

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))  
総合研究報告書

リンケージデータだからこそ示すことのできる要介護発生前から死亡までの軌跡  
—要介護発生の背景、医療介護費用に着目した分析

研究代表者 伊藤 智子 筑波大学医学医療系

研究要旨

本研究では、リンケージデータに関するレビューとリンケージデータを用いたユースケースを示すことで、公衆衛生におけるリンケージデータの今後の利活用を検討した。その結果、我が国のデータリンケージにおいては、官庁横断的で専門的な運営体制、データマネージに長けた人材と設備が必要であり、また、資金面における安定的なバックアップも重要であると考えられた。またユースケースからは、各自治体が保有する複数のデータをリンケージ活用することの有用性が示唆され、各自治体が独自にリンケージデータ分析を行い、地域に合った政策を検討する上で欠かせない方法であると考えられた。

研究分担者

筑波大学医学医療系 田宮菜奈子  
筑波大学医学医療系 森隆浩

A. 研究目的

本研究では、リンケージデータを用いた海外の研究のレビューと、リンケージデータを用いた研究のユースケースを示すことで、リンケージデータの今後の利活用に向けた提言を目指した。そのための研究として令和元年度および令和2年度の2年間において、下記のとおり、レビュー研究4課題、ユースケース研究4課題を行った。

レビュー1 海外における公的死亡データ研究目的利用の実態—利用方法・リンケージ可能性・研究実績

レビュー2 医学主要誌におけるリンケージデータを用いた研究の動向—JAMA・NEJM・BMJ・LANCETの2019年出版分のレビューから

レビュー3 高齢者ヘルス領域における海外データリンケージの状況

レビュー4 韓国の公衆衛生分野における公的データリンケージの状況

ユースケース1 リンケージデータによる介護予防サービス利用の予防効果の検討

ユースケース2 リンケージデータによる入院中初回要介護認定者の特定

ユースケース3 要介護認定前の医療サービスの利用実態

ユースケース4 死亡前1年間におけるサービス費用

B. 研究方法

レビュー1

公的データの突合が活発に行われている米国、英国、オランダ、デンマーク、スウェーデン、フィンランド、ノルウェー、オーストラリアにおける公的死亡データを挙げ、その突合可能性について方法や突合実績を踏まえてレビューした。

レビュー2

レビュー対象は、高いImpact Factorを誇る主要4誌 (Journal of American Medical

Association: JAMA, The New England of Medicine: NEJM, British Medical Journal: BMJ, The Lancet) の 2019 年発行分とした。検索語は使用せず、発行されたすべての原著論文を対象とし、研究で使用されているデータがリンケージデータである論文を選出した。

### レビュー3

高齢者ヘルス領域におけるリンケージデータを用いた研究の実績について、PubMed を用いて、検索語「Elderly Care Linkage」によって文献を検索した。検索にヒットした文献のうち、主要なリンケージデータソースを抽出し、そのデータ提供の方法や研究実績についてレビューした。

### レビュー4

韓国における公的データリンケージ活用の実態について、韓国の公衆衛生領域における公的ビッグデータ管理の仕組み、データリンケージの方法、関連する研究実績の紹介、研究における限界等の観点から情報収集した。

### ユースケース 1

2012 年 4 月から 2015 年 3 月までの千葉県柏市の行政データを用いた。データは、①後期高齢者医療保険、②介護保険のレセプトデータ、③要介護認定のための認定調査、④住民データ、④介護保険料についての行政データを用いた。各データは、研究目的に新たに付与された ID 番号を用いてリンケージされた。対象において、初回認定から 6 か月間を曝露期間とし、この期間における介護保険の介護予防サービスの利用が一度でもあった者を介入があった群とした。その後の要介護度 1 以上に悪化するまでの期間を従属変数とした Cox 回帰モデルを分析した。

### ユースケース 2

千葉県柏市（関東圏、地方都市）の医療レセプトデータ（国民健康保険および後期高齢者）、介護保険のレセプトデータ（2012～2013 年）および認定調査データ（2008～2016

年）を用いた。各データ間のリンケージは研究用に付与された ID を用いて行った。初回要介護認定は、認定調査データにおいて、新規で情報が作成されており、かつ同一の ID 内で遡って認定情報がないケースとした。また初回要介護認定が特定された集団の中で、初回要介護認定が入院中に行われた者とそれ以外の者に群別し、1 年後の生存をロジスティック回帰モデルにより比較した。

### ユースケース 3

A 市（人口 5 万人規模、高齢化率全国平均との差+5%）の医療保険データと介護保険データとの個人単位での連結データ（以下、医療介護データ）（2012～2016 年）を用いた。75 歳以上の対象において、初回要介護認定を受けた者を認定群とした。対照群は、認定を受けていない者から性別、年齢、認定時期によって認定群とマッチングした者を選定した（1:1 マッチング）。認定群と対照群において、認定時期（対照群においてはマッチングされた時期）の前 1 年間における医療サービスの利用状況を比較した。

### ユースケース 4

B 市（人口 40 万人規模、高齢化率全国平均との差-3%）の医療保険データと介護保険データとの個人単位での連結データ（以下、医療介護データ）、認定調査データおよび死亡/転出データ（2012～2014 年）を用いた。2013 年 4 月～2015 年 3 月の間に死亡した高齢者を対象とし、死亡月から 12 か月分の医療保険請求額（ただし、医科請求分のみ）、介護保険請求額、および医療と介護の合計額を算出し、性別、死亡時年齢および死亡時要介護度別にみた。

（倫理面への配慮）

本研究で用いるデータは、個人情報を含まない。また本研究は筑波大学医学医療系倫理委員会の承認（承認日：2019 年 12 月 2 日、承認番号：1448）を得て実施した。

## C. 研究結果

### レビュー1

本研究では、海外8か国における公的死亡データの研究利用について調査した。その結果、各国は公的な死亡データの研究目的利用を可能にしており（英国は部分的）、データを用いて出産、小児から高齢者といったあらゆるライフステージの対象について分析を行っていた。

### レビュー2

2019年の医学主要4誌において、リンケージデータを用いた研究は発行論文数の中で0.5-3%と決してメジャーな研究方法ではなかった。

### レビュー3

検索の結果、254件（2021年2月時点）が該当し、そのうち、リンケージデータに関する文献は96件であった。SEER-Medicare linked database（米国）、Clinical Practice Research Datalink（英国）、Survey of Health, Ageing and Retirement in Europe（EU）、The Centre for Health Record Linkage（オーストラリア、ニューサウスウェールズ州）、ICES（カナダ、オンタリオ州）、Rochester Epidemiology Project medical record（米国、ミネソタ州オルムステッド郡）の7種類のデータベースが見いだされ、各データベースにおける運営体制やデータ内容、データ利用方法などを比較した。報告事例の7件中3件が米国における事例であり、英国1件、EU1件、オーストラリア1件、カナダ1件である。データ間のリンケージ方法は社会保障番号などの個人番号を用いた **Deterministic Linkage** が4件であった。また、データの利用料については、4件が有料であった。データベース提供事業の運営組織は様々であり、国立研究所や大学内の特別部局、あるいはNPO組織などによるものであった。

### レビュー4

韓国では NHIS（National Health Insurance Service）のレセプトデータを中心

にその他のデータともリンケージして、研究目的にデータ提供しており、その規模は全被保険者をカバーする悉皆データとなっている。NHIS レセプトデータは2017年で700件以上の提供実績があり、様々な観点から研究利用されている。NHIS のデータは National Health Insurance Sharing Service（NHISS）によってサンプルコホートデータベース（DB）、カスタマイズDB、健康疾患指数DBなどとして提供され、国民健康情報データの適用から研究成果の共有に至るまで、利用者のアクセスと利用の利便性を高めている。おおまかな研究テーマに沿うように既にいくつかのデータリンケージがなされセットになって提供可能となっている点が特徴であり、例えば高齢者に関するデータセットとしては、死亡情報+障害の程度+社会的経済的情報+医療サービスの利用状況+介護サービスの利用状況がリンケージされている。データリンケージは公的個人番号による **Deterministic linkage** である。近年では、こうしたNHISS提供のリンケージデータを利用した報告が出てきており（J Bahk, et al. J Korean Med Sci. 2017、SG Seong, et al. Int J Epidemiol. 2017）、これからさらに研究が加速されるものと考えられる。

### ユースケース1

本研究では医療介護のリンケージデータを用いて、介護予防サービス利用の効果を検証した。その結果、全対象においては、介護予防サービス利用は有意な効果を示さなかったが、85歳以上かつ要支援1という限定された対象においては、有意な予防効果を示した。

### ユースケース2

最終分析対象は5,811人であり、初回要介護認定が入院中に行われた者は876人（15.1%）であった。そのうち1年後に死亡していた者は284人（32.4%）であり、入院中でなかった者における死亡数に比して多かった（オッズ比2.65、95%信頼区間2.26-3.13）。

これは基本属性や認定状況（要介護度や必要な医療処置）によって調整した結果でも同様であった（調整済みオッズ比 1.52、95%信頼区間 1.26-1.84）。

#### ユースケース 3

本分析においては、認定群 1,342 人、対照群 1,342 人の計 2,684 人が対象であった。ポアソン回帰による分析の結果、要介護認定された者は通院日数が多い傾向がみられた（相対リスク比 1.27、95%信頼区間 1.25-1.29）。また、要介護認定者の中で、要介護度別に比較したところ、要支援 1-2 に対して、要介護 1-2 (0.75、0.73-0.77) および要介護 3-5 (0.72、0.70-0.74) では通院日数が少なくなる傾向がみられた。また、ポアソン回帰による分析の結果、要介護認定された者は入院日数が多い傾向がみられた（相対リスク比 3.04、95%信頼区間 2.92-3.15）。また、要介護認定者の中で、要介護度別に比較したところ、要支援 1-2 に対して、要介護 1-2 (1.20、1.14-1.27) および要介護 3-5 (2.42、2.29-2.54) では入院日数が多くなる傾向がみられた。

#### ユースケース 4

最終分析対象は 3,468 人であった。各請求額について、初めに介護保険請求額では、死亡時年齢が高いほど、女性、そして死亡時要介護度が重度であるほど、請求額が高くなる傾向がみられた。一方で、医療保険請求額では、死亡時年齢が高いほど、男性、死亡時要介護度が重度であるほど、請求額が低くなる傾向がみられた。介護保険請求額と医療保険請求額との合計額については、死亡時年齢が高いほど、男性、死亡時要介護度が重度であるほど、請求額が低くなる傾向がみられた。重回帰分析の結果、介護保険請求額については、死亡時年齢が高いほど、女性、死亡時要介護度が重度であるほど請求額が高くなる傾向がみられた。医療保険請求額では、死亡時年齢が 80-85 歳の者に対し、95-99 歳、および 100+ 歳における請求額が低くなる傾向がみられた。また、女性で低くなる傾向があ

り、死亡時要介護度が重度であるほど請求額も高くなる傾向がみられた。合計額では、死亡時年齢では 100+ 歳のみ有意に請求額が低くなる傾向がみられ、死亡時要介護度が重度であるほど請求額が高くなる傾向がみられた。性別に有意な差はみられなかった。

#### D. 考察

##### レビューより考察

##### **公的リンケージデータの持つ研究上の利益**

広く調査された公的データを用いることは、ほぼ全国民の死亡を把握できるという点から、悉皆性のある研究を可能にする。悉皆性のある研究では、選択バイアスをほぼなくすることができるという、疫学上、大変重要な利点を得ることができ、疫学研究で議論されやすい結果の代表性や一般化可能性といった課題をデータの特性によって解消することができる。そして、リンケージするデータによっては、超長期的な観察期間を設定することが可能であり、特定の因子や介入による長期的アウトカムの検証が可能になる。

また、介入研究には倫理上、実施困難な研究仮説がある。例えば、レビュー 2 で紹介した妊娠中のインフルエンザワクチン接種の児への影響を検証した研究は、妊娠時にインフルエンザワクチンの接種の有無を研究目的にコントロールできるとは考えにくい。そして、この研究仮説を実証可能にしているのは、インフルエンザ接種歴を含む母親のデータと出生した児の健康状態のデータとのリンケージであると言える。また同様に、収入と平均余命との関連や貧困状態と医療サービス支出との関連といった、社会経済的地位 (Socio-economic Status, SES) を要因として扱う研究は、通常、一つのデータセットになりにくい SES と健康状態や生存の 2 つの要素を、データリンケージによって一つのデータセットにして実施されている。つまり、曝露をデータ A から、アウトカムをデータ B から特定し、そしてデータ A とデータ B と

をリンケージすることで、長期縦断的なデータセットを基に因果関係の検証を行うことができている。異なるデータそれぞれの強みを活かしたリンケージデータを作成して、これまで無かったデータセットを生み出せることは、創造的な研究仮説の検証につながっているものと考えられる。

### 公的リンケージデータの整備方法

リンケージ方法については、フィンランド、デンマーク、オランダ、ノルウェー、韓国では国民に付与されている個人番号を使用した **Deterministic linkage** が行われており、従前から示されている北欧諸国の個人番号の強みが改めて示された形である。一方、アメリカ、カナダでは個人識別情報（氏名、性別、生年月日など）の組み合わせを用いた **Probabilistic linkage** を行っていた。こうした違いの背景には、個人情報保護に関する法的問題がある。一方で、個人番号による個人特定が有効であるかというように、個人番号が適切に国民に付与されているかという個人番号の整備状況にも大きく関係しているとみられた。

また、レビュー3で報告した世界有数のデータベースのリンケージ事例では、それぞれのデータベース管理およびリンケージ処理は、半公的あるいは第三セクター的な位置付けの専門組織によって行われていた。こうした運営方法は、我が国のように各官庁の既存の組織内の限られたソースの中で管理されていることに比較して、多くの案件を迅速に処理することを可能にしているとみられた。また、データマネージに長けた人材と設備が不可欠であり、資金面における安定的なバックアップも重要であると考えられた。その結果、提供されたリンケージデータを用いてタイムリーな研究結果に繋がり、データ活用の有効なフローの構築に貢献するものと考えられる。

特にレビュー4で報告した韓国の公衆衛生

分野で公的データのリンケージ状況は特筆すべきであり、国民健康保険公団の「国民健康情報 DB」基盤のビッグデータ、4つの公共機関（疾病管理庁、国民健康保険公団、健康保険審査評価院、国立がんセンター）が協力して作った「保健医療ビッグデータプラットフォーム」を通じて、研究者たちが公益目的で活用できるデータを提供するために国を挙げての取り組みがなされていた。その結果、国民健康保険公団の国民健康情報 DB ビッグデータの場合、研究支援資料の提供がかなり速い速度で増加し、研究者のアクセシビリティが向上しているといえる。

しかし、我が国においては、例えばレビュー1で示したような公的死亡データとのリンケージは許されておらず、最大かつ最終のアウトカムである「死」の検証が、我が国では行き届かないでいる。今後、我が国のデータヘルスをより促進するためには、こうした人口動態統計といった健康政策上重要なデータのリンケージ活用が望まれるところであり、そのためには研究者への利用の門戸を開くとともに、データ運営上の安全を担保するための仕組みの整備が求められるところである。

### ユースケースより考察

#### 自治体リンケージデータの有用性と限界

本研究では、自治体が所有する公的データのリンケージデータの有効性を議論すべく、4件のユースケース研究を行った。こうした自治体データを用いた研究では、特定の限定されたコホートとは異なって、対象の選択バイアスをできるだけ減少できる研究デザインとなっている。また、医療保険データをリンケージすることで、対象の医療的な側面による調整を可能にした。さらに保険料データがリンケージされたことで、社会経済的要因（SES）による調整が可能になったことは大変重要である。こうした行政データを個人間でリンケージして、扱うことのできる変数を

多様にすることは、結果の強固性につながることであり、リンケージデータ研究の強みであると言える。また、ユースケース 1、2、3 においては、対象の選定等で住民基本台帳データによる死亡の同定を行っている。死亡は、競合リスクや母集団の設定において、欠かせない情報であり、自治体データでは該当の自治体の許諾のもと可能になっている。レビュー1 では、国ベースの人口動態統計のリンケージが制限されていると言及したが、自治体データにおけるこうしたデータ利用の利便性の高さは着目するところである。またユースケース 3 や 4 で報告したように、介護保険と医療保険は、特に介護が必要になった後には互いに補完するような形になっていると言え、介護保険データと医療保険データのリンケージ使用は、生涯のヘルスケアを検証する上では、欠かせない方法である。

一方で、自治体単位での分析には研究上の制限もある。まずは一地域に基づく結果であり、対象の選択バイアスに基づく一般化における限界がある。次に、自治体データにおける個人特定は、自治体内で付与された個人識別番号（住民基本台帳データに基づく番号）によって行われ、自治体の転出・転入に伴って番号が変更されると、同一個人としての識別ができなくなる。これは国単位のデータであっても、個人番号（いわゆるマイナンバー）が付与されない限り、住民票のある自治体が付与した個人番号が識別番号となり、自治体とセットでないと個人が識別できない。

#### 自治体リンケージデータ活用上の課題

こうしたレセプトデータに基づく検証は、医療介護のリンケージデータであるからこそ可能となるものである。現在、地方自治体が自らデータをリンケージして分析し、自らの地域に応じた施策を講じることが求められている。しかしながら、そうした人材を自治体が獲得することは現状困難であると推察され、いくつかの自治体では大学との連携

のもとデータ活用の動きがみられている。他機関との連携は自治体の特徴によって左右されているとみられ、全国一律の動きではない。今後、さらなる地域格差の拡大を抑制する上でも、こうした状況に対し、さらなるリンケージデータ研究が普及・発展し、リンケージ方法やそのデータの分析方法についてプラットフォーム化されるような展開が求められるところである。

#### E. 結論

本研究では、リンケージデータに関するレビューとリンケージデータを用いたユースケースを示すことで、公衆衛生におけるリンケージデータの今後の利活用を検討した。その結果、我が国のデータリンケージにおいては、官庁横断的で専門的な運営体制、データマネージに長けた人材と設備が必要であり、また、資金面における安定的なバックアップも重要であると考えられた。またユースケースからは、各自治体が保有する複数のデータをリンケージ活用することの有用性が示唆され、各自治体が独自にリンケージデータ分析を行い、地域に合った政策を検討する上で欠かせない方法であると考えられた。

#### F. 研究発表

##### 1. 論文発表

Tomoko Ito, Takahiro Mori, Hideto Takahashi, Natsumi Shimafuji, Katsuya Iijima, Satoru Yoshie, Nanako Tamiya. Prevention services via public long-term care insurance can be effective among a specific group of older adults in Japan. BMC Health Services Research. In press.

##### 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