

201202005A

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)

厚生労働統計データの利用促進等に関する研究

平成 24 年度 研究報告書

主任研究者 祖父江友孝

平成 25(2013)年 4 月

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)

厚生労働統計データの利用促進等に関する研究

平成 24 年度 研究報告書

主任研究者 祖父江友孝

平成 25(2013)年 4 月

目 次

I. 総括研究報告

- 厚生労働統計データの利用促進等に関する研究……………1
祖父江友孝 大阪大学大学院医学系研究科環境医学教授

II. 分担研究報告

1. 公的統計利用（オーダーメイド集計、匿名データの提供）検討……………7
橋本修二 藤田保健衛生大学医学部衛生学講座教授

人口動態統計の二次利用のありかたに関する研究

2. 人口動態統計の二次利用のありかたに関する研究……………14
岡山 明 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教授
原田亜紀子 Japan Arteriosclerosis Longitudinal Study 事務局

3. がんおよびその他の疾患に関する厚生労働統計データの解析……………19
片野田耕太 独立行政法人国立がん研究センターがん対策情報センターがん情報統計部

4. 母子保健に関する厚生労働統計データの解析および外部データとの連結解析・26
山縣然太郎 山梨大学大学院医学工学総合研究部 社会医学講座 教授
鈴木孝太 山梨大学大学院医学工学総合研究部 社会医学講座 准教授

5. 糖尿病に関する厚生労働統計データの解析……………47
中村好一 自治医科大学教授
坪井聡 自治医科大学助教

III. 研究成果の刊行に関する一覧表

I. 総括研究報告

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））
総括研究報告書

厚生労働統計データの利用促進等に関する研究

研究代表者 祖父江友孝 大阪大学大学院医学系研究科環境医学教授

研究要旨

公的統計の二次的利用申請に基づくデータ解析によって、公的統計の行政活用および発展的な活用の方法を探索し、その運用上、技術上の問題点と解決策を提示することを目的とし、以下の3点について検討した。

【公的統計二次的利用方法の整理と課題検討】疫学研究における公的統計の利用目的・方法として、記述的疫学的と分析疫学的なものに整理された。公的統計の二次的利用の手続等には大きな課題がなかった。利用上の課題として、オーダーメイド集計では対象統計と集計方法が、匿名データの提供では対象統計と情報内容が挙げられた。

【公的統計と外部データの連結】JALS (Japan Arteriosclerosis Longitudinal Study) 統合研究において、人口動態統計を用いて死因同定を行ったところ、性、生年月日、死亡年月日、死亡時市町村コードすべてのキーを用いて一旦照合した上で不一致事例に限定して活用すれば、99.5%以上の一致率が確保できた。

【公的統計二次的利用のデータ解析例】(1) 医療施設調査（静態）、(2) 人口動態調査（出生票）と21世紀出生児縦断調査(3) 患者調査、医療施設調査および受療行動調査を用いて、それぞれ、がん、母子保健、糖尿病領域の課題を検討することにより、その運用上、技術上の問題点を検討した。

研究分担者氏名・所属機関名・職名

中村 好一 自治医科大学公衆衛生学教室 教授

橋本 修二 藤田保健衛生大学医学部障害衛生学講座 教授

岡山 明 結核予防会第一健康相談所 所長

山縣然太郎 山梨大学大学院社会医学講座 教授

大橋清雄 東京大学大学院医学系研究科公共健康医学
専攻生物統計学教

片野田耕太 国立がん研究センターがん統計研究部室長

A. 研究目的

わが国では、新統計法の下制度的な整備が行われたが、必ずしも公的統計のデータ活

用が進んでいるとは言えない。本研究では、二次的利用の申請に基づくデータ解析によって、公的統計の行政活用および発展的な活用の方法を探索し、その運用上、技術上の問題点と解決策を提示することを目的とした。

B. 研究方法

【公的統計二次的利用方法の整理と課題検討】公的統計の利用形式として、下記の4つを対象とした。①公表された集計データの利用、②統計法33条に基づく個票データ

の利用、③オーダーメイド集計の利用、④匿名データの提供の利用、を対象とした。公的統計の疫学研究への利用事例を基礎資料とした。利用事例として、過去(①と②)と最近(①～④)のものを参照した。

【公的統計と外部データの連結】

JALS (Japan Arteriosclerosis Longitudinal Study) 統合研究は全国各地で統一の調査票に基づき調査調査統合して大規模コホートを構成した研究である。本研究ではその中で死亡をエンドポイントとして設定している 21 コホートの内からすでに死因を把握している 1 コホートを除いた対象者計 86870 名から追跡中に観察された死亡例 5114 名について人口動態統計死亡票の磁気情報と照合し商号結果の一致度がキー変数の数や部分使用によってどのように変化するかを検討した。

H24 年度の申請に基づき、人口動態統計を用いて JALS 事務局が行った死因同定の手順を整理し、JALS 事務局での初回の同定作業に要する作業を図式化した。一方米国の NDI 制度を参考に我が国で行う場合の同定の手順書を作成する。死因を同定する対象者数は全部で 5114 名であった。対象とする地域は 144 地区であり、これらの地区で観察期間中に記録された人口動態死亡票の件数は 761294 件であった。データの結合にはコホートデータと人口動態統計死亡票磁気情報から抽出した性別、生年月日、死亡年月日、死亡時市町村コードとした。これらを用いてすべて照合した場合、一部を用いた場合の一致率を検討した。

【公的統計二次的利用のデータ解析例】

(1) 医療施設調査(静態)

医療施設調査(静態)は、保健統計調査の基幹統計および一般統計として、3 年毎に

実施されている。本研究では、疾病対策の評価への利用可能性を検討することを目的として、平成 20 年医療施設調査(静態)を用いて、がんに関わる医療施設に関する解析を行った。対象 8,814 施設のうち、一般病床のある 6,133 施設を解析対象とした。

(2) 人口動態調査(出生票)と 21 世紀出生児縦断調査

人口動態調査データと 21 世紀出生児縦断調査データを連結し、集計・解析を行う。人口動態調査-第 1 回縦断調査-第 2 回...-第 10 回縦断調査という形で連結した、1 つのデータセットを作成することにより、都道府県別の集計結果表および児の発育に関する図を作成する。

上記データセットを作成するときには、データの存在しない調査回については欠損値として扱い、データ消去は行わない。

(3) 患者調査、医療施設調査および受療行動調査

平成 20 年の患者調査(患者の特性(性、年齢、受療の状況、診療費等支払方法など))、医療施設調査(医療施設の特性(健診・保健指導の状況、禁煙外来の状況など))、受療行動調査(患者の行動レベルの状況(待ち時間、診察時間、満足度など))の調査項目を網羅的に吟味し、糖尿病対策に有用と思われる評価指標の候補を選定した。

(倫理面への配慮)

オーダーメイド集計、匿名データの提供の利用については、連結不可能匿名化された既存の統計資料のみを用いるため、個人情報保護に係る問題は生じない。人口動態統計との照合については、適切な個人情報安全管理のもとに、倫理審査委員会の承諾を得て行った。患者調査・受療行動調査・医療施設静態調査の照合データについ

ては、匿名化データを用いて解析を行った。

C. 研究結果

【公的統計二次的利用方法の整理と課題検討】

(1) 疫学研究での公的統計の利用目的・方法の整理

疫学研究における公的統計の利用目的は、記述疫学的なものや分析疫学的なものがある。記述疫学的な利用目的としては、疾病発生状況の記述等が挙げられる。その利用方法としては、疾病の有病率や死亡率についての対象者の属性別、年次推移や地域分布の観察等である。利用形式としては、①公表された集計データの利用と③オーダーメイド集計の利用が中心的である。分析疫学的な利用目的としては、疾病発生と要因曝露の関連性評価等が挙げられる。その利用方法としては、疾病発生と要因曝露についての集団単位の相関分析、統計の詳しい解析等があり、また、異なる統計間や研究と統計間のリンクを基礎とすることがある。利用形式としては、②統計法33条に基づく個票データの利用と④匿名データの提供の利用が中心的である。

(2) 疫学研究での公的統計の利用上の課題の検討

統計の二次的利用の手続等には大きな課題がない。たとえば、申請の手引書、相談窓口、申請書類、申請から受取までの期間、提供データ形式、費用等である。

③オーダーメイド集計の利用には、いくつかの課題がある（利用事例：平成20年患者調査、2012年10月）。対象統計としては、受療行動調査の追加等である。集計方法としては、人口動態統計の死因分類の細分化等である。④匿名データの提供の利用には、

いくつかの課題がある（利用事例：平成16年国民生活基礎調査、2011年12月）。対象統計としては、国民健康・栄養調査や新しい年次分の追加等である。情報内容としては、匿名化に伴う情報減少の緩和等である。

【公的統計と外部データの連結】

JALS事務局で行った死因の同定手順によると、各コホートからの提出データがすべて正しいとした場合の単純な照合でも97.5%が照合可能であった。更にコホートデータの生年月日と年齢の再照合を行ったところ、最も単純な照合方法で一致したのは99.3%であった。死亡年月日を死亡年月まで限定して照合すると99.5%が照合可能であった。生年月日を性年月までを用いて照合した場合も99.4%が照合可能であった。

結合キーを限定した場合の重複割合は実際の一致事例に対する不一致者の割合はそれぞれ0.26%、0.16%と限定的であった。更に照合キーを限定した際の一致状況の検討も行ったところ、性別を除外して結合すると生年月、死亡年月を用いた場合と同様の効果があることが推測された。以上からすべてのキーを用いて一旦照合した上で不一致事例に限定して活用すれば、99.5%以上の一致率が確保できることが明らかになった。以上からキー項目の扱いに熟知すれば、比較的短時間に照合処理が可能であることが明らかになった。

【公的統計二次的利用のデータ解析例】

(1) 医療施設調査（静態）

がん診療連携拠点病院とそれ以外の病院とで、受動喫煙防止、ニコチン依存症管理料算定、遠隔医療、高精度放射線治療、緩和ケアなどの状況を比較した結果、がん診療連携拠点病院ではそれ以外の病院より、こ

これらの医療体制が整備されている傾向があった。また、がん診療連携拠点病院の中でも、都道府県がん診療連携拠点病院においてより医療体制が整備されている傾向があった。

(2) 人口動態調査(出生票)と21世紀出生児縦断調査

昨年12月20日より、厚労省の担当者と、メールで連絡を取りあい、申請を進めている(2月6日現在)。今年度中にデータを入手し、上記の解析を終える予定である。

(3) 患者調査、医療施設調査および受療行動調査

選定された指標候補の実用性を検討するために、統計法(平成19年法律第53号)第33条の規定に基づき、患者調査、医療施設調査、受療行動調査の二次的利用の申請を行った(平成25年1月7日)。データが入手でき次第、速やかに二次的利用の申請に基づく実際のデータ解析に着手する

(平成25年1月22日現在)。なお、申請によって得られたデータは厳重な注意のもとで扱う。

D. 考察

【公的統計二次的利用方法の整理と課題検討】

「公的統計の疫学研究への二次的利用に関する調査」が、日本疫学会統計利用促進委員会と、平成23年度厚生労働科学研究費補助金による「厚生労働統計データの疫学研究への二次的利用と他のヘルスケアデータとの連携について」研究班により実施された。統計の疫学研究への二次的利用の現状と課題を明らかにすることが目的である。日本疫学会会員766人(電子メールアドレス公開者)を対象とし、2011年9～10月にインターネットを介して547人

(71.6%)から回答を得た。統計法33条に基づく個票データについて、利用者の割合は2007年の日本疫学会会員調査のそれと変わらなかったが、申請から許可までの期間は大幅に短縮していた。オーダーメイド集計と匿名データの提供について、利用希望者の割合は大きく、また、国民健康・栄養調査と国民生活基礎調査などの利用希望が多かった。

この調査結果から、疫学研究における公的統計の二次的利用の大きな進展が期待される。利用希望者の多くにおいて、利用目的・方法が具体化され、実際の利用にむすびつけられることが望まれる。その際に、本研究による利用目的・方法の整理の結果または利用事例が参考になるものと思われる。一方、公的統計の二次的利用の運用側においては、利用希望の統計が考慮されることが望まれる。また、本研究による利用上の課題が参考になるものと思われる。

【公的統計と外部データの連結】

本研究で用いた人口動態死亡票データセットは当該地域・当該年度のすべてのデータとなる。真に必要とするデータは提供されたデータの0.67%にとどまった。現在の統計情報の二次利用の制度ではこれら大部分の研究とは直接関連のない個人の死因情報が研究者に送付されることを防ぐことはできない。

研究者が研究と直接関連のない情報を取り除き対象者の死因同定を行う必要が生じている。NDIを用いた場合の手順では研究対象以外の死因情報を研究者が扱うことはあり得ない。NDIが可能となればこうした個人識別情報を含む研究と関連しない多数の情報は提供されず漏洩の危険性がきわめて低くなるメリットがある。また、

NDS 事務局には照合のノウハウが蓄積し、照合に要する時間もきわめて短時間になることが推測される。

E. 結論

疫学研究における公的統計の利用目的・方法として、記述的疫学的と分析疫学的なものに整理された。公的統計の二次的利用の手続等には大きな課題がなかった。利用上の課題として、オーダーメイド集計では対象統計と集計方法が、匿名データの提供では対象統計と情報内容が挙げられた。

NDI 制度は研究者、行政共にメリットの大きい制度であり疫学研究推進のためには是非実現すべき制度と考えられた。

今後、実際のデータ解析により選定した指標の妥当性を検討し、がん対策、母子保健対策、糖尿病対策に関する公的統計の活用方法や、集計に伴う技術上の課題等を検討する。

F. 研究発表

1. 論文発表

- 1) Sobue T. Scientific approach to radiation-induced cancer risk. Fukushima J Med Sci. 2011;57(2):90-2.
- 2) Higashi T, Nakamura F, Saruki N, Takegami M, Hosokawa T, Fukuhara S, Nakayama T, Sobue T. Evaluation of newspaper articles for coverage of public reporting data: a case study of unadjusted cancer survival data. Jpn J Clin Oncol. 2013;43(1):95-100.
- 3) Chihara D, Ito H, Matsuda T, Katanoda K, Shibata A, Saika K, Sobue T, Matsuo K. Decreasing trend in mortality of chronic myelogenous leukemia patients after introduction of imatinib in Japan and the US. Oncologist. 2012;17(12):1547-50.
- 4) McCarthy BJ, Shibui S, Kayama T, Miyaoka E, Narita Y, Murakami M, Matsuda A, Matsuda T, Sobue T, Palis BE, Dolecek TA, Kruchko C, Engelhard HH, Villano JL. Primary CNS germ cell tumors in Japan and the United States: an analysis of 4 tumor registries. Neuro Oncol. 2012;14(9):1194-200.
- 5) Chihara D, Ito H, Katanoda K, Shibata A, Matsuda T, Tajima K, Sobue T, Matsuo K. Increase in incidence of adult T-cell leukemia/lymphoma in non-endemic areas of Japan and the United States. Cancer Sci. 2012;103(10):1857-60.
- 6) Ikeda N, Inoue M, Iso H, Ikeda S, Satoh T, Noda M, Mizoue T, Imano H, Saito E, Katanoda K, Sobue T, Tsugane S, Naghavi M, Ezzati M, Shibuya K. Adult mortality attributable to preventable risk factors for non-communicable diseases and injuries in Japan: a comparative risk assessment. PLoS Med. 2012;9(1):e1001160.
- 7) Matsuda T, Marugame T, Kamo K, Katanoda K, Ajiki W, Sobue T; Japan Cancer Surveillance

Research Group. Cancer incidence and incidence rates in Japan in 2006: based on data from 15 population-based cancer registries in the monitoring of cancer incidence in Japan (MCIJ) project. Jpn J Clin Oncol. 2012;42(2):139-47.

- 8) Katanoda K, Ajiki W, Matsuda T, Nishino Y, Shibata A, Fujita M, Tsukuma H, Ioka A, Soda M, Sobue T. Trend analysis of cancer incidence in Japan using data from selected population-based cancer registries. Cancer Sci. 2012;103(2):360-8.
- 9) 橋本修二, 川戸美由紀, 山田宏哉, 谷脇弘茂, 栗田秀樹. 国民生活基礎調査の匿名データによる健康状態と喫煙の解析. 厚生指標, 2012;59(13):27-31.
- 10) 大橋靖雄, 島本和明, 佐藤眞一, 他 日本動脈硬化縦断研究(JALS)グループ: 肥満を含む循環器リスクファクターの重積と脳卒中発症リスクの検討. 日本公衆衛生雑誌 2011 ; 58(12) : 1007-1015.
- 11) 大橋靖雄, 田栗正隆: メタアナリシスの統計学. 動脈硬化予防 2012;11(2):53 - 60
- 12) 大橋靖雄, 原田亜紀子, 岡山 明, 他 日本動脈硬化縦断研究(JALS)グループ: 人口動態調査の調査票情報を用いた大規模コホート研究における死因照合作業の問題点の検討. 厚生指標 2012;59(11):1 - 8

2. 学会発表

- 1) 橋本修二. 公的統計の二次的利用および疫学会アンケート結果について. 日本疫学会学術総会. 2013年1月. 大阪.
- 2) 片野田耕太. 公的統計の二次的利用について—申請する立場から—. 日本疫学会学術総会. 2013年1月. 大阪.
- 3) 山縣然太郎. 個人情報保護・医療分野個別法について. 日本疫学会学術総会. 2013年1月. 大阪.
- 4)

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

II. 分担研究報告

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））
分担研究報告書

公的統計利用（オーダーメイド集計、匿名データの提供）検討

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

疫学研究における公的統計の利用拡大に向けて、利用目的・方法の整理と利用上の課題の検討を行うことを目的とした。利用形式としては、公表された集計データ、統計法 33 条に基づく個票データ、オーダーメイド集計、匿名データの提供とした。利用目的・方法は記述疫学的と分析疫学的なものに整理し、利用方法ごとに利用事例を提示した。利用の手続などには大きな課題がなかった。利用上の課題として、オーダーメイド集計では対象統計と集計方法が、匿名データの提供では対象統計と情報内容が挙げられた。

A. 研究目的

公的統計として、多種多様な情報が収集され、人口動態統計、患者調査、国民生活基礎調査、国民健康・栄養調査、医療施設調査、受療行動調査などの多くの健康に関わる統計（保健統計）が作成されている。

統計法が改正され、2009 年度から施行されている。それによって、公的統計は「行政のため」から「社会の情報基盤」の統計へと転換され、二次的利用の促進が規定された。利用形式として、公表された集計データ、統計法 33 条に基づく個票データ（従前の目的外使用）とともに、新たにオーダーメイド集計と匿名データの提供が開始されている。

従来から、疫学研究では保健統計が比較的良好に利用されてきた。その利用拡大は保健統計の二次的利用の促進に大いに寄与するとともに、疫学と医学の発展および国民の健康増進につながるこ

と期待される。

本研究では、疫学研究における公的統計の利用拡大に向けて、利用目的・方法の整理と利用上の課題の検討を行うことを目的とした。

B. 研究方法

基礎資料として、公的統計の疫学研究への利用事例を取り上げた。公的統計の利用形式として、次の 4 つを対象とした。①公表された集計データの利用：報告書冊子またはホームページ（e-Stat）からファイルをダウンロード。②統計法 33 条に基づく個票データの利用：従前の目的外使用。③オーダーメイド集計の利用：研究者が集計方法を指定し、国の機関などが集計し、その集計結果表を受け取る。④匿名データの提供の利用：匿名化された個票情報を国の機関などから提供され集計する。

(倫理面への配慮)

本研究では、個人情報扱わないため、個人情報保護に係る問題は生じない。

C. 研究結果

(1) 疫学研究での公的統計の利用目的・方法の整理

表 1 に、疫学研究での公的統計の利用目的・方法を示す。疫学研究の公的統計の利用目的としては、記述疫学的なものと同分析疫学的なものに整理した。

記述疫学的な利用目的としては、疾病発生状況の記述などが挙げられた。その利用方法としては、疾病の有病率や死亡率についての対象者の属性別、年次推移や地域分布の観察などである。利用形式として、①公表された集計データの利用と③オーダーメイド集計の利用が中心である。

分析疫学的な利用目的としては、疾病発生と要因曝露の関連性評価などが挙げられた。その利用方法としては、疾病発生と要因曝露についての(a)集団単位の相関分析、(b)統計の詳しい解析などがあり、また、(c)異なる統計間のリンクや(d)研究と統計間のリンクを基礎とすることがある。利用形式として、②統計法 33 条に基づく個票データの利用と④匿名データの提供の利用が中心的である。

(2) 疫学研究での公的統計の利用事例
分析疫学的な利用目的において、前述の(a)～(d)の利用方法を例示するために、それぞれの利用事例を示す。

図 1 は利用方法が(a)集団単位の相関分析の事例である¹⁾。検討目的としては、平均自立期間と介護予防事業費の関連

性評価である。基礎資料として、介護給付費実態調査と介護保険事業状況報告などを用いている。方法として、都道府県を単位として平均自立期間と介護予防事業費の相関分析である。

図 2 は利用方法が(b)統計の詳しい解析の事例である²⁾。検討目的としては、日常生活動作(起床、食事、入浴など)の制限と疾病(脳血管疾患、認知症など)の有病の関連性評価である。基礎資料として、国民生活基礎調査の個票データ(2007 年)を用いている。方法として、日常生活動作と疾病有病の性・年齢別のクロス集計である。

図 3 は利用方法が(c)異なる統計間のリンクの事例である³⁾。検討目的としては、高齢者の介護の要否と栄養摂取状況の関連性評価である。基礎資料として、国民生活基礎調査(介護)と国民健康・栄養調査(栄養)を用いている。方法として、世帯・性・生年月日により、個人単位の統計間をリンクしている。対象者は 14,240 人で、リンク率は 93%である。

図 4 は利用方法が(d)研究と統計間のリンクの事例である⁴⁾。検討目的としては、高齢者の要介護の発生と体格の関連性評価である。基礎資料として、大崎コホート研究(高齢者 1.6 万人を 2007～2010 年に追跡)を用いている。方法として、各人ごとに、研究情報と人口動態統計、介護保険の情報をリンクしている。

(3) 疫学研究での公的統計の利用上の課題の検討

表 2 に、疫学研究での公的統計の利用上の課題を示す。統計の二次的利用の手続などには大きな課題がない。たとえば、

申請の手引書、相談窓口、申請書類、申請から受取までの期間、提供データ形式、費用などである。

③オーダーメード集計の利用には、いくつかの課題がある（利用事例：平成20年患者調査、2012年10月）。対象統計としては、受療行動調査の追加などである。集計方法としては、人口動態統計の死因分類の細分化などである。

④匿名データの提供の利用には、いくつかの課題がある（利用事例：平成16年国民生活基礎調査、2011年12月）⁵⁾。対象統計としては、国民健康・栄養調査や新しい年次分の追加などである。情報内容としては、匿名化に伴う情報減少の緩和などである。

D. 考察

平成23年度厚生労働科学研究費補助金による「厚生労働統計データの疫学研究への二次的利用と他のヘルスケアデータとの連携について」研究班によって、「公的統計の疫学研究への二次的利用に関する調査」が日本疫学会の統計利用促進委員会と共同して実施された⁶⁾。

この調査では、統計の疫学研究への二次的利用の現状と課題の一端が明らかにされた。日本疫学会会員766人（電子メールアドレス公開者）を対象とし、2011年9～10月にインターネットを介して547人（71.6%）から回答を得た。統計法33条に基づく個票データについて、利用者の割合は2007年の日本疫学会会員調査のそれと変わらなかったが、申請から許可までの期間は大幅に短縮していた。オーダーメード集計と匿名データの提供について、利用希望者の割合

は大きく、また、国民健康・栄養調査と国民生活基礎調査などの利用希望が多かった。

この調査結果からみると、疫学研究での公的統計の二次的利用には大きな進展の可能性がある。利用希望者の多くがその利用目的・方法を具体化すれば、実際の利用にむすびつけられるであろう。その際に、本研究による利用目的・方法の整理の結果または提示した利用事例が参考になるものと思われる。とくに、利用目的としては、疾病発生の記述から要因の関連性評価への拡大を期待したい。

公的統計の利用拡大に向けて、利用側とともに、運用側による課題の対応が重要と考えられる。これには、本研究で提示した利用上の課題が参考になるものと思われる。とくに、利用側の希望を十分に考慮して、利用対象統計が拡充されることが望まれる。

将来的な課題としては、疫学研究に向けて、統計の二次的利用の新たな展開を目指すことが重要であろう。その方向としては、たとえば、異なる統計間および研究と統計間の個人単位のリンクの促進などが挙げられる。

E. 結論

疫学研究での利用目的・方法は記述疫学的と分析疫学的なものに整理し、利用方法ごとに利用事例を提示した。利用の手続などには大きな課題がなかった。利用上の課題として、オーダーメード集計では対象統計と集計方法が、匿名データの提供では対象統計と情報内容が挙げられた。

引用文献

- 1) 尾島俊之, 野田龍也. 健康寿命の関連要因のマクロ面からの評価. 平成 22 年度厚生労働科学研究費(循環器疾患等生活習慣病対策総合研究事業)による「健康寿命の年次推移、地域分布と関連要因の評価に関する研究(研究代表者: 橋本修二)」平成 22 年度総括・分担研究報告書, 2011;16-39.
- 2) Hashimoto S, Kawado M, Yamada H, Seko R, Murakami Y, Hayashi M, Kato M, Noda T, Ojima T, Nagai M, Tsuji I. Gains in disability-free life expectancy from elimination of diseases and injuries in Japan. *J Epidemiol*, 2012;22:199-204.
- 3) 川戸美由紀, 橋本修二, 松村康弘, 小栗重統, 岡山 明, 中村好一, 柳川洋. 要介護高齢者と介護者の栄養摂取状況—国民生活基礎調査と国民栄養調査に基づく検討—. *日本公衆衛生雑誌*, 2002;49:922-928.
- 4) 辻 一郎, 永井雅人, 川戸美由紀. 健康寿命の関連要因のミクロ面からの評価. 平成 22 年度厚生労働科学研究費(循環器疾患等生活習慣病対策総合研究事業)による「健康寿命の年次推移、地域分布と関連要因の評価に関する研究(研究代表者: 橋本修二)」平成 22 年度総括・分担研究報告書, 2011;10-15.
- 5) 橋本修二, 川戸美由紀, 山田宏哉, 谷脇弘茂, 栗田秀樹. 国民生活基礎調査の匿名データによる健康状態と喫煙の解析. *厚生*の指標, 2012;59(13):27-31.

- 6) 祖父江友孝, 中村好一, 橋本修二, 岡山明, 辻一郎, 山縣然太郎, 安村誠司, 味木和喜子, 郡山千早. 公的統計の疫学研究への二次的利用に関する調査. 日本疫学会ホームページ (<http://jeaweb.jp/news/pdf/20120202houkokusho.pdf>) 2013.2.20.

F. 研究発表

1. 論文発表
なし
2. 学会発表
1) 橋本修二. 公的統計の二次的利用および疫学会アンケート結果について. *日本疫学会*, 2013;23 (Supple 1) :48.

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. 疫学研究での公的統計の利用目的・方法

<p>・利用目的：記述疫学的なもの（疾病発生状況の記述等）。</p> <p>利用方法：疾病の有病率や死亡率について、</p> <ul style="list-style-type: none">(a) 対象者の属性別の観察(b) 年次推移の観察(c) 地域分布の観察、等。 <p>利用形式：①公表された集計データの利用、</p> <ul style="list-style-type: none">③オーダーメイド集計の利用が中心的。
<p>・利用目的：分析疫学的なもの（疾病発生と要因曝露の関連性評価等）。</p> <p>利用方法：疾病発生と要因曝露について、</p> <ul style="list-style-type: none">(a) 集団単位の相関分析(b) 統計の詳しい解析(c) 異なる統計間のリンクに基づくもの(d) 研究と統計間のリンクに基づくもの、等。 <p>利用形式：②統計法 33 条に基づく個票データの利用、</p> <ul style="list-style-type: none">④匿名データの提供の利用が中心的。

表 2. 疫学研究での公的統計の利用上の課題

<p>・統計の二次的利用の手続等には大きな課題がない。</p> <p>たとえば、申請の手引書、相談窓口、申請書類、 申請から受取までの期間、提供データ形式、費用等。</p>
<p>・③オーダーメイド集計の利用には、いくつかの課題がある。</p> <p>(利用事例：平成 20 年患者調査、2012 年 10 月)</p> <p>対象統計：受療行動調査の追加等</p> <p>集計方法：人口動態統計の死因分類の細分化等</p>
<p>・④匿名データの提供の利用には、いくつかの課題がある。</p> <p>(利用事例：平成 16 年国民生活基礎調査、2011 年 12 月)</p> <p>対象統計：国民健康・栄養調査や新しい年次分の追加等</p> <p>情報内容：匿名化に伴う情報減少の緩和等</p>

図1. 利用事例：(a) 集団単位の相関分析

平均自立期間（年）（女性、65歳、2009年）

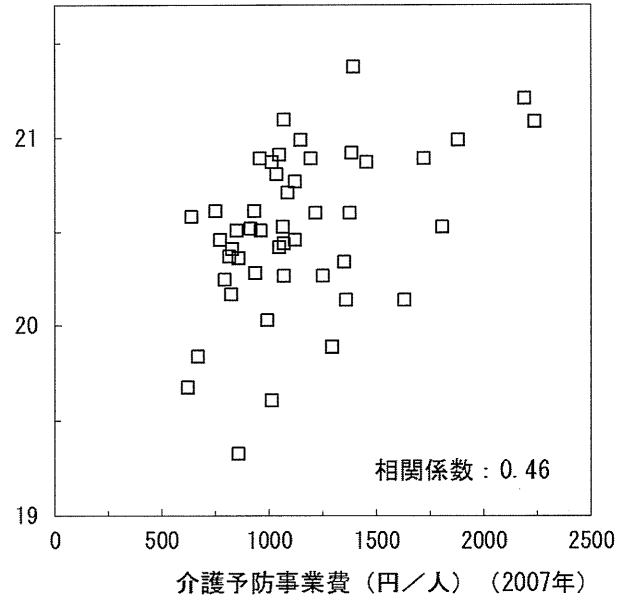


図2. 利用事例：(b) 統計の詳しい解析

日常生活動作の制限あり (%) 男性、2007年

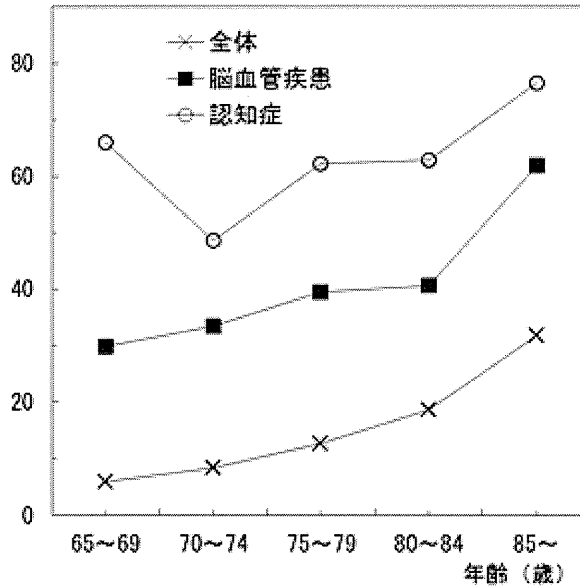


図3. 利用事例：(c)異なる統計間のリンク

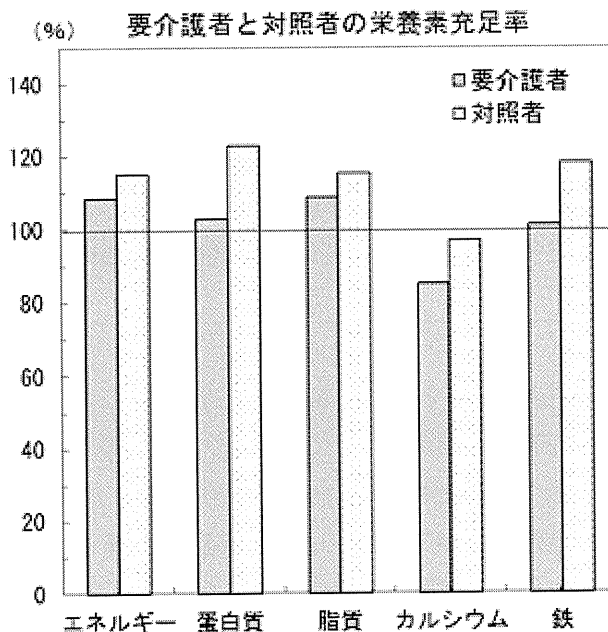
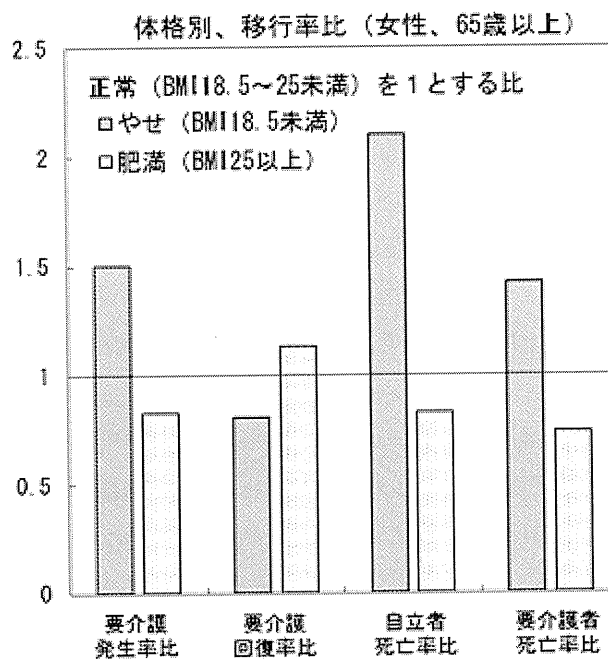


図4. 利用事例：(d)研究と統計間のリンク



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））
分担研究報告書

人口動態統計の二次利用のありかたに関する研究

研究分担者 岡山 明 結核予防会第一健康相談所 所長
研究分担者 大橋 靖雄 東京大学大学院医学系研究科教授 健康科学・看護学専攻
研究協力者 原田亜紀子 Japan Arteriosclerosis Longitudinal Study 事務局

研究要旨

厚生労働統計の二次利用を促進するための要因を明らかにする目的で、実際の疫学研究データを用いて、死因同定の手順に沿って作業を行い、現在我が国で実施されている手順と、米国をはじめとして各国で用いられている、ナショナルデスインデックス (NDI) を用いたと仮説をおいた場合の手順を比較した。現在の手順では研究者が本来必要のないデータを含め入手した上で、必要なデータ以外を消去することになっている。NDI の手順に沿って行えば、研究者が不必要な情報を入手する必要はなく、情報漏洩の可能性が低くなる。照合手順も単純になり、統計情報のさらなる活用が可能となって今後の疫学研究を発展させるものと考えられた。

A. 研究目的

統計法の改正により人口動態統計の二次利用が促進され、疫学研究での人口動態統計死亡票をもちいて死因を同定し、種々の生活習慣病率の解析に用いることが疫学研究の手法として広く普及し始めている。一方死因の同定には、種々のノウハウが必要とされ、研究者が申請により人口動態統計情報を入手しただけでは、十分な活用が困難な場合があることを昨年度の報告で明らかにした。

人口動態統計二次利用を促進するための手法について、米国等で実用化されている NATIONAL DEATH INDEX (NDI) を用いた場合と、現行の制度に基づく死因の同定作業を比較しその優劣を比較して記述することで我が国における NDI 推進に向

けた資料とする。

B. 研究方法

JALS 統合研究は全国各地で統一の調査票に基づき調査を統合して大規模コホートを構成した研究である。本研究ではその中で死亡をエンドポイントとして設定している 21 コホートの内からすでに死因を把握している 1 コホートを除いたコホートの追跡対象者計 86870 名の中から観察された死亡例 5114 名について人口動態統計死亡票の磁気情報と照合し、照合結果の一致割合がキー変数の数や部分使用によってどのように変化するかを検討した。H24 年度の申請に基づき、人口動態統計を用いて JALS 事務局が行った死因同定の手順を整理し、JALS 事務局での初回の同定作業に要する

作業を図式化した。一方米国の NDI 制度を参考に我が国で行う場合の同定の手順書を作成した。

対象者数は全部で 5114 名であり、照合対象とする地域は 144 地区であり、これらの地区で観察期間中に記録された人口動態死亡票の件数は 761294 件であった。データの結合にはコホートデータと人口動態統計死亡票磁気情報から抽出した性別、生年月日、死亡年月日、死亡時市町村コードとした。これらを用いてすべて照合した場合、一部を用いた場合の一致率を検討した。照合のフローを図にまとめた(図 1)。

一方 NDI 制度を用いた場合は研究者は死因同定に必要な情報を NDI 事務局に提出し、照合作業を依頼する。照合できなかったデータセットについて、再度照合キーの確認を行った上で再提出し、照合可能であったデータを受領する仕組みとした(図 1)。

C. 結果と考察

JALS 事務局で行った死因の同定手順によると、各コホートからの提出データがすべて正しいとした場合の最も単純な照合方法でも 97.5%が照合可能であった。更にコホートデータが誤っている可能性を考慮して各コホートに問い合わせを行い、生年月日と年齢の再照合を行ったところ、同様の照合方法で一致率 99.3%に向上した。

照合キーの一部情報を限定し、死亡年月日を死亡年月までに限定して照合すると 99.5%が照合可能であった。生年月日を性年月までを用いて照合した場合も 99.4%が照合可能であった。実際の一致事例に対して誤って照合されたものの割合はそれぞれ

0.26%、0.16%と限定的であった。以上からすべての照合キーデータを用いても一致しなかったものについて、照合キーを限定することで、照合率が更に高められると考えられた。

次に照合キーの数を限定した際の一致状況の検討も行ったところ、性別を除外して結合すると生年月、死亡年月を用いた場合と同様の効果があることが推測された。以上からすべてのキーを用いて一旦照合した上で不一致事例に限定して活用すれば、

99.5%以上の一致率が確保できることが明らかになった。以上からキー項目の扱いに熟知すれば、比較的短時間に照合処理が可能であることが明らかになった。

一方本研究で利用申請によって提供された人口動態死亡票データセットはコホート研究を実施した地域の当該年度のすべてのデータであった。コホート研究で真に必要とするデータは提供されたデータの 0.67%にとどまった。現在の二次利用の制度ではこれらの大部分の研究とは直接関連のない個人の死因情報が研究者に送付されることを防ぐことはできない。研究と直接関連のない情報を取り除き、対象者の死因同定を行う業務が研究者に課されている。研究者にとって疫学研究で死因の同定を行う作業は重要ではあるが、頻度は多くなく、せっかく身につけたノウハウを活用する場面も少ない。

NDI 制度の場合には、研究者が同定に必要な情報を照合を行う機関に提出し、照合結果を入手することになるので、研究対象以外の死因情報を研究者が扱うことはあり得ない。NDI 制度が可能となればこうした個人識別情報を含みしかも研究と関連しない