×転帰×死因疾病大分類	5229	2083
5	追跡期間(月)×転帰×死因疾病大分類	1593	439
6	追跡期間(月)×転帰×(死因削除)	243	0

表2 アンケートカテゴリの組み合わせによって得られた各パターンにおける分類数とユニークセル数(標本数別)

アンケートカテゴリ	個人特性	追跡	N=100629			n=10000			n=1000			n=100		
			分類数 K	ユニークセル数 S ₂	S ₂ /K	分類数 K	ユニークセル数 S ₂	S ₂ /K	分類数 K	ユニークセル数 S ₂	S ₂ /K	分類数 K	ユニークセル数 S ₂	S ₂ /K
なし(ベースラインのみ)	0	1	20176	16631	0.82	2710	2499	0.92	300	283	0.94	43	37	0.86
なし(ベースラインのみ)	0	2	11398	7936	0.70	2040	1611	0.79	284	256	0.90	41	33	0.80
なし(ベースラインのみ)	0	3	6225	3654	0.59	1394	1014	0.73	236	197	0.83	35	31	0.89
なし(ベースラインのみ)	0	4	5229	2083	0.40	1647	1078	0.65	263	225	0.86	41	33	0.80
なし(ベースラインのみ)	0	5	1593	439	0.28	721	295	0.41	202	153	0.76	34	29	0.85
なし(ベースラインのみ)	0	6	243	0	0.00	241	1	0.00	155	88	0.57	29	26	0.90
なし(ベースラインのみ)	1	0	673	19	0.03	609	47	0.08	413	145	0.35	90	82	0.91
なし(ベースラインのみ)	1	1	31208	29062	0.93	3876	3245	0.84	682	498	0.73	93	88	0.95
なし(ベースラインのみ)	1	2	28383	24606	0.87	3823	3154	0.83	682	498	0.73	93	88	0.95
なし(ベースラインのみ)	1	3	27949	24168	0.86	3691	3128	0.85	642	457	0.71	93	88	0.95
なし(ベースラインのみ)	1	4	26280	21880	0.83	3785	3083	0.81	680	494	0.73	93	88	0.95
なし(ベースラインのみ)	1	5	25700	21390	0.83	3643	3038	0.83	640	453	0.71	93	88	0.95
なし(ベースラインのみ)	1	6	22743	18272	0.80	3261	2655	0.81	608	407	0.67	93	88	0.95
なし(ベースラインのみ)	2	0	120	0	0.00	120	0	0.00	109	4	0.04	61	38	0.62
なし(ベースラインのみ)	2	1	27899	25715	0.92	3248	3013	0.93	437	332	0.76	78	66	0.85
なし(ベースラインのみ)	2	2	22025	17925	0.81	3104	2773	0.89	435	328	0.75	77	64	0.83
なし(ベースラインのみ)	2	3	21475	17273	0.80	3058	2746	0.90	411	319	0.78	76	63	0.83
なし(ベースラインのみ)	2	4	18812	13770	0.73	2968	2568	0.87	432	322	0.75	77	64	0.83
なし(ベースラインのみ)	2	5	18152	12951	0.71	2900	2511	0.87	408	313	0.77	76	63	0.83
なし(ベースラインのみ)	2	6	13749	7258	0.53	2587	2220	0.86	360	262	0.73	75	61	0.81
自覚症状	0	0	860	188	0.22	308	114	0.37	108	62	0.57	28	14	0.50
自覚症状	0	1	28210	24958	0.88	3586	3208	0.89	473	412	0.87	71	59	0.83
自覚症状	0	2	22813	18730	0.82	3339	2828	0.85	470	407	0.87	71	59	0.83
自覚症状	0	3	17440	13779	0.79	2687	2216	0.82	388	343	0.88	59	50	0.85
自覚症状	0	4	19276	14262	0.74	3185	2613	0.82	467	401	0.86	71	59	0.83
自覚症状	0	5	12724	8835	0.69	2411	1854	0.77	377	326	0.86	59	50	0.85
自覚症状	0	6	8846	5484	0.62	1976	1357	0.69	343	293	0.85	53	43	0.81
自覚症状	1	0	19239	11191	0.58	4602	3050	0.66	832	709	0.85	97	94	0.97
自覚症状	1	1	46922	40305	0.86	6719	5592	0.83	906	830	0.92	98	96	0.98
自覚症状	1	2	46049	38839	0.84	6700	5554	0.83	906	830	0.92	98	96	0.98
自覚症状	1	3	44772	37922	0.85	6480	5353	0.83	891	810	0.91	98	96	0.98
自覚症状	1	4	45325	37684	0.83	6693	5540	0.83	906	830	0.92	98	96	0.98
自覚症状	1	5	43948	36633	0.83	6472	5337	0.82	891	810	0.91	98	96	0.98
自覚症状	1	6	41003	33822	0.82	6250	5042	0.81	882	794	0.90	98	96	0.98
自覚症状	2	0	9253	4740	0.51	2566	1584	0.62	562	377	0.67	82	79	0.90
自覚症状	2	1	38234	33756	0.88	5170	4388	0.85	740	618	0.84	92	87	0.95
自覚症状	2	2	35941	30299	0.84	5120	4305	0.84	740	618	0.84	92	87	0.95
自覚症状	2	3	35103	29835	0.84	4956	4184	0.84	717	598	0.83	92	87	0.95
自覚症状	2	4	34499	28387	0.82	5090	4247	0.83	740	618	0.84	92	87	0.95
自覚症状	2	5	33534	27614	0.82	4923	4120	0.84	717	598	0.83	92	87	0.95
自覚症状	2	6	30759	24836	0.81	4646	3843	0.83	696	569	0.82	92	87	0.95
既往	0	0	3379	2106	0.62	790	532	0.67	161	111	0.69	32	19	0.59
既往	0	1	30350	26980	0.89	3745	3363	0.90	472	414	0.88	71	56	0.79
既往	0	2	25222	20994	0.83	3533	3038	0.86	465	404	0.87	69	52	0.75
既往	0	3	21390	17371	0.81	3144	2668	0.85	414	362	0.87	56	44	0.79
既往	0	4	21750	16631	0.76	3390	2838	0.84	461	397	0.86	69	52	0.75
既往	0	5	17017	12568	0.74	2911	2359	0.81	406	349	0.86	56	44	0.79
既往	0	6	12964	8989	0.69	2560	1954	0.76	367	307	0.84	53	42	0.79
既往	1	0	23102	15583	0.67	4739	3277	0.69	827	699	0.85	97	94	0.97
既往	1	1	48119	42161	0.88	6582	5498	0.84	896	817	0.91	98	96	0.98
既往	1	2	47305	40784	0.86	6560	5459	0.83	896	817	0.91	98	96	0.98
既往	1	3	46844	40499	0.86	6451	5374	0.83	887	804	0.91	98	96	0.98
既往	1	4	46738	39832	0.85	6557	5453	0.83	896	817	0.91	98	96	0.98
既往	1	5	46208	39444	0.85	6446	5364	0.83	887	804	0.91	98	96	0.98
既往	1	6	43482	36711	0.84	6253	5095	0.81	882	794	0.90	98	96	0.98
既往	2	0	13350	8639	0.65	2845	1931	0.68	566	394	0.70	88	76	0.86
既往	2	1	40149	35853	0.89	5176	4478	0.87	731	611	0.84	92	84	0.91
既往	2	2	38057	32690	0.86	5132	4403	0.86	731	611	0.84	91	82	0.90
既往	2	3	37753	32397	0.86	5075	4377	0.86	717	598	0.83	91	82	0.90
既往	2	4	36738	30883	0.84	5111	4364	0.85	731	611	0.84	91	82	0.90
既往	2	5	36374	30557	0.84	5051	4333	0.86	717	598	0.83	91	82	0.90
既往	2	6	33648	27715	0.82	4791	4053	0.85	697	565	0.81	91	82	0.90
身長・体重	0	0	7635	4020	0.53	2282	1241	0.54	597	392	0.66	95	91	0.96
身長・体重	0	1	46327	38730	0.84	7054	5733	0.81	933	876	0.94	98	96	0.98
身長・体重	0	2	45863	37892	0.83	7046	5717	0.81	933	876	0.94	98	96	0.98
身長・体重	0	3	37393	32621	0.87	5087	4242	0.83	783	653	0.83	97	95	0.98
身長・体重	0	4	45128	36688	0.81	7031	5688	0.81	932	874	0.94	98	96	0.98
身長・体重	0	5	35668	30119	0.84	5051	4182	0.83	782	651	0.83	97	95	0.98
身長・体重	0	6	31829	26180	0.82	4694	3785	0.81	757	612	0.81	97	95	0.98
身長・体重	1	0	77778	62845	0.81	9666	8346	0.97	992	984	0.99	100	100	1.00
身長・体重	1	1	89452	81268	0.91	9840	9686	0.98	996	992	1.00	100	100	1.00
身長・体重	1	2	89429	81222	0.91	9840	9686	0.98	996	992	1.00	100	100	1.00
身長・体重	1	3	87748	78678	0.90	9821	9648	0.98	996	992	1.00	100	100	1.00
身長・体重	1	4	89399	81163	0.91	9839	9684	0.98	996	992	1.00	100	100	1.00
身長・体重	1	5	87710	78605	0.90	9819	9644	0.98	996	992	1.00	100	100	1.00
身長・体重	1	6	86625	76847	0.89	9811	9628	0.98	996	992	1.00	100	100	1.00
身長・体重	2	0	49270	30959	0.63	8651	7576	0.88	979	959	0.98	99	98	0.99
身長・体重	2	1	72093	59609	0.83	9372	8837	0.94	989	978	0.99	99	98	0.99
身長・体重	2	2	71999	59427	0.83	9372	8837	0.94	989	978	0.99	99	98	0.99
身長・体重	2	3	69315	56671	0.82	9259	8646	0.93	988	976	0.99	99	98	0.99
身長・体重	2	4	71874	59192	0.82	9369	8831	0.94	988	976	0.99	99	98	0.99
身長・体重	2	5	69153	56376	0.82	9255	8638	0.93	987	974	0.99	99	98	0.99
身長・体重	2	6	66804	53244	0.80	9197	8529	0.93	987	974	0.99	99	98	0.99
まるめ身長・体重	0	0	2570	595	0.23	1429	439	0.31	535	311	0.58	93	88	0.95
まるめ身長・体重	0	1	41602	34719	0.83	6637	5224	0.79	927	864	0.93	98	96	0.98
まるめ身長・体重	0	2	41023	33671	0.82	6628	5206	0.79	927	864	0.93	98	96	0.98
まるめ身長・体重	0	3	32987	28977	0.88	4510	3676	0.82	751	605	0.81	96	94	0.98
まるめ身長・体重	0	4	40223	32363	0.80	6612	5175	0.78	926	862	0.93	98	96	0.98
まるめ身長・体重	0	5												

アンケートカテゴリ	個人 特性	追跡	N=100629			n=10000			n=1000			n=100		
			分類数 K	ユニークセル数 S ₂	S ₂ /K	分類数 K	ユニークセル数 S ₂	S ₂ /K	分類数 K	ユニークセル数 S ₂	S ₂ /K	分類数 K	ユニークセル数 S ₂	S ₂ /K
まるめ身長・体重	1	3	85794	75499	0.88	9796	9598	0.98	996	992	1.00	100	100	1.00
まるめ身長・体重	1	4	87497	78026	0.89	9815	9636	0.98	996	992	1.00	100	100	1.00
まるめ身長・体重	1	5	85752	75419	0.88	9794	9594	0.98	996	992	1.00	100	100	1.00
まるめ身長・体重	1	6	84559	73490	0.87	9783	9573	0.98	996	992	1.00	100	100	1.00
まるめ身長・体重	2	0	42689	23044	0.54	8420	7182	0.85	977	955	0.98	99	98	0.99
まるめ身長・体重	2	1	67968	54296	0.80	9253	8624	0.93	988	976	0.99	99	98	0.99
まるめ身長・体重	2	2	67850	54068	0.80	9253	8624	0.93	988	976	0.99	99	98	0.99
まるめ身長・体重	2	3	65128	51273	0.79	9137	8428	0.92	987	974	0.99	99	98	0.99
まるめ身長・体重	2	4	67719	53822	0.79	9250	8618	0.93	987	974	0.99	99	98	0.99
まるめ身長・体重	2	5	64956	50963	0.78	9133	8420	0.92	986	972	0.99	99	98	0.99
まるめ身長・体重	2	6	62388	47570	0.76	9067	8298	0.92	986	972	0.99	99	98	0.99
胸部レントゲン	0	0	205	30	0.15	125	33	0.26	62	22	0.35	26	12	0.46
胸部レントゲン	0	1	27170	24524	0.90	3312	2994	0.90	438	367	0.84	68	59	0.87
胸部レントゲン	0	2	20850	16390	0.79	3122	2671	0.86	435	362	0.83	68	59	0.87
胸部レントゲン	0	3	15898	11684	0.73	2615	2156	0.82	355	307	0.86	56	46	0.82
胸部レントゲン	0	4	16102	10616	0.66	2939	2397	0.82	428	349	0.82	68	59	0.87
胸部レントゲン	0	5	10104	5803	0.57	2303	1714	0.74	344	291	0.85	56	46	0.82
胸部レントゲン	0	6	6042	2754	0.46	1780	1116	0.63	298	240	0.81	52	40	0.77
胸部レントゲン	1	0	11866	4686	0.39	4086	2276	0.56	821	679	0.83	99	98	0.99
胸部レントゲン	1	1	42604	35904	0.84	6492	5171	0.80	904	824	0.91	99	98	0.99
胸部レントゲン	1	2	41898	34698	0.83	6481	5151	0.79	904	824	0.91	99	98	0.99
胸部レントゲン	1	3	40854	34007	0.83	6289	4968	0.79	891	799	0.90	99	98	0.99
胸部レントゲン	1	4	41124	33496	0.81	6473	5136	0.79	904	824	0.91	99	98	0.99
胸部レントゲン	1	5	40007	32689	0.82	6281	4953	0.79	891	799	0.90	99	98	0.99
胸部レントゲン	1	6	36615	29235	0.80	6063	4668	0.77	884	786	0.89	99	98	0.99
胸部レントゲン	2	0	4197	1324	0.32	1795	762	0.43	537	340	0.63	88	78	0.89
胸部レントゲン	2	1	34563	30806	0.89	4732	3858	0.82	733	599	0.82	93	87	0.94
胸部レントゲン	2	2	32592	27701	0.85	4700	3805	0.81	733	599	0.82	93	87	0.94
胸部レントゲン	2	3	32008	27274	0.85	4568	3724	0.82	706	568	0.80	92	85	0.92
胸部レントゲン	2	4	30911	25352	0.82	4667	3745	0.80	732	597	0.82	93	87	0.94
胸部レントゲン	2	5	30226	24840	0.82	4532	3658	0.81	705	566	0.80	92	85	0.92
胸部レントゲン	2	6	26646	20947	0.79	4220	3313	0.79	687	541	0.79	92	85	0.92
胃レントゲン	0	0	569	152	0.27	281	107	0.38	108	46	0.43	35	23	0.66
胃レントゲン	0	1	26930	24062	0.89	3387	3014	0.89	470	387	0.82	71	59	0.83
胃レントゲン	0	2	20896	16530	0.79	3148	2620	0.83	468	385	0.82	70	57	0.81
胃レントゲン	0	3	16876	12716	0.75	2711	2198	0.81	398	327	0.82	62	52	0.84
胃レントゲン	0	4	16533	11246	0.68	2954	2368	0.80	458	369	0.81	70	57	0.81
胃レントゲン	0	5	11854	7590	0.64	2388	1775	0.74	386	313	0.81	62	52	0.84
胃レントゲン	0	6	8021	4532	0.57	1896	1251	0.66	346	267	0.77	59	49	0.83
胃レントゲン	1	0	15394	6708	0.44	4589	2855	0.62	829	704	0.85	97	95	0.98
胃レントゲン	1	1	45546	37820	0.83	6822	5612	0.82	912	848	0.93	97	95	0.98
胃レントゲン	1	2	44785	36449	0.81	6811	5590	0.82	912	848	0.93	97	95	0.98
胃レントゲン	1	3	44229	36155	0.82	6702	5507	0.82	907	838	0.92	97	95	0.98
胃レントゲン	1	4	43816	35038	0.80	6800	5569	0.82	912	848	0.93	97	95	0.98
胃レントゲン	1	5	43190	34649	0.80	6698	5481	0.82	907	838	0.92	97	95	0.98
胃レントゲン	1	6	40035	31532	0.79	6446	5177	0.80	897	822	0.92	97	95	0.98
胃レントゲン	2	0	5854	2198	0.38	2226	1064	0.48	571	393	0.69	87	78	0.90
胃レントゲン	2	1	36032	31512	0.87	5082	4124	0.81	764	651	0.85	92	86	0.93
胃レントゲン	2	2	33830	28004	0.83	5050	4068	0.81	764	651	0.85	92	86	0.93
胃レントゲン	2	3	33513	27733	0.83	4984	4029	0.81	751	636	0.85	91	85	0.93
胃レントゲン	2	4	32069	25695	0.80	4992	3970	0.80	764	651	0.85	92	86	0.93
胃レントゲン	2	5	31672	25352	0.80	4920	3921	0.80	751	636	0.85	91	85	0.93
胃レントゲン	2	6	28322	21655	0.76	4621	3619	0.78	727	604	0.83	91	85	0.93
健康診査	0	0	187	47	0.25	104	37	0.36	47	21	0.45	16	7	0.44
健康診査	0	1	22934	19579	0.85	3035	2745	0.90	385	333	0.86	59	48	0.81
健康診査	0	2	16044	12350	0.77	2534	2100	0.83	373	313	0.84	58	46	0.79
健康診査	0	3	12250	8708	0.71	2161	1721	0.80	326	281	0.86	49	40	0.82
健康診査	0	4	10569	6270	0.59	2287	1765	0.77	357	285	0.80	58	46	0.79
健康診査	0	5	6498	3501	0.54	1726	1123	0.65	305	248	0.81	49	40	0.82
健康診査	0	6	3334	1530	0.46	1166	573	0.49	263	200	0.76	44	35	0.80
健康診査	1	0	6142	2157	0.35	2595	1205	0.46	673	478	0.71	95	91	0.96
健康診査	1	1	37800	32898	0.87	5513	4393	0.80	824	705	0.86	96	93	0.97
健康診査	1	2	35450	29225	0.82	5473	4320	0.79	824	705	0.86	96	93	0.97
健康診査	1	3	34319	28474	0.83	5288	4138	0.78	814	687	0.84	96	93	0.97
健康診査	1	4	33972	27282	0.80	5450	4277	0.78	823	703	0.85	96	93	0.97
健康診査	1	5	32768	26425	0.81	5265	4095	0.78	813	685	0.84	96	93	0.97
健康診査	1	6	29709	23436	0.79	4948	3746	0.76	796	661	0.83	96	93	0.97
健康診査	2	0	2120	718	0.34	966	340	0.35	362	183	0.51	78	64	0.82
健康診査	2	1	31509	28241	0.90	4139	3530	0.85	624	489	0.78	87	80	0.92
健康診査	2	2	26809	22125	0.83	4021	3334	0.83	622	485	0.78	86	78	0.91
健康診査	2	3	26189	21693	0.83	3888	3246	0.83	598	457	0.76	86	78	0.91
健康診査	2	4	24533	19198	0.78	3939	3215	0.82	620	481	0.78	86	78	0.91
健康診査	2	5	23807	18677	0.78	3804	3123	0.82	596	453	0.76	86	78	0.91
健康診査	2	6	20129	14188	0.70	3487	2805	0.80	565	418	0.74	86	78	0.91
子宮がん	0	0	75	21	0.28	42	9	0.21	20	8	0.40	7	3	0.43
子宮がん	0	1	24116	21309	0.88	2972	2751	0.93	343	302	0.88	55	43	0.78
子宮がん	0	2	15129	11206	0.74	2517	2067	0.82	334	287	0.86	54	41	0.76
子宮がん	0	3	9278	6007	0.65	1868	1406	0.75	290	251	0.87	42	37	0.88
子宮がん	0	4	9348	4894	0.52	2228	1679	0.75	322	269	0.84	54	41	0.76
子宮がん	0	5	3853	1760	0.46	1318	712	0.54	266	220	0.83	41	35	0.85
子宮がん	0	6	1565	634	0.41	721	203	0.28	228	170	0.75	36	32	0.89
子宮がん	1	0	2559	631	0.25	1352	396	0.29	537	293	0.55	94	90	0.96
子宮がん	1	1	33659	30388	0.90	4553	3629	0.80	767	612	0.80	95	92	0.97
子宮がん	1	2	31304	26620	0.85	4509	3553	0.79	767	612	0.80	95	92	0.97
子宮がん	1	3	30644	26126	0.85	4343	3464	0.80	737	576	0.78	95	92	0.97
子宮がん	1	4	29407	24100	0.82	4479	3495	0.78	765	608	0.79	95	92	0.97
子宮がん	1	5	28621	23552	0.82	4304	3389	0.79	735	572	0.78	95	92	0.97
子宮がん	1	6	25523	20354	0.80	3935	3001	0.76	708	534	0.75	95	92	0.97
子宮がん	2	0	701	130	0.19	402	86	0.21	209	65	0.31	76	59	0.78
子宮がん	2	1	29240	26747	0.91	3564	3151	0.88	518	388	0.75	86	78	0.91
子宮がん	2	2	24153	19832	0.82	3443	2949	0.86	516	384	0.74	86	78	0.91
子宮がん	2	3	23584	19249	0.82	3365	2907	0.86	493	372	0.75	86	78	0.91
子宮がん	2	4	21179	15936	0.75	3329	2778	0.83	493	378	0.74	86	78	0.91
子宮がん	2	5	20486	15165	0.74	3232	2709	0.84	490	366	0.75	86	78	0.91