
一般口演 | データ二次利用／国際状況

一般口演21

データ二次利用／国際状況

2019年11月24日(日) 10:40 ～ 12:10 F会場 (国際会議場 3階中会議室302)

[4-F-2-02] 医療情報の二次利用の制度に関する現状—デンマークとオランダを例に—

○伊藤 伸介¹、木村 映善²（1. 中央大学, 2. 国立保健医療科学院）

キーワード：Register-Based Statistical System, Microdata, Administrative Data

デンマークやオランダといった国々においては、個人・家族に関する人口・社会的な行政記録情報や医療・健康に関する情報等、行政上の登録簿(レジスター)を用いた、レジスター(登録簿)ベースの統計作成システムが展開されている。具体的には、レジスターに収集・保管されている行政記録データ(以下「レジスターデータ」)が公的統計の作成のために自動的に統計作成部局に移送され、保管されるシステムが確立されている。デンマークを例に挙げると、保健省(ministry of health)のデンマーク健康データ局(Danish Health Data Authority)で保管されている、患者の病名や病歴等の様々なレジスターデータが、デンマーク統計局にオンラインで送られる。これらのレジスター情報は、人口情報(居住者情報)のレジスター、住所・建物情報のレジスター、および企業情報のレジスターといった各種のレジスターと、個人識別番号を介して連結される構造となっている。デンマークやオランダでは、個票データの保護に関する法制度、さらにはデータを提供する政府機関と認証された機関との間の信頼関係をもとに、レジスターデータの二次利用が展開されてきた。利用者は、個票データの利用申請を行うことによって、医療・健康を含む様々な分野を対象に、IDを用いてリンケージされた個票データにアクセスできる。個票データの利用にあたっては、直接的な識別子の削除と個人識別番号の仮名化を適用した上で、セキュアなリモートアクセスの環境の下で、可能になっている。また、分析結果の持ち出しにおいては、分析結果から個人情報が見えなくなるように、分析結果のチェックが行われている。本報告では、主として公的統計作成の観点から見たデンマークやオランダにおける医療情報の収集・保管・管理の現状、および医療情報の二次利用制度の特徴について述べることにしたい。

医療情報の二次利用の制度に関する現状

－ デンマークとオランダを例に －

伊藤 伸介^{*1}、木村映善^{*2}、

*1 中央大学、*2 国立保健医療科学院、

Current Situation of the System of Secondary Use

about Medical Information

- The Cases for Denmark and Netherlands -

Shinsuke Ito^{*1}, Eizen Kimura^{*2}

*1 Chuo University, *2 National Institute of Public Health

Denmark and the Netherlands have launched systems for generating statistics based on register data such as demographic and administrative information (including medical and health care information) on both individuals and families. Data from administrative records (register data) are collected and stored in a registry are also automatically transferred to and maintained by the country's statistics bureau, with the purpose of producing official government statistics. In Denmark, various register data such as the names and medical histories of patients kept by the Danish Health Data Authority of the Ministry of Health are transmitted online to Statistics Denmark. The register data is structured in a way that allows for it to be easily shared between different registries, including those that store demographic information (resident information), address and building information, corporate information, and personal identification numbers. In Denmark and the Netherlands, register data has been made available for use by third parties under an arrangement between the government agencies that collect and maintain the data and certified organizations that wish to use the data; this arrangement is carried out in accordance with laws regarding the use and protection of personal data. Users can log in with an ID and apply for access to personal data in various fields such as medicine and health care. After direct identifier information has been deleted and personal identification numbers pseudonymized, personal data can be used in a secure remote access environments. Results of analyses that involve personal data are checked in order to ensure that no identifying information is exposed. This report describes the current status of collection, storage, and management of medical information in Denmark and the Netherlands, along with the features of the system that allows third parties to access utilize this information. This report focuses especially on the creation of official statistics.

Keywords: Register-Based Statistical System, Microdata, Administrative Data

1. 背景

1970年代以降、ヨーロッパ諸国の多くでは、市民のプライバシー意識の顕在化に伴い、すべての世帯に調査票を配布することによって実施する全数調査(センサス)の実施が困難な状況にあった。そのため、市民が行政機関に出生、死亡、結婚、離婚等の個人・家族に関する人口・社会的情報や、課税所得を記録した税務情報等の経済的な情報等、行政上の記録を用いて作成するセンサスの実施可能性を模索した。¹⁾

デンマークは、最初に調査票ベースのセンサスから行政記録から作成するレジスター(登録簿)型センサス(register-based census)に移行した国として知られている。行政機関は国民の個人情報行政記録として管理するが、国民に対して行政サービスの提供を行うという国・行政機関と国民との間の信頼関係が、こうしたレジスター型センサスの成立を可能にしている。²⁾

なぜ、デンマークではレジスター・ベースの統計に移行することが可能になったのだろうか。当時、ノルウェー統計局の職員で、後にベルゲン大学の教授となった Svein Nordbotten 氏によってレジスターによる統計システムの構想が発表されたのは、1960年における北欧の統計局長会議の場においてであった。³⁾ Nordbotten 氏を始めとする先駆者は、統計を作

成する上では中央集権的なレジスター・ベースの統計システムのほうが効率的であることを強く認識していた。こうしたことから、レジスター・ベースの統計に移行する前は、各省庁がそれぞれ必要な統計を作成していたものの、各種のレジスターを創設し、個人や企業の情報を行政記録としてレジスターに登録し、その情報を収集するシステムが確立されるだけでなく、レジスターに保管されている行政記録データのリンケージを通じて、様々な統計目的で行政記録を再利用するためのシステムが指向された。その一方で、当時は、現在のようなIT技術もなかったことから、こうしたレジスターに収集・保管された行政記録データ(以下「レジスターデータ」と呼称)の保管と移送を可能にするための仕組みづくりには、時間を要したと言える。

デンマークにおいては、中央人口レジスターが、個人の人口社会的属性に関する行政記録情報の一元的な管理を行っている。中央人口レジスターは個人識別番号を管理しており、個人や世帯の人口社会的属性として、名前、年齢、性別、結婚状態、市民権、出生地と住所を把握しているだけでなく、家族情報(母親、父親と配偶者の個人識別番号)もレジスターに登録している。中央人口レジスターが管理する個人識別番号だけでなく、建物・住居レジスターが保有する住所番号や企業・事業所レジスターが管理する勤務先企業番号は、個人

ひとり 1 人のレコードに含まれることから、これらのレジスター間の連結は可能である。また、世帯に関する識別子は存在しないが、個人識別番号と住所番号によって、世帯を特定することもできる。それによって、同一の世帯に含まれる世帯構成員同士をリンクすることも可能であることから、世帯構成員の属性(ex. 世帯主の子供や配偶者の属性)を把握することが可能になっている。

この個人識別番号は、税務情報、社会給付状況、教育情報、雇用情報といった社会人口的な属性情報だけでなく、個々の医療データや健康データを含む行政記録データ(以下「レジスターデータ」)に関しても個々のレコードに付与されている。このことから、医療健康に関するレジスターデータとの連結も可能になっている。

レジスターは、各担当省庁で管理しているが、統計を作成するために、必要なレジスターデータはすべて自動的にデンマーク統計局に集められ、デンマーク統計局内に保管されるシステムが確立されている。デンマークにおける医療保健データに関しても、保健省(ministry of health)のデンマーク健康データ局(Danish Health Data Authority)で保管されている様々なレジスターデータが移送される。したがって、患者の病名や病歴などを含む情報についても、統計作成のために必要な情報は、すべてオンラインで送られている。

オランダも、調査票ベースの統計からレジスター・ベースの統計に移行した国の1つである。オランダにおいては、1971年に、オランダ統計局によって、調査票ベースのセンサスが実施されたが、調査における非回答やコストの削減の観点から、レジスターに基づく統計の作成に徐々に移行するようになった。1981年と1991年においては、人口レジスターと調査データを組み合わせることによって、センサスが実施された。⁴⁾

その後、1990年代にレジスターが整備されると、レジスターに保管されている様々な行政記録情報がオランダ統計局に移送された後に、オランダ統計局に集められたレジスターデータに基づいて人口センサス等の公的統計が作成されてきた。レジスターデータに該当しない項目にのみ、オランダ統計局は、標本調査を実施することが認められている。こうしたことから、オランダ統計局では、統計を作成するためにレジスターデータの整備が進められてきた。

本稿の目的は、オランダとデンマークを例に、行政記録データの利用状況を明らかにするだけでなく、オランダとデンマークにおける医療健康データの二次利用の方向性を洞察することである。

2. 方法

本研究では、デンマークとオランダに関する伊藤の先行研究⁵⁾に基づきながら、文献調査によって、デンマークとオランダにおける医療健康データに関する二次利用について情報収集を試みた。その後、筆者の1人である伊藤がデンマークとオランダを訪問し、関係する部局および専門家に対してヒアリング調査を実施した。2019年2月21日～26日にデンマーク、2019年2月27日と3月1日にオランダを対象に、それぞれ関係部局の職員(元職員を含む)および専門家に対して調査を行った。本稿は、デンマーク統計局(Statistics Denmark)とデンマーク保健データ局の担当職員、コペンハーゲン・ビジネススクール(Copenhagen Business School=CBS)の研究者に対するヒアリング調査、さらには、オランダ統計局(Statistics Netherlands)の担当職員2名、Vektisの関係者1名に行ったヒアリング調査の結果をもとに作成したものである。

なお、伊藤は、総務省統計局で開催されたSDC(Statistical

Disclosure Control)に関する研修会(2019年2月19日～22日)に参加しており(1日目のみ参加)、本稿の作成にあたっては、研修会で用いられたオランダ統計局の担当職員による発表資料も参考にした。

なお、本研究は、内閣官房(健康・医療戦略室)の調査事業として実施された匿名加工医療情報の利活用に関する調査の内容に基づいている。⁶⁾

3. 結果

3.1 デンマークにおける医療健康データの二次利用

公的統計のマイクロデータに関しては、さまざまな形態による提供が進められてきた。具体的には、複数の属性と属性値から構成されるレコード群であるマイクロデータに関しては、①住所や氏名といった直接的な識別子は除かれているものの、それ以外の匿名化処理が施されていない個票データ(confidential data)の提供サービス、②匿名化マイクロデータ(個票データに匿名化処理が施されたデータ、anonymized microdata)の作成・提供がなされるだけでなく、さらに、③オーダーメイドによる集計結果表の提供、④オンデマンドによる集計サービス(リモート集計、リモートエグゼキューション(remote execution))という形態がある。

デンマークにおいては、1980年代半ばから、学術研究のためのマイクロデータの利用に対する関心が高まってきた。そこで、1988年にデンマーク統計局内にセキュアな環境であるオンサイト施設が設置されたが、2001年から、研究者が個票データへの直接のアクセスを可能にするリモートアクセスのサービスが開始された。その後、2008年には、オンサイト施設は閉鎖され、リモートアクセスのサービスが拡充されている。デンマークの場合、リモートアクセスを通じてのみ、個票データ(非識別データ(deidentified data))の利用サービスが行われていることが特徴的であって、デンマークでは、他の欧米諸国のように、匿名化マイクロデータのような形でのマイクロデータの提供はなされていない。そのため、ひとり1人に備わっている個人識別番号(ID)を仮名化されたIDに置き換えることが、デンマークにとって唯一といってよい秘匿の方法だと言える。

デンマークにおいては、「認証された機関(authorized institution)」を通じてのみ、マイクロデータにアクセスすることが可能である。大学等の学術研究機関が認証された機関となるためには、安定した研究環境(stable environment)があり、責任のある管理者とマイクロデータを扱った経験を持つ研究者がいることが要件であって、そのためにデンマーク統計局に申請をしなければならない。そして、一般データ保護規則(GDPR)にしたがって、デンマーク統計局は認証された機関と協定を結んでいる。マイクロデータにアクセスするには、認証された機関に所属している研究者及びその共同研究者であることが求められる。なお、デンマークの場合、研究者は、研究室の外からでも、ラップトップのPCを用いて、リモートアクセスを行うことが可能である。認証された機関に所属する研究者が外国に滞在している場合でも、リモートアクセスを行うことができる。

個票データの利用を希望する研究者は、個票データの利用申請を行い、誓約書に署名した上で、利用することが可能である。さらに、人口社会、保健衛生、所得、教育、労働市場、企業経済等の様々な分野の個票データが利用可能である。2000以上の研究プロジェクトによって個票データの利用が行われている(2019年2月現在)。

医療健康データについては、1977 年から現時点まで、入院状況(hospitalization)や診断内容(diagnosis)等が把握可能な国民患者レジスター(National Register of Patients)、1995 年～2018 年までの総人口データや薬の処方箋(prescription medicines)が捕捉されている医薬品統計レジスター(Register of Medical Product Statistics)といったレジスターが存在しており、医者への受診状況等の健康保険情報から得られるデータも研究のためにアクセスすることができる。デンマーク統計局では、利用を希望する研究プロジェクトに対して、「知る必要性(need to know)」原則に基づいて、利用者において必要なマイクロデータのみアクセスを認めている。

デンマーク統計局においては、リモートアクセスによる個票データの分析結果(集計表、回帰分析等)に関するチェックについては、すべての分析結果ではなく、一部の結果をランダムに選んだ上で、職員が事後的にマニュアルでチェックを行う。職員が分析結果をチェックする上では、セルに含まれる度数は一定の数値(例えば 3)以上でなければならないという閾値ルール(thresholding rule)のような基準が採用されている。また、審査担当者は、研究者が分析結果を持ち出す場合のデータの容量をチェックしており、容量が大きいデータについては、個票データあるいはそれに準じた高次元の集計表である可能性が高いと考えられる。ただ基本的には、デンマークにおいても、分析結果のチェックに関しては統計作成部局と研究者との信頼関係が重視されている。

2018 年から、デンマーク統計局によって開発された分析結果のスキャンシステムによって、研究者が分析結果のチェックを自動的に行うことが可能になった。具体的な基準としては、作成した集計表において、セルに含まれる度数は 3 以上であるといった閾値ルールが指摘される。ヒアリングを行った CBS のガイドラインによれば、スキャンシステムの基準として、集計表に含まれるセルの閾値を 5 に設定している。また、平均や標準偏差といった基本統計量の計算や回帰分析等の推定結果の算出においても、対象となる観測数を 5 以上とすることを定めている。なお、CBS のガイドラインでは、最大値、最小値、中位数、パーセンタイルさらには散布図を出力結果として持ち出すことは認められていない。

分析結果をチェック後に、自分のメールアドレスに添付して、送付することができる。分析結果については、個体を特定することが不可能な分析結果のみを研究プロジェクトのサーバーから送付することができる。このスキャンシステムにおいては、マイクロデータに含まれる個体が特定されないように処理されたあとのデータを利用者の責任で分析結果を持ち出すことが可能な仕組みとなっている。

なお、デンマーク統計局の他にも、

3.2 オランダにおける医療健康データの二次利用

オランダの場合、オランダ統計局で保管されているレジスターデータは、公的統計のマイクロデータという形で提供されている。先述の通り、公的統計マイクロデータの提供形態は様々であるが、以下では、匿名化マイクロデータと個票データの提供に焦点を当てる。

匿名化マイクロデータについては、例えば、1971 年、2001 年と 2011 年の人口センサスが 1%の標本データとして利用可能である。匿名化マイクロデータの作成においては、主として非攪乱的な手法(non-perturbative method)が適用されているが、攪乱的な手法(perturbation)を適用している場合もある。また、オランダ統計局は、 μ -Argus と呼ばれるマイクロデータの秘匿処

理用のソフトウェアを用いて、public use microdata files の作成を行っている。⁷⁾

個票データについては、オンサイト施設(オランダ統計局内部に 2 か所)やリモートアクセスによる利用が可能である。1998 年から、オランダ統計局においてオンサイト施設による個票データの提供が始まった。リモートアクセスについては、2005 年にパイロット版のシステムが開始され、2006 年から本格運用が開始された。現在は、基本的には研究室に備えられている PC からリモートアクセスで個票データにアクセスするという方式が一般的になっている。リモートアクセスでは、安全性の観点から、citrix というリモートデスクトップ環境を提供するシステムを経由して、研究室の PC から個票データのアクセスがなされるので、画面上でデータを閲覧・操作はできるが、研究室の PC にデータは残らない。現在、年間で数百人の利用者がリモートアクセスを通じて、個票データのアクセスを行っている。

利用者は、オランダ統計局の職員の同レベルの(名前と住所は削除されている)個票データにアクセスすることができる。ただし、分析に必要な変数のみがアクセス可能となっている。レジスターデータのリンケージについては、利用者が ID に直接アクセスすることができないデンマークとは異なり、オンサイト施設およびリモートアクセスのいずれにおいても、オランダ統計局が仮名化された ID を用いてリンケージを行った上で、個票データを提供することもできるし、研究者が自らリンケージを行うことも可能になっている。

リモートアクセスやオンサイト施設で個票データを分析した後、研究者が分析結果を持ち出したい場合に、秘匿の観点から分析結果のチェックが必要になる。集計表や回帰分析の結果等の分析結果のチェックに関しては、オランダ統計局における 5 名の担当職員が、マニュアルによるチェックを 2～3 日で行っている。オランダ統計局において数多くの判断基準があり、基準が満たされていれば、分析結果が利用者に提供するという仕組みがとられている。

オランダ統計局では、収集されたレジスターデータを用いて、様々な形態での二次利用が展開されている。この中には医療健康に関するレジスターデータも含まれる。病院で記録された医療健康情報は、DHD(=Dutch Hospital Data)という組織を通じて一元的に集められており、健康統計(health statistics)の作成に必要な情報は、DHD からオランダ統計局に移送される。医療健康データに含まれるセンシティブな情報についても、オランダでは、オンサイト施設やリモートアクセスの形態で、個票データとして研究者に提供されている。したがって、医療健康データに関しても、オランダ統計局の職員がアクセスするレジスターデータと同じレベルの詳細な情報を有するレジスターデータを利用することができる。なお、研究者が、医療健康データにおける分析を行った後に、分析結果を持ち出す際には、他の社会人口の属性や経済的属性を用いて分析した場合と同様の基準でチェックが行われている。

なお、オランダ統計局以外で行われている医療健康データの提供の事例として、(非営利の組織である)Vektis による病院の請求書(invoice)データの提供を指摘することができる。病院の請求書がオランダの民間保険会社に送られた後、請求書に関するデータは、Vektis にオンラインで移送され、Vektis 内部で保管されている。病院の請求書データには、個人の識別番号が含まれているが、Vektis 内部で個人の識別番号が仮名化されている。なお、この請求書に関する情報の一部は、オランダ統計局にも Vektis を通じて送られているが、Vektis においても、保管している仮名化された請求書データ

を研究者に提供するための仕組みが存在する。具体的には、識別番号が仮名化されているだけでなく、地域の分類区分を粗くし、出生年月を出生年に粗くするといった匿名化の措置が取られた上で、請求書データが研究者に提供され、請求書データに含まれる医療・健康情報を用いて、学術研究目的で実証研究を行うことが可能になっている。

4. 考察

本稿では、デンマークとオランダにおける医療情報の二次利用の制度に関する現状について明らかにした。デンマーク統計局の事例によれば、個票データの保護に関する法制度、さらにはデータを提供する政府機関と認証された機関に所属する研究者との間の信頼関係に基づいて、直接的な識別子の削除と個人識別番号の仮名化のみが適用された個票データの利用が、リモートアクセスの形で可能になっている。さらに、デンマーク統計局で設けられた基準にしたがって分析結果のチェックを行うことによって、個体情報が露見しないような形で分析結果を持ち出すこともできる。

オランダ統計局の場合、デンマークとは異なり、オンサイト施設での利用あるいはリモートアクセスでの利用のいずれにおいても、研究者自身が仮名化されたIDにアクセスして直接リンケージを行うことが可能なことは、オランダにおける重要な特徴と言えよう。それに対して、オランダでは、デンマークのような研究者自身が分析結果のチェックを行うスキームシステムは存在しない。そのため、研究者が集計表や回帰分析の結果等の分析結果を持ち出す上で、オランダ統計局で設けられた基準にしたがって、統計局の職員がマニュアルで分析結果のチェックを行い、分析結果から個体情報が漏洩しないような措置がとられている。

デンマークにおいては、GDPR に沿った形で 2018 年デンマークデータ保護法(Danish Act on Protection of Data 2018)が制定されたが、デンマーク統計局等が保有するレジスターデータの二次利用は、デンマークデータ保護法においても認められている。その理由としては、GDPR にも同様に明記されているが、2018 年デンマークデータ保護法に基づいて、統計目的で集められた情報を他の統計目的や研究目的のために加工することが可能になっている(個人情報を経営目的や学術研究目的で利用することは GDPR の適用除外という扱いになる)からである。

デンマークとオランダの事例は、わが国において二次利用の観点から行政記録データのさらなる利活用を図る上では、参考になる点が少なくないと思われる。その一方で、わが国の公的統計のマイクロデータのさらなる利用促進の観点からは、公的統計の個票データと行政記録データ、あるいは行政記録データ同士のリンケージを可能にするためのデータの整備が必要になるだろう。また、次世代医療基盤法に基づく医療健康データの利活用推進を図るための制度設計を目指すうえで、オンサイト施設やリモートアクセスによる医療健康データの利用可能性、さらには医療健康関連のレジスターデータを利用した後の分析結果のチェック方法も検討課題となりうるだろう。

参考文献

- 1) 工藤弘安. レジスター・ベースの人口・住宅センサス. 成城大学経済研究. 1995; 127: 71-87.
- 2) United Nations Economic Commission for Europe *Register-Based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics*, United Nations. 2007
- 3) Thygesen, L. The Importance of the Archive Statistical Idea for the

Development of Social Statistics and Population and Housing Censuses in Denmark. 2010

https://www.researchgate.net/publication/302449598_The_importance_of_the_archive_statistical_idea_for_the_development_of_social_statistics_and_population_and_housing_censuses_in_Denmark【2019年3月7日アクセス】

- 4) 森博美. オランダの virtual census について. 熊本学園大学経済論集. 2009; 15(3・4): 35-58.
- 5) 伊藤伸介. 公的統計における行政記録データの利活用についてーデンマーク, オランダとイギリスの現状ー. 経済学論纂(中央大学). 2017; 58(1): 1-17.
- 6) 内閣官房. 匿名加工医療情報の利活用に関する調査報告書. 2019
- 7) Nordholt, E.S.. Access to Microdata in the Netherlands: from a Cold War to Cooperation Projects. Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Ottawa, Canada. 2013: 1-11.