
厚生労働科学研究費補助金

政策科学総合研究事業（統計情報総合研究事業）

死亡統計データベースの作成とその研究利用のあり方に関する研究

（課題番号 H20 - 統計 - 一般 - 001）

平成 21 年度総括・分担研究報告書

研究代表者 安村 誠司

平成 22（2010）年 3 月

目 次

I. 総括研究報告

- 死亡統計データベースの作成とその研究利用のあり方に関する研究…………… 1
(研究代表報告：安村誠司)

II. 分担研究報告

1. 死亡統計の研究利用における統計的課題
—レコードリンケージと匿名化の方法— (橋本修二) …………… 11
2. 人口動態統計の前向き研究への適用と解析用病名作成に関する研究
(岡山 明、大橋靖雄) …………… 19
(資料1) …………… 29
(資料2) …………… 32
(資料3) …………… 35
3. プリオン病サーベイランスを通じたわが国における National Death Index の必要性の検討と、導入に関連する問題点 (中村好一) …………… 36
4. 循環器疾患を対象としたコホート研究における死亡統計の研究利用について
(岡村智教) …………… 39
(資料4) …………… 42
5. 地域がん登録における人口動態死亡統計の利用と手続きに関する研究 (味木和喜子) …… 45
6. デンマークにおける保健医療関連データベースに関する調査研究 (辻 一郎) …………… 48
7. 死亡者データベースの疫学研究における有用性の検討
—諸外国における死亡者データベースの利用状況から— (安村誠司) …………… 54

I. 総括研究報告

厚生労働科学研究費補助金（統計情報総合研究事業）
総括研究報告書

死亡者データベースの作成とその研究利用のあり方に関する研究

研究代表者 安村誠司 福島県立医科大学医学部公衆衛生学講座教授

研究要旨 研究データと死亡データのレコードリンケージにおいて、死亡情報のデータベース化とレコードリンケージの技術面にはそれほど大きな問題がないと考えられ、個人情報を除く死亡データの医学研究への利用には、匿名化よりもオーダーメイド集計などの検討を進めることが大切であろう。

大規模コホート（JALS）における死因確定の業務を通じ、①人口動態統計申請から許可までの作業上の課題の把握（申請側）、②上記対象地域において実施された追跡開始時調査受診者で死亡が把握されたものについて人口動態調査の調査票情報と照合し原死因を確定する作業、③原死因を確定する際に、人口動態調査の調査票情報と住民票から得た情報の一致率と誤差要因の検討、④人口動態調査情報から得た原死因から、疫学的研究に必須の事項に絞った二次病名の分類基準の作成を行った。

厚生労働科学研究費補助金 難治性疾患克服研究事業プリオン病及び遅発性ウイルス感染症に関する調査研究班で実施しているプリオン病サーベイランスのデータを用いて、もし、日本にNDIシステムが存在し、本データセットに応用していたら、どの程度登録データが改善するか、という点を明らかにした。加えて、日本にNDIシステムを導入する際の課題も検討した。

コホート研究の維持のためには、住民票請求や人口動態統計による網羅的な追跡は重要であり、国民のプライバシーは最大限保護されるべきであるが、長期的な国民の利益のためには、統計情報の学術研究への柔軟な利用について、今後、法整備も含め検討していく必要があるものと考えられる。

地域がん登録事業において、がんの罹患数・率を計上し、その信頼性を評価し、がん患者の生存率を計測するために、人口動態死亡統計の利用が不可欠である。改正統計法下での地域がん登録事業からの標準的な申請書を作成し、承認を得た。がん登録資料の適正な整備と利用が図られるよう、引き続き、関係機関との調整を進めていく。

デンマーク政府機関による保健医療関連データベースのなかで使用頻度が最も多い5種類のデータベースについて、その構築までの経緯・歴史、システムの概要、データ管理を行う機関、年間登録数、個人情報保護の手続き、研究者が利用申請してから承認されるまでの手続き・期間、実際の研究活用の事例などについて、調査を行った。デンマークでは、国の機関がデータベースを管理・監督し、研究者に短期間でデータを提供するシステムが確立されている。これにより多くの有用な研究成果が世界に向けて発信され、人々の健康と福祉に貢献していた。

死亡者データベースの疫学研究における有用性を検討する目的で、文献データベース（pubmed）を用いて、死亡者データベースを用いた研究を検索した。検索された文献の検討から、死亡者データベースの導入により、導入された国においては、研究が振興していると考えられた。

分担研究者

橋本修二・藤田保健衛生大学医学部衛生学講座・教授

岡山 明・財団法人結核予防会第一健康相談所・所長

大橋靖雄・東京大学大学院医学系研究科公共健康医学・教授

中村好一・自治医科大学地域医療学センター公衆衛生学部門・教授

岡村智教・国立循環器病センター予防検診部・部長

味木和喜子・国立がんセンターがん対策情報センターがん情報・統計部・室長

辻 一郎・東北大学大学院医学系研究科公衆衛生学分野・教授

A. 研究目的

医学研究において、死亡統計の利用はきわめて重要である。本研究の目的は、死亡統計の研究利用における統計的課題を明らかにするとともに、その検討を行うことにある。本年度は人口動態調査死亡票データに対して、研究データとのレコードリンケージの方法と匿名化の方法について検討した。

本研究では、実際の疫学研究で人口動態調査を活用する過程を追従し、疫学研究で人口動態調査データに使用していく上での課題を明らかにすることにした。この際に、多くの研究者が関与し、地域も多地域にわたる大規模な研究を扱う方が様々な課題が明らかになると考えた。そこで、全国 33 研究施設が、地域住民を対象とした動脈硬化性疾患（脳卒中、虚血性心疾患、閉塞性動脈硬化症等）の発症と死亡の追跡を行っている Japan Arteriosclerosis Longitudinal Study (JALS) を検討対象に、新統計法のもとで人口動態調査死亡データの目的外使用申請から照合作業を実際に行い、人口動態統計を研究活用する際の課題を検討した。

疫学研究において特定の疾患の罹患者を追跡する様な研究では、場合によっては対象者が全国におり、死亡小票の閲覧は現実問題として不可能である。また、多くの臨床研究では、患者の追跡を行う場合に、観察を行っている医療機関を受診している患者だけを追跡し、当該医療機関を受診しなくなった場合、最終受診日で観察打ち切りとするのが通例である。このような場合に、米国の National Death Index (NDI) のような仕組みが日本にも存在すれば、対象者の生死の確認と死亡に関するデータの入手が容易になり、研究は促進する。本研究では日本におけるプリオン病の登録研究を通じて、もし NDI があればどの程度登録データが改善されるかを示し、併せて日本に NDI を導入する際の問題点を議論する。

日本でにおける循環器疾患をエンドポイントとした研究では、発症をエンドポイントとし、1) 研究を市町村事業の一環として研究を取り入れ、市町村の業務の提携先、受託先として異動情報入手、2) エンドポイントは発症とし、悉皆的な質問紙調査と近隣の医療機関での診療録調査（または情報提供）、3) コホートによっては保健所の許可を得て死亡小票を閲覧、という手法をとってきた。このモデルは非学術的な外部要因の影響を受けやすいという問題や、質問紙調査への未回答や診療録閲覧の同意が得られない、保健所管轄外転出者などの多くが登録漏れになるなど種々の問題を抱えてい

る。したがってより悉皆性の高い追跡や全国的なコホート研究を行なう際には、人口動態統計を用いた追跡システムを構築する必要がある。

このような状況で、本研究では循環器疾患の発症情報と人口動態統計の死因とを突合可能な吹田研究のコホートデータを用いて、人口動態統計における死因利用の有用性を検討した。

第 3 次対がん総合戦略事業開始前においては、各県が死亡情報を独自の方法で利用し、各県からの人口動態死亡目的外申請の記載内容が統一されておらず、承認を得るために多大な労力と時間を要した。地域がん登録の標準化を目指す第 3 次対がん「がん罹患・死亡の実態把握」班（研究代表者：祖父江友孝）では、平成 19 年度において、地域がん登録からの人口動態死亡目的外申請のひな形となる記載方法を山形県の協力を得て作成し、承認を得た。平成 20 年度は、他県において、山形申請書の記載に準拠して申請手続きを進めるとともに、改正統計法全面施行後にも、同様の利用が可能となるよう、その手続きについて検討を進めた。本年度は、改正統計法下での申請手続きを実際に進めた。

政府機関による保健医療データベース構築に関してもっとも進んだ国の一つであるデンマークについて、調査を実施している。本年度は、使用頻度がもっとも多い 5 つの主要な登録データベース（がん登録、死因登録、患者登録、出産登録、出生力）について、データベース構築までの経緯・歴史、システムの概要、データ管理を行う機関、年間登録数、個人情報保護の手続き、研究者が利用申請してから承認されるまでの手続き・期間、実際の研究活用の事例などを調査した。

すでに死亡データベースを有している米国、デンマーク、オーストラリア、ブラジル、韓国におけるデータベースを用いた研究の実施状況を明らかにし、死亡データベース導入の有効性を下記らかにすることを目的とした。また、韓国における死亡データベースの概要についても一部紹介する。

B. 研究方法

1. 人口動態調査死亡票の項目について、人口動態調査オンライン報告システムの入力状況、および、人口動態調査の磁気媒体への転写状況を整理した。個人識別情報の入力・転写状況ごとに、研究データとのレコードリンケージの方法を検討した。匿名化の方法、および、統計データへの適用状況を整理するとともに、人口動態統計の死亡データの匿名化の方法を議論した。

2. 検討対象となるコホート研究 JALS と人口動態統計目的外使用申請

1) JALS は、平成 13 年から開始された公益信託日本動脈硬化予防研究基金（主務官庁：厚生労働省）による研究であり、全国各地の循環器コホート研究の個人データ統合し、循環器疾患発症リスクとリスク因子の影響を定量的に評価することを目的としている。全国 33 研究施設が参加し、2002 年より開始し、2006 年 3 月末で終了している（118,239 名（男性 54,349 名、女性 63,890 名））。ベースライン調査以後、動脈硬化性疾患（脳卒中、虚血性心疾患、閉塞性動脈硬化症等）の発症と死亡に関する観察を行っており、2010 年 1 月末現在、死亡 3,351 例、脳卒中発症 1,619 例、急性心筋梗塞 354 例、急性死 145 例が登録されている。

2) 人口動態統計目的外使用申請

人口動態統計目的外使用の申請に必要な資料を作成した。その後、人口動態統計目的外使用の申請を行った。

3) 人口動態調査の調査票情報と照合し原死因の確定

①照合作業前の準備として、JALS 対象者で死亡が確定している者の性別、生年月日、死亡年月日、居住市町村名（コード）などのリストをあらかじめ作成した。

②照合作業では、人口動態統計使用許可後、性、生年月日、死亡年月日、死亡前の居住市町村名をキー変数とし、人口動態統計データと照合し原死因を確定した（第 1 回照合）。この作業で、照合不能であった例については、各コホートの研究者にデータの照合を行った。その後、データの修正事項を反映させ、再度照合作業を行った（第 2 回照合）。2 回の照合作業を実施しても死因を結合できなかった例については、さらに検索を行い、該当候補の絞込みを行った（検索水準を落とした目視検索）。上記作業を通じ、照合に影響を与える要因を検討した。

4) 人口動態調査情報から得た原死因から、二次病名の分類基準の作成

コホート研究として必要な病名の集約のあり方について検討した。基本死因から循環器疾患及び癌関連、腎疾患、外因死を抽出し、これ以外の死因をその他の死因とした。対象地区の対象期間の全死因（346424 件）と照合し、死因の上位 50 位以内の単独死因として抽出されている死因と比較した。50 位以内で生活習慣病関連死因を検索して、新たな死因区分の必要性を検討して、死因区分を作成した。

3. 厚生労働科学研究費補助金 難治性疾患克服研究事業プリオン病及び遅発性ウイルス感

染症に関する調査研究班では 1999 年よりヒトのプリオン病（クロイツフェルト・ヤコブ病、ゲルストマン・ストロイスラー・シャインカー病、致死性家族性不眠症）の登録（「サーベイランス」という名称だが、実質は登録者の死亡まで追跡）を行っている。特定疾患臨床調査個人票や感染症法による届け出を端緒の情報とし、全国の専門医の協力を得て情報収集し、年 2 回のサーベイランス委員会で、プリオン病と認めたものを登録している。その後は半年に 1 度、主治医に調査票を送付し、現状報告を求めている。登録された患者は死亡が確認されるまで追跡（主治医への現状報告の依頼）を行っている。本研究では、プリオン病登録において、解析段階で死亡が確認されていない対象者の割合を明らかにし、その上でこれらの者の最終の状況確認年月日から解析時点までの経過日数の分布を明らかにし、もし、日本に NDI システムが存在し、本データセットに応用していたら、どの程度登録データが改善するか、という点を明らかにする。

4. 吹田研究は、平成元年に吹田市の住民台帳より 12,200 名を無作為抽出し、その中で同意が得られた 6,485 名が一次コホート集団として設定されている。

1) 吹田研究の概要

本研究は都市部における循環器コホート研究である。ベースライン調査（初回健診）は 1989 年から 2 年ごとに対象者の循環器健診を老人保健法に基づく基本健康診査として実施しており、それに合わせて糖負荷検査、アンケート調査、頸部超音波検査などの研究検査を本人の同意のもとに実施している。現在、1989～1993 年の初回受診日をベースラインとし、その後の心筋梗塞および脳卒中の発症者の追跡調査を継続中である。

2) 心筋梗塞・脳卒中発症者の確認方法

2 年ごとに実施している循環器健診の際に各対象者の既往歴を確認している。また、年に 1 回、質問紙による調査を全員に行っている。こ発症時に受診した病院のカルテ調査について同意を得た上で、カルテ調査を行っている（カルテ調査同意率 86.2%）。

3) 人口動態統計の目的外利用申請と人口動態統計有用性の検討

吹田研究では人口動態の目的外利用申請を行い、死亡した追跡対象者の原死因を把握している。発症データと人口動態統計の原死因データとを突合し、人口動態統計の死亡情報の利用状況などその有用性を検討した。

5. 「統計法第 33 条の運用に関するガイドライ

ン」に基づいて、人口動態調査票の利用申請への記載事項が2009年4月改訂された。昨年度中に山形申請書に基づく申請書を提出し、承認を得る前に改正統計法の適用となった栃木県の協力を得て、改正統計法下での申請書を作成し、承認を得た。

6. デンマークの5つのデータベースについて、以下の項目を検討する。

- 1) データベースが構築されるまでの経緯・歴史、
- 2) システムの概要：情報の収集・データベース化に至るまでの手続き
- 3) データを管理している機関（ホームページアドレス等の情報の所在）
- 4) 年間登録数
- 5) 個人情報保護の手続き
- 6) 研究者が利用を申請してから承認されるまでの手続き、期間
- 7) 実際の研究活用の事例

上記の点を明らかにすることで、デンマーク国家規模で実施されている登録データベースがどのように研究利用されているのかを把握できる。さらに、研究利用の観点から日本とデンマークの登録データベースとの差異を明らかにし、日本の登録データベースを研究利用する際の問題点を明確にする。

7. 文献データベース (pubmed) を用い、次の二種類のキーワードで検索した。

- 1) "national death index" AND (Australia OR australian)

下線部は各国を示す用語とする(オーストラリア、アメリカ合衆国、デンマーク、日本、韓国、ブラジル)

- 2) (national OR nationwide) AND (database) AND (death OR "cause of death" OR "cause death" OR "death certificates" OR mortality) AND (Australia OR australian) これは、辻の方法にならった。
- 3) 検索した文献のうち、「死亡を結果としないもの、フィールド・データが外国のもの、総説・論説、invitro・動物の実験」を除いた論文の抄録を読み、論文を分類した。
- 4) 韓国における死亡者データベースについては、研究協力者の李相侖氏の協力により実施した。ネット上から得られた情報については、李相侖氏の翻訳をもとに分析した。また、韓国統計庁社会統計局人口動向課イ ジョン事務官へは李相侖氏が電話による取材を行い、得られた情報を翻訳した情報を分析した。

C. 研究結果

1. 人口動態調査死亡票の項目

1) 人口動態調査オンライン報告システムでは、死亡票の19項目が入力され、70変数に及んでいた。個人識別情報として、氏名(文字;漢字)、性、生年月日、住所(文字;丁目や番地を含む)が含まれていた。人口動態調査の磁気媒体では、性、生年月日、住所(市町村符号)が含まれ、氏名や詳しい住所は含まれていない。死亡に関する項目としては、死亡したとき、死亡したところ、および、死亡の原因であった。

2) レコードリンケージの方法

研究データの人口動態調査死亡票データとのレコードリンケージの実施可能性では、研究データが死亡の事件簿番号、死亡の有無と死亡年月日を含んでいれば、人口動態調査のオンライン報告システムの入力データ、人口動態調査の磁気媒体の転写データともに、レコードリンケージが実施可能である。死亡の事件簿番号がなくとも、死亡の有無と死亡年月日を含んでいれば同様である。研究データが死亡の事件簿番号、死亡の有無と死亡年月日を含んでいなくとも、氏名と住所(詳細)によって、人口動態調査のオンライン報告システムの入力データとはレコードリンケージがほぼ実施可能(転居がない場合)である。人口動態調査の磁気媒体の転写データには氏名と住所(詳細)がないために、レコードリンケージが実施可能でない。

3) 匿名化の方法

匿名データを作成された4統計の匿名化の方法として、レコードのリサンプリング(レコードの一部を削除)は4統計とも実施され、リサンプリング率が3統計で80%、1統計調査(標本の抽出率が高い)で10%であった。識別情報の削除は調査区番号等の削除とレコードの無作為な並び替えが実施された。匿名化の方法として、誤差の導入、データスワッピングは適用されなかった。

2. 人口動態統計目的外使用申請、及び、現死因の確定

1) 申請前の準備、訪問、申請

2009年7月31日に、安村、岡山、原田(研究協力者)で、厚生労働省大臣官房統計情報部企画課審査解析室を訪問し、人口動態統計目的外使用の申請書を提出した。

2) 申請作業

仮申請日当日、厚生労働省大臣官房統計情報部企画課審査解析室統計審査第一係担当官より、E-mailにて初回の照会事項の送付を受けた。JALS側の実質の対応は、同年9月末より開始となったが、以後E-mailによるやりとり

を開始した。照会・返答の事項は、①申請書に記載の事項については、死因情報を付加したデータの利用方法、利用場所、利用する環境、保管場所及び管理方法、転写書類（転写 CD-R）及び死因情報を付加したデータの利用後の処置に関する修正や追記であった。データの使用者、管理方法（セキュリティー）、利用後の処置などについて具体的な記載を求められた。②申請書以外の提出書類についての指示であった。2009年12月21日に内諾が得られた後、正式書類を厚生労働省に提出した。

3) 許可

2010年1月13日に正式な使用許可がおりた。受領した人口動態統計死亡データは、JALS 当該市町村に限定した2002年1月1日から2008年12月31日までに死亡した346,429件分であった（5件の重複データを含んでいたため削除して、以後の検討に用いた）。

4) 照合による原死因の確定

JALS 研究の対象者で死亡が確定している3,351例のうち、今回申請対象に含まれない職域コホート分、期間内の死亡調査データが確定していないコホートを除いた3,258例を検討対象とした。性、生年月日、死亡年月日、死亡前の居住市町村名をキー変数とし、人口動態統計死亡データと照合作業を行った。

①第1回照合作業により、人口動態統計データと一致が見られたのは3,144件（96.9%）であった。結合不能であった100件については、JALS の各コホート研究者に対し照会を行った。修正が必要な情報は、生年月日6件、死亡日25件、性別1件、市町村情報13件であり、死亡日に関する誤りが多かった。情報に間違いなしと報告があったものが37件、期日までの確認が難しかったものが18件であった。

②第2回照合作業を行ったところ、人口動態統計データと一致が見られたのは3,194件（98.5%）であ、結合不能例は50件となった。

5) 検索水準を落とした目視検索

結合不能例について、対象市町村ごとに、目視にて該当候補の絞込みを行った。照合不能例65件に対し、抽出された該当候補を照合不能の原因として考えられるエラー別に分類した。コホートデータと人口動態統計のデータで月と日が逆になっていたもの、生年月日の月日と死亡日の月日が同じ入力であったものが各1件、死亡の月日を人口動態の生年月日の”月日”に入力していたものが5件、性別違いが6件、年月日にズレがあったもの9件、明らかな入力間違い6件、元号変換ミスと思われるものが1件みられた。目視によっても該当候補が抽出で

きなかったものが21件であった。抽出された該当候補については、各コホート研究者に再度照会し、精査したのち、コホート側の間違いが否定できたものについては、両者を同一人物のデータと判断して死因を照合する予定である。

3. 解析時点（2009年11月27日現在）までに登録されたプリオン病の患者は1,305人で、うち1045人（80.1%）が死亡し、残りの260人（19.9%）の最終報告時の状況は、通院25人（1.9%）、入院198人（15.2%）、在宅15人（1.1%）、不明22人（1.7%）であった。現状報告がなされてから解析時点までの経過時間（日数）の分布は、0～199日81人（死亡が確認されていない者の31.2%）、200～299日36人（13.8%）、300～399日18人（6.9%）、400日以上125人（48.1%）であった。最大は4077日（状況は「不明」）であった。

プリオン病の患者の90%以上が発病後1年以内に死亡するとされている。追跡調査は半年に1度の頻度で行っているため、現状報告から解析時点まで199日以下の者は適切に報告されており、生存していると考えても大きな問題は生じない。一方、400日以上の方は追跡不能になっていると考えられ、また、プリオン病の生命予後を考えると多くの者は死亡していると推測されるので、もしNDIシステムが日本に存在すればこれらの者の中で死亡している者については、死亡の事実と死亡年月日が確認できるであろう。400日以上の方と期間不明の方の死亡が確認されたとすると、死亡1170人（89.7%）、通院23人（1.8%）、入院102人（7.8%）、在宅10人（0.8%）となる。

4. 吹田研究では、使用した人口動態統計は、平成18年度末までの異動情報の確認によって死亡が把握された吹田研究対象者の1165名である。1165名のうち原死因が心筋梗塞の者は94人、脳卒中の者は103人（脳梗塞60人、脳出血22人、くも膜下出血11人、その他10人）であった。一方、コホート研究のエンドポイントとしては心筋梗塞が124人で、54人がdefinite（確実例）（カルテ調査により検査データ等からMONICAプロジェクトの基準を満たす者）、70人がprobable（ほぼ確実例）（カルテ調査の同意が得られず、人口動態統計の死因データなどから情報を得た者）であった。また、脳卒中は214人で、155人がdefinite（確実例）

（カルテ調査により画像データ等からThe National Survey of Strokeの基準を満たす者、内訳：脳梗塞111人、脳出血27人、くも膜下出血14人、その他3人）、59人がprobable（ほぼ確実例）（カルテ調査の同意が得られず、人

口動態統計の死因データなどから情報を得た者、内訳：脳梗塞 31 人、脳出血 5 人、くも膜下出血 8 人、その他 15 人) であった。コホート研究の発症データベースと人口動態統計の死因のデータベースを突合した結果、コホート研究で心筋梗塞 definite と判定されている 54 人のうち死亡動態統計に含まれるのは 5 人 (9%)、心筋梗塞 probable と判定されている 70 人中では 64 人 (91%) であった。同様に脳卒中 definite と判定されている 155 人のうち 17 人 (11%)、脳卒中 probable と判定されている 59 人のうち 48 人 (81%) が死亡動態統計に含まれていた。

5. 旧統計法では調査年次の翌年末 (例：平成 20 年死亡では平成 21 年 12 月末) まで認められていた利用期間が、改正統計法では承認日より最長 1 年以内となった。

地域がん登録においては、既登録患者の予後情報を得るためには、死亡後速やかなデータ利用が望ましい。一方で、登録漏れの把握のためには、医療機関からの届出情報が出揃うのを待ち、死亡年次の翌年の利用が必要となる。「利用期間が 1 年以内」という改正統計法の制限化では、死亡情報を 2 年に渡って利用する従来の方法が困難となった。そこで、複数年の死亡情報の利用を申請し、それを毎年継続申請する方針とした。一方で、がん診療連携拠点病院の院内がん登録による全国集計のために、2007 年 (平成 19 年) に診断された患者情報が、平成 21 年になって、多数、地域がん登録室に提出された。平成 19 年死亡の利用期限は平成 20 年末で終了していることから、平成 19 年以降の死亡情報の利用を、毎年、申請する方針とした。

6. デンマークの 5 つのデータベースが構築されるまでの経緯等

① Danish Cancer Registry (デンマークがん登録データベース)

Danish Cancer Society (デンマーク対がん協会)、国家保健委員会、Danish Medical Association (デンマーク医師会) が共同し 1942 年に設立した。1987 年 1 月以降には各医療機関は国家保健委員会へのがん登録の報告が義務化された。1997 年 1 月にはがん登録は、デンマーク対がん協会から国家保健委員会に移管され、2004 年 1 月以降、それまでの紙媒体の届出から電子媒体による届出に変更した。本データベースは、全ての新規がん診断情報及び関連疾患の情報を含んでいる。

② Danish Registry of Cause of Death (デンマーク死因登録データベース)

国家保健委員会が 1973 年に設立した。1871 年 1 月以降には各医療機関は死亡届の提出が義務化された。2004 年 1 月以降、それまでの紙媒体での届出から電子媒体での届出に変更した。1974 年 1 月に死因登録は、Statistics Denmark (デンマーク統計局) から国家保健委員会に移管された。本データベースは故人の名前、死亡時の年齢、住所、死亡日、死因、方法、死亡場所の情報を含んでいる。

③ Danish National Patient Registry (デンマーク患者登録データベース)

1977 年に設立され、精神疾患以外の入院患者の全データが登録されている。1995 年以降は、上記に加え、精神疾患患者、外来患者、救急患者の全データが登録されている。全てのデータは、国家規格に基づいた形式で登録されている。本データベースの主な目的は、統計データを提供すること、病院の利用率を監視すること、治療計画過程を援助することである。しかし、近年は、疾病・治療の監視のみならず医学研究や病院部門における医療の質の向上に資するために使用されている。

④ Danish Medical Birth Registry (デンマーク出産登録データベース)

1968 年に設立され、国立保健委員会により維持・継続されている。1973 年より電子化された。本データベースは、社会経済的状況、出生児の医学的状況、死産・出生後死亡に関する情報を含んでいる。

⑤ Danish Fertility Database (デンマーク出生力データベース)

1992 年に設立され、Statistics Denmark (デンマーク統計局) により維持・継続されている。1992 年より電子化されている。本データベースは、デンマークの男女における出生時の男女の特性 (年齢、住所、婚姻歴、家族構成、出産歴)、社会的背景 (教育歴、職業・収入、公的給付金受給の有無、住宅)、子供に関する情報 (出生登録場所、性別、双生児、出生時体重、死亡年齢、死因) を含んでいる。

7. 文献データベースで検索された文献

1) 該当した論文の数

キーワード 1) で該当した論文は全体で、米国では 328、オーストラリアでは 48、ヨーロッパ全体で 5、デンマークで 1 であった。キーワード 2) で該当した論文は全体で、オーストラリアでキーワード 1 で検出された論文と一致したものは 10 件 (うち 1 件はデータ分析でなし)、デンマークで 45、日本で 40、韓国で 27、ブラジルで 18 であった。なお、少なくとも、NATIONAL DEATH INDEX で文献がヒットする国

(米国、オーストラリア)では、死亡状況の把握が「NATIONAL DEATH INDEX によって」と、明確に示されていた。NDI で拾えない国の調査は、同列には論じられないが、これらの国々でも NATIONAL DATABASE から把握・LINKAGE して把握することができる。しかし、社会保険番号など個人識別したかどうかは、抄録からは把握できなかった。

- 2) 死亡者データベースに関する論文の推移
- ①NDI 以前と以後の比較では、米国で論文数は1981 まで1 (1972)、1982 以降 339 であった。NDIplus 導入後、論文数は大きく増加していた。
- ②デンマーク、オーストラリア、ブラジル、韓国いずれの国でも、文献数が増加傾向にある。
- 3) 韓国における死亡者データベースについて
韓国の死亡原因統計 (2007)
- <概要>統計名：死亡原因統計
種類：指定・加工統計、保健分野
法的根拠：家族関係の登録等に関する法律第84 条 統計法指定統計 10154 号
作成目的：国民の正確な死亡規模と原因、地理的分布を把握し人口及び保健福祉政策樹立のための基礎資料を提供
作成周期・期間：毎年、1 月 1 日～12 月 31 日
使用資料：死亡申告書資料 (死亡者住所、生年月日、死亡年月日、死亡場所、当時の職業、婚姻状態、教育水準、死亡原因等)
作成方法：
- ①死亡者発生後、家族関係登録法により 1 ヶ月以内に邑、面、洞の役所および市・区役所の死亡申告書および医師が作成した死亡診断書の提出
- ②邑、面、洞および市・区役所では申告された内容を確認し、ウェブ入力システムである人口動態システムに入力
- ③統計庁は死亡申告入力内容を検討し、死亡関連の各種の行政資料を連携し、死亡申告の不正確な死亡原因を補う
- ④死亡原因は韓国の標準疾病死因分類 (第4 次) により約 2 千コードと分類
- ⑤WHO の勧告案である 103 項目と、そ韓国の状況に合わせて細分化した 236 項目に作表し、公表
- 適用分類：韓国標準疾病・死因分類
<調査方法論>推定技法
<資料提供>公表方法：
KOSIS (国統計ポータル(portal))：1983～最近
MDSS (マイクロデータ)：1991～最近
公表範囲：集計単位；全国、市・都、市・郡・区
公表周期：毎年 9 月
刊行物名：死亡原因統計年報

資料検索：

ア. 報道資料

2007 年の死亡および死亡原因における統計結果. PDF, KOSIS, 死亡原因統計 DB

イ. デジタル刊行物

死亡原因統計 (全国編、地域編)

D. 考察

1. 人口動態調査死亡票データに対して、研究データとのレコードリンケージの方法を議論した。研究データとして、氏名、住所 (詳細)、性と生年月日を含むことを前提とした。研究データが死亡の事件簿番号、死亡の有無と死亡年月日を含まなくとも、人口動態調査オンライン報告システムの入力データとはレコードリンケージがほぼ実施可能 (転居がない場合) であった。これは、住民基本台帳や死亡小票などと照合しなくとも、レコードリンケージがほぼ実施可能であることを意味する。National Death Index に近いシステムとも言えよう。死亡情報のデータベース化とレコードリンケージの技術面には、現在、それほど大きな問題がないと考えられる。

匿名化の方法については、最近、日本でもガイドラインとしてまとめられている。現在、利用可能な匿名データは、いずれも、標本抽出を用いた統計調査である。匿名化の方法としては、匿名性の確保をやや優先している印象を受ける。匿名性の担保の必要性はいうまでもないが、データの利用性をさらに拡大する余地が小さくない。匿名化されたデータの利用拡大に向けて、多くの経験と実績の積み重ねが基礎であり、今後、医学研究での寄与が大いに期待される。

死亡統計の医学研究への利用にあたっては、主として全数データが前提となる。一方、匿名化の確保にあたっては、全数でなく、抽出が最も重要な要件の 1 つと位置づけられる。死亡情報の匿名データの作成は、かなり限定的なことを考える。むしろ、死亡データの二次利用には、オーダーメイド集計がより適切と考えられ、今後の重要な検討課題であろう。

2. 疫学研究で人口動態調査データを使用していく上での課題を明らかにするため、現在進行中の大規模疫学研究を題材に、人口動態調査データの申請から死因の照合までの過程を追従し検討した。申請から死因照合までの過程を通して、多施設で実施の研究ならではの課題も多くみられた。今回、国内で比較的規模の大きい本研究で、現行の制度下で、申請から死因照合までの一連の作業を行えたということは、少数の研究者、限れた地域で実施される研究におい

ては、実施・運用のハードルはこれより低いものになると考えられた。

1) 人口動態統計目的外使用申請

実質の審査期間は3ヶ月であった。上述の申請書が基本的な体裁を充足していることを前提とした場合、3ヶ月という審査期間はある程度参考となる数値と思われる。

2) 人口動態調査の調査票情報との照合

死因照合の作業では、多施設研究ならではの課題、人口動態統計側の問題点もみられた。不一致の原因としては、追跡精度、追跡体制の違いなどコホート側の要因が大きい。第一に、転出者の把握が十分でない可能性が考えられる。検索は、市町村を限定して実施しており市外転出者は検索されない。このため、転出後も追跡を行っているコホートでは、事務局に対し転出の情報を報告せずに、転出情報の報告前に死亡の報告を登録している可能性も考えられる。第二に、JALS 研究では死亡情報の確認に、住民基本台帳（住民票）での確認を必須としているが、それ以外のデータに基づく登録を行っている可能性が考えられる。ルールを遵守して収集されたデータのみで再度検討したならば、人口動態統計データとの照合率は、99%近くに達することが予想される。

一方、人口動態統計側の問題もいくつかみられた。死亡日の“月日”が、人口動態統計の生年月日の“月日”に入力されている例が散見された。市町村における入力段階でのエラーと考えられる。人口動態調査票の報告は、OCR 調査票による報告からオンライン報告システムに切り替わってきた経緯があり、OCR の読み取りエラーから、市町村担当者が直接入力することでのエラーへの対策も必要と思われた。人口動態統計では、生年月日入力のエラーチェックの方法について検討する必要がある。

3) 二次病名

コホート研究では分析仮説を立てるのはデータセットが完成してからの場合も多い。そうした場合、原病名が不明であれば、再分類がきわめて困難となる。しかし、原死因を保持することは匿名化されていてもまれな疾患の場合にはデータ管理に強い制限を受ける。また、本研究のような大規模なデータセットであっても、検出力を考慮すると、コホート研究を行う際の対象疾患として最低でも 0.5% (15 例) 程度の頻度の疾患である必要がある。従来の対象疾患に加え、糖尿病死、腎疾患以外の腎症を含む腎死、肝疾患及び肝硬変を対象とするのが望ましいと考えられた。人口動態統計の照合結果は研究期間を満了したとき削除することが求

められている。しかし、データの公衆衛生的価値は一定の期間にとどまらない。死因情報の二次病名化を考えることはできないだろうか。さらには今後発展が期待される分野の二次病名についても検討する必要があるだろう。

3. 日本に NDI システムが存在すれば、現在のプリオン病サーベイランスの結果において、死亡者の割合が 10% 弱上昇することを示した。日本では、地域単位のコホート研究では、厚生労働省の許可のもとに保健所に保管されている死亡小票などから対象者の生死と死亡の場合には死亡に関する情報を入手している。しかし、プリオン病サーベイランスのように対象者が全国に点在し、また、個人情報保護の観点から制限された個人情報（氏名 [イニシアルのみ]、生年月日、性別、住所 [都道府県名のみ]）しか収集していない状況では、保健所の死亡小票を用いた生死の確認と死亡情報の収集は不可能である。ここで NDI システムがあれば、利用できる情報は生年月日、性別、都道府県名のみだが、生死（死亡届の提出）の確認と死亡情報の入手が可能となり、より正確なデータベースを構築することが可能となることが示された。

NDI を日本に導入することに対する障壁として、個人情報保護法である。個人情報保護法では個人情報を生存している個人の情報（第 2 条）と定義しており、NDI で使用する死亡診断書に基づく情報は個人情報には該当しない。個人情報保護法の問題は解決できたとしても、対象者本人や遺族のプライバシー問題は残る。これについては、NDI を使用した研究の公表は統計的な集計結果など個人を特定できない形でしか認めず、違反した場合には当該研究者の罰則規定を定めておく必要がある。研究者が所属する施設内倫理審査委員会の承認を原則とし、違反した場合には施設単位で使用停止することも検討に値する。NDI に要する経費は NDI を使用する研究者から適切な使用料を徴収することで解決を図ることができる。

4. 吹田研究では心筋梗塞、脳卒中いずれにおいても、definite と判定されている者では、約 10% しか人口動態統計の死亡情報に含まれていないが、一方で probable と判定されている者では 80-90% が含まれていた。発症後まもなく死亡した者では definite と判定することは非常に難しく、必然的に人口動態統計などから情報を得ることになり、probable と判定されるものが多いと推察された。人口動態統計を活用しなかった場合、心筋梗塞では半数以上、脳卒中では 4 分の 1 以上の情報を失う可能性があり、人口動態統計の利用は必須である。住

民票請求や人口動態統計による網羅的な追跡は依然として重要である。

5. 信頼性の高いがんの罹患統計とがん患者の生存率を整備するために、地域がん登録において、人口動態死亡統計の利用は不可欠である。

改正統計法下で人口動態統計の利用に関する判断主体が総務省より厚生労働省に移ることにより、申請から承認を得るまでの期間には相当な短縮が見られて、人口動態調査死亡情報の利用が今後促進されるものと期待される。

地域がん登録と院内がん登録の両者において、信頼性の高い生存率を計測し、がん医療の均てん化に資することが重要な課題である。これは、現行の統計法と人口動態調査の下で、調査情報の二次利用という整理では実現困難である。がん患者の正確な生存率の整備するためには、人口動態調査令施行細則に「都道府県知事は、人口動態調査票、出生小票及び死亡小票を県の運営資料として利用することができる」とする条項を追加することが必要と考える。

デンマークの5つの主要なデータベースについて、その詳細を調査した。国の機関がこれら全てのデータベースを管理・監督し、研究者に短期間でデータを提供するシステムが確立されている。また、これにより多くの有用な研究成果を世界に向けて発信している。デンマークの人口が500万人程度と比較的少数であり、高福祉国家であるため、このようなデータベース化が可能となっていると考える。しかし、研究利用に関して個々にインフォームド・コンセントを行っておらず、国民のデータ利用に関する意識や理解については不明である。日本でも大規模データベースを構築することで、疾病・治療の監視のみならず医学研究や病院部門における医療の質の向上にも繋がると考えられる。そのためには、人口規模や、国民の意識を考慮に入れた政策が必要であると考えられる。

6. 文献データベースで検索された文献

死亡者データベースに関する文献の検索の結果、データベースを導入している諸国においては、データベースを活用した研究が実施され、その研究成果が文献となっており、その数は増加傾向にあった。保健・医療・福祉施策を策定する上で、重要な根拠となりうる文献が死亡者データベースを活用されていることが伺えた。

また、導入することで研究者がデータの信頼性を評価し、かつ、その信頼性のもとで研究を実施できる環境が生まれていることも容易に想像できる。

また、韓国における死亡者データベースは研究活用が可能となっていた。情報源が韓国語

(ハングル文字)での文書、また、電話による取材であったため、研究協力者の李氏の協力のもとに情報収集を行った。導入のメリットとして、①個人識別能力が高く、資料連携の効率が高い。②人口規模がかなり大きいと、公共資料を用いた連携研究を進める際、ごく稀な疾患に関しても研究がやりやすく、相対的に検出力が高くなるので、部分集団における分析、及び要因別分析が可能となる。③健康保険資料を用いた追跡研究が容易になり、また完全性が保障される、などが挙げられていた。一方、悉皆性に問題があるとの指摘の他、デメリットも何点か指摘されていた。ただ、韓国ではどのような経過で死亡者データベースが導入されたかについては、今回の調査では明らかにならなかった。今後さらに、情報収集を行う必要がある。

E. 結論

1. 研究データと死亡データのレコードリンケージにおいて、死亡情報のデータベース化とレコードリンケージの技術面にはそれほど大きな問題がないと考えられ、将来に向けて様々な面からの検討・議論の進展を望まれる。個人情報を除く死亡データの医学研究への利用には、匿名化よりもオーダーメイド集計などの検討を進めることが大切であろう。

2. 大規模コホート(JALS)における死因確定の業務を通じ、①人口動態統計申請から許可までの作業上の課題の把握(申請側)、②上記対象地域において実施された追跡開始時調査受診者で死亡が把握されたものについて人口動態調査の調査票情報と照合し原死因を確定する作業、③原死因を確定する際に、人口動態調査の調査票情報と住民票から得た情報の一致率と誤差要因の検討、④人口動態調査情報から得た原死因から、疫学的研究に必須の事項に絞った二次病名の分類基準の作成を行った。

3. 日本にNDIが存在すれば、さらに疫学研究が発展していくことをプリオン病サーベイランスのデータを利用して示した。

4. コホート研究である吹田研究のデータを用いて、人口動態統計の死亡情報が非常に重要であることを示した。国民のプライバシーは最大限保護されるべきであるが、長期的な国民の利益のためには、統計情報の学術研究への柔軟な利用について、今後、法整備も含め検討していく必要があるものと考えられる。

5. 地域がん登録における人口動態死亡統計の利用と申請書の記載方法について、改正統計法下での標準化を定めた。申請のための事務手続きの労力と期間の軽減のみならず、信頼性の高

いがん統計の整備と活用に大きく寄与した。今後、地域がん登録と院内がん登録の両者が連携しながら、信頼性の高い生存率を計測し、がん医療の均てん化に資するよう、引き続き、関係機関との調整を進めていく。

6. デンマーク政府機関による保健医療関連データベースのなかで使用頻度がもっとも多い5種類のデータベースについて、その詳細を調査した。デンマークでは、国の機関がデータベースを管理・監督し、研究者に短期間でデータを提供するシステムが確立されている。これにより多くの有用な研究成果が世界に向けて発信され、人々の健康と福祉に貢献している。

6. 韓国における死亡者データベースの検討から、導入による問題点の発生はなかった。メリットのみならず、デメリットの指摘もあった。今後さらに導入の経過等について情報収集する必要がある。

F. 健康危険情報

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

研究協力者

原田亜紀子 千葉県衛生研究所健康疫学研究室 研究員

渡邊 至 国立循環器病センター予防検診部 医師

小久保喜弘 国立循環器病センター予防検診部 室長

中谷直樹 日本学術振興会 海外特別研究員、デンマーク対がん協会 研究員

中谷久美 デンマーク対がん協会 研究員

中野匡子 福島県立医科大学医学部公衆衛生学講座 博士研究員

李 相侖 大分大学福祉科学研究センター 講師

顧問

児玉和紀 主席研究員（放射線影響研究所）・日本疫学会前理事長

II. 分担研究報告

厚生労働科学研究費補助金（統計情報総合研究事業）
分担研究報告書

死亡統計の研究利用における統計的課題
—レコードリンケージと匿名化の方法—

研究分担者 橋本修二 藤田保健衛生大学医学部衛生学講座教授

研究要旨 死亡統計の研究利用に関する主な統計的課題として、個人情報に基づく研究データと死亡データのレコードリンケージ、および、個人情報を除く死亡データの匿名化が挙げられた。研究データと死亡データのレコードリンケージにおいて、死亡情報のデータベース化とレコードリンケージの技術面にはそれほど大きな問題がないと考えられ、将来に向けて様々な面からの検討・議論の進展を望まれる。匿名化されたデータの利用拡大には多くの経験と実績の積み重ねが基礎であり、今後、医学研究での寄与が大いに期待される。個人情報を除く死亡データの医学研究への利用には、匿名化よりもオーダーメイド集計などの検討を進めることが大切であろう。

A. 研究目的

医学研究において、死亡統計の利用はきわめて重要である。本研究の目的は、死亡統計の研究利用における統計的課題を明らかにするとともに、その検討を行うことにある。昨年度は2年計画の初年度として、統計的課題を整理するとともに、死亡関連統計の現状把握を行った。統計的課題としては、個票形式の死亡データでは、主として個人情報を使用する研究データとのレコードリンケージ、および、個人情報を削除する匿名化が挙げられた。死亡関連統計としては、人口動態統計と住民基本台帳の現状を把握した。

本年度は2年計画の最終年度として、人口動態調査死亡票データに対して、研究データとのレコードリンケージの方法と匿名化の方法について検討した。

B. 研究方法

人口動態調査死亡票の項目について、人口動態調査オンライン報告システムの入力状況、および、人口動態調査の磁気媒体への転写状況を整理した。なお、磁気媒体へ転写された項目が人口動態統計に主に利用される。

個人識別情報の入力・転写状況ごとに、研究データとのレコードリンケージの方法を検討した。匿名化の方法、および、統計データへの適用状況を整理するとともに、人口動態統計の死亡データの匿名化の方法を議論した。

C. 研究結果

1. 人口動態調査死亡票の項目

表1と表2に、人口動態調査死亡票の項目について、人口動態調査オンライン報告シス

テムの入力状況、および、人口動態調査の磁気媒体への転写状況を示す。人口動態調査オンライン報告システムでは、死亡票の19項目が入力され、70変数に及んでいた。

個人識別情報として、氏名（文字；漢字）、性、生年月日、住所（文字；丁目や番地を含む）が含まれていた。人口動態調査の磁気媒体では、性、生年月日、住所（市町村符号）が含まれ、氏名や詳しい住所は含まれていない。

死亡に関する項目としては、死亡したとき、死亡したところ、および、死亡の原因であった。死亡の原因として、人口動態調査のオンライン報告システムでは原死因と直接死因などが含まれ、磁気媒体では原死因のみであった。

それ以外の項目としては、世帯の主な仕事、死亡したところの種別などであった。生後1歳未満の病死では、出生時体重、妊娠週数、母の年齢などであった。

2. レコードリンケージの方法

人口動態調査死亡票データと研究データの間で、個人識別情報を用いて、個人単位にレコードリンケージする方法を議論する。ここで、リンクとしては、確率的一致でなく、完全一致を想定する。研究データは氏名、住所（詳細）、性、生年月日を含むことを前提とする。また、個人識別情報の間違いは考慮しない。

表3に、研究データの個人識別情報の保有状況別に、人口動態調査死亡票データとのレコードリンケージの実施可能性を示す。研究データが死亡の事件簿番号、死亡の有無と死

亡年月日を含んでいれば、人口動態調査のオンライン報告システムの入力データ、人口動態調査の磁気媒体の転写データともに、レコードリンケージが実施可能である。死亡の事件簿番号がなくとも、死亡の有無と死亡年月日を含んでいれば同様である。

研究データが死亡の事件簿番号、死亡の有無と死亡年月日を含んでいなくとも、氏名と住所（詳細）によって、人口動態調査のオンライン報告システムの入力データとはレコードリンケージがほぼ実施可能（転居がない場合）である。人口動態調査の磁気媒体の転写データには氏名と住所（詳細）がないために、レコードリンケージが実施可能でない。

3. 匿名化の方法

表4に、匿名データを作成された4統計の概要を示す。4統計としては、全国消費実態調査、社会生活基本調査、就業構造基本調査、住宅・土地統計調査であった。調査の年次は最近のもの（平成17年以降）を含んでいなかった。調査の基本単位は社会生活基本調査と就業構造基本調査が個人で、全国消費実態調査と住宅・土地統計調査が世帯または住戸・世帯であった。いずれも標本調査であり、標本の抽出率（標本の概数／対応する母集団）は住宅・土地統計調査が6.5%、それ以外の3統計が0.1～0.8%であった。

表5に、匿名化の方法と統計データへの適用状況を示す。匿名化の方法として、レコードのリサンプリング（レコードの一部を削除）は4統計とも実施され、リサンプリング率が3統計で80%、住宅・土地統計調査（標本の抽出率が高い）で10%であった。識別情報の削除は調査区番号等の削除とレコードの無作為な並び替えが実施された。裾切りによるレコードの削除は世帯人員8人以上または三つ子以上のいる世帯が削除された。トップ・ボトムコーディングは年齢の上限値が85歳以上とし、全国消費実態調査と住宅・土地統計調査で住宅の規模の上限値・下限値が適用された。リコーディングは地域区分が3統計で三大都市圏・その他の地域の2区分とし、住宅・土地統計調査で47都道府県となり、年齢の区分が15歳以上を5歳階級となった。匿名化の方法として、誤差の導入、データスワッピング（レコード間の識別情報の一部を交換）は適用されなかった。

D. 考察

人口動態調査死亡票データに対して、研究

データとのレコードリンケージの方法を議論した。研究データとして、氏名、住所（詳細）、性と生年月日を含むことを前提としたが、死亡データとのリンケージするためには当然であろう。ここでは、リンクとして確率的一致でなく、完全一致を想定した。死亡データとのレコードリンケージを行う場合、医学研究では、通常、予後や要因のリスク評価を目的とするためである。

住民票コード（住民基本台帳で使用する11桁の数値）については検討しなかった。現在、このコードが研究に利用できず、また、死亡票に記載されていないためである。かりに利用できれば、生死の情報をはじめ、他の個人識別情報がなくともレコードリンケージが可能となるが、その利用に向けては様々な議論を要するものと思われる。

研究データが死亡の事件簿番号、死亡の有無と死亡年月日を含まなくとも、人口動態調査オンライン報告システムの入力データとはレコードリンケージがほぼ実施可能（転居がない場合）であった。これは、住民基本台帳や死亡小票などと照合しなくとも、レコードリンケージがほぼ実施可能であることを意味する。いわゆる、National Death Indexに近いシステムともいえよう。その実現に向けて、様々な検討や議論を要することはいうまでもないが、死亡情報のデータベース化とレコードリンケージの技術面には、現在、それほど大きな問題がないと考えられる。

匿名化の方法については、過去に様々な検討と議論が行われており、最近、日本でもガイドラインとしてまとめられている。現在、利用可能な匿名データをみると、いずれも、標本抽出を用いた統計調査である。匿名化の方法としては、データの利用率を保ちつつも、匿名性の確保をやや優先している印象を受ける。匿名性の担保の必要性はいうまでもないが、その上でもデータの利用率をさらに拡大する余地が小さくないと考える。匿名化されたデータの利用拡大に向けて、多くの経験と実績の積み重ねが基礎であり、今後、医学研究での寄与が大いに期待される。

死亡統計の医学研究への利用にあたっては、主として標本抽出でなく、全数データが前提となる。一方、匿名化の確保にあたっては、全数でなく、抽出が最も重要な要件の1つと位置づけられる。このような点を考慮すると、死亡情報の匿名データの作成については、かなり限定的なことと考える。むしろ、死亡データの二次利用には、オーダーメイド集計が

より適切と考えられ、その点は、今後の重要な検討課題であろう。

E. 結論

死亡統計の研究利用に関する主な統計的課題として、個人情報に基づく研究データと死亡データのレコードリンケージ、および、個人情報を除く死亡データの匿名化が挙げられた。研究データと死亡データのレコードリンケージにおいて、死亡情報のデータベース化とレコードリンケージの技術面にはそれほど大きな問題がないと考えられ、将来に向けて様々な面からの検討・議論の進展を望まれる。匿名化されたデータの利用拡大には多くの経験と実績の積み重ねが基礎であり、今後、医学研究での寄与が大いに期待される。個人情報を除く死亡データの医学研究への利用には、匿名化よりもオーダーメイド集計などの検討

を進めることが大切であろう。

F. 研究発表

1. 論文発表
なし。

2. 学会発表
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
なし。

2. 実用新案登録
なし。

3. その他
なし。

表1. 人口動態調査死亡票におけるオンライン報告システムの入力状況と磁気媒体への転写状況(1)

死亡票の 項目番号	番号別 通し番号	内容	入力 の形式#1	転写の 有無#2
○	1	市町村符号	記号	有
○	1	保健所符号	記号	有
○	1	事件簿番号	記号	有
○	1	市町村受付年月日	数値	
○	1	保健所受付年月日	数値	
1	1	氏名	文字	
2	1	男女別	選択	有
3	1	生年月日	数値	有
4	1	死亡したとき	数値	有
5	1	死亡したところ	選択	有
6		死亡した人の住所		
	1	届出地と同じ市区町村・届出地以外の市区町村	選択	有
	2	外国・不詳	選択	有
	3	都道府県	選択	
	4	市, 郡, 東京都の区	選択	
	5	町, 村, 指定都市の区	選択	
	6	指定都市の町, 字, 丁目, 番地, 番号	文字	
	7	アパート・マンション, 様方	文字	
	8	市町村符号	記号	有
	9	保健所符号	記号	有
7	1	死亡した人の国籍	選択	有
8,9		死亡した人の夫または妻		
	1	いる・いない(未婚・死別・離別)・不詳	選択	有
	2	満年齢	数値	有
10	1	死亡したときの世帯の主な仕事	選択	有
11	1	死亡したときの職業・産業	記号	有
12,13	1	死亡したところの種別	記号	有
	2	施設の名称	文字	
○	1	原死因符号	記号	有
○	1	外因の状況符号	記号	有
○	1	発生したところの符号	記号	有
○	1	傷害発生したところの符号	記号	有
○	1	母側符号	記号	有

#1: 人口動態調査オンライン報告システムの入力の形式

#2: 人口動態調査の磁気媒体の転写の有無

○: 死亡票の番号のない項目

表2. 人口動態調査死亡票におけるオンライン報告システムの入力状況と磁気媒体への転写状況(2)

死亡票の 項目番号	番号別 通し番号	内容	入力 の 形式#1	転写の 有無#2
14		死亡の原因		
	1	I (ア) 直接死因	文字	
	2	発病(発症)又は受傷から死亡までの期間	文字	
	3	I (イ) (ア) の原因	文字	
	4	発病(発症)又は受傷から死亡までの期間	文字	
	5	I (ウ) (イ) の原因	文字	
	6	発病(発症)又は受傷から死亡までの期間	文字	
	7	I (エ) (ウ) の原因	文字	
	8	発病(発症)又は受傷から死亡までの期間	文字	
	9	II. I 欄に影響を及ぼした傷病名等	文字	
	10	発病(発症)又は受傷から死亡までの期間	文字	
	11	手術 有無	選択	有
	12	手術 部位及び主要所見	文字	
	13	手術年月日	数値	
	14	解剖 有無	選択	有
	15	解剖 主要所見	文字	
15	1	死因の種類	選択	有
16		外因死の追加事項		
	1	傷害が発生したとき	数値	
	2	傷害が発生したところの種別	選択	
	3	傷害が発生したところの種別 その他	文字	
	4	傷害が発生したところ 都道府県	選択	
	5	傷害が発生したところ 市, 郡, 東京都の区	選択	
	6	傷害が発生したところ 町, 村, 指定都市の区	選択	
	7	手段及び状況	文字	
17		生後1年未満で病死した場合の追加事項		
	1	出生時体重	数値	有
	2	単胎・多胎の別	選択	有
	3	多胎の数	数値	有
	4	多胎の順位	数値	有
	5	妊娠週数	数値	有
	6	妊娠・分娩時における母体の病態又は異状 有無	選択	有
	7	妊娠・分娩時における母体の病態又は異状 内容	文字	
	8	母の生年月日	数値	有
	9	前回までの妊娠結果 出生児数	数値	有
	10	前回までの妊娠結果 妊娠満22週以降の死産児数	数値	有
18	1	その他特に付言すべきことがら 有無	選択	有
	2	その他特に付言すべきことがら	文字	
19	1	施設の所在地又は医師の住所及び氏名 住所	文字	
	2	施設の所在地又は医師の住所及び氏名 氏名	文字	
○	1	備考 有無	選択	有
○	1	備考	文字	

#1: 人口動態調査オンライン報告システムの入力の形式

#2: 人口動態調査の磁気媒体の転写の有無

○: 死亡票の番号のない項目

表3. 死亡データと研究データのレコードリンケージの実施可能性

	研究データの状況		
	①	②	③
研究データの個人識別情報の保有状況			
氏名	含む	含む	含む
性	含む	含む	含む
生年月日	含む	含む	含む
住所（詳細）	含む	含む	含む
死亡の有無・死亡年月日	含む	含む	含まない
死亡の事件簿番号	含む	含まない	含まない
研究データの人口動態調査死亡票とのレコードリンケージの実施可能性			
死亡票のオンライン報告の入力データ	可能	可能	ほぼ可能
死亡票の磁気媒体の転写データ	可能	可能	不可能