

人口の健康・疾病構造の変化にともなう複合死因の分析手法の開発と その妥当性の評価のための研究

研究代表者 別府志海 国立社会保障・人口問題研究所

研究要旨

日本の死亡統計である『人口動態統計』は、死亡診断書に記載された複数の死因をもとに、WHOが定める方法により「原死因」を一つに特定し、これを掲載している。しかし現代では一人がいくつもの疾病を同時に罹患していることが珍しくなくなっており、死因を一つに限定することで疾病間の関連など逆に見えなくなる情報も存在する。こうした複数の死因（複合死因）に関する人口ベースのマクロの研究は、データが得られなかったこともあって我が国ではほとんど行われていない。

本研究は、原死因を決定する以前の死亡診断書データ（死亡個票データ）を用い、そこに記載されている各死因を用いた分析手法を探るとともに、分析結果の妥当性について評価し、長寿化を進展する要因を複合死因の視点から分析を試みるものである。これらの結果は健康・疾病構造の変化にともなった医療・介護・福祉への効果的な介入方法の検討に資する基礎資料となる。

初年度目は、新型コロナウイルス感染症の感染拡大に配慮しつつ研究活動を行った。研究を進める上で必須となる複合死因データについて、厚生労働省「人口動態統計」の死亡票および「死亡個票」の二次利用によるデータの提供を受けた。このうち死亡個票データはすべてクリーニングの行われていないテキストデータであることから、提供を受けたデータについて独自にコード化して複合死因についての試行的な分析を行った。

研究2年目は、令和2年度までの死亡個票データを用いて、コード化の精緻化、新型コロナウイルス感染症（COVID-19）の複合死因分析、複合死因のネットワーク分析、死因簡単分類による複合死因の分析を行った。

研究3年目は、コード化の精緻化、死因簡単分類に振り分けて原死因別の原死因死亡割合、複合死因割合の推移と原死因別複合死因順位を算定、糖尿病および新型コロナウイルス感染症（COVID-19）の複合死因分析、複合死因についての指標であるSRMU・CDAIやネットワーク分析を適用し、年次変化等を含めた手法の評価を行った。また複合死因ネットワーク（Multicausal network）が開催され、各国の複合死因研究者と情報交換を行った。

複合死因研究は国際的にもまだ開発途上であり、また分析を行う上でのハードルは低いものの、得られる情報は豊富である。本研究の進展が貢献できる余地は小さくないといえよう。なお、これらの研究結果・データは国立社会保障・人口問題研究所ホームページ上で公表している。

研究分担者：

林玲子 国立社会保障・人口問題研究所
副所長
石井太 慶應義塾大学経済学部 教授

篠原恵美子 東京大学大学院 医学系研究
科 医療 AI 開発学講座 特任
助教

A. 研究目的

日本の死亡統計である『人口動態統計』は、死亡診断書に記載された複数の死因をもとに、WHO が定める方法により「原死因」を一つに特定し、これを掲載している。しかし現代では一人がいくつもの疾病を同時に罹患していることが珍しくなくなっており、死因を一つに限定することで疾病間の関連など逆に見えなくなる情報も存在する。こうした複数の死因（複合死因）に関する人口ベースのマクロの研究は、データが得られなかったこともあって我が国ではほとんど行われていない。

本研究は、原死因を決定する以前の死亡個票データを用いることにより、死因間の関連を分析し、長寿化を進展する要因を死因構造から分析するものである。すでに申請者らはわが国の複合死因集計・分析に着手し、フランス国立人口研究所 (INED) 等を中心に結成されている複合死因研究の国際ネットワーク (MultiCause network) での研究交流を通じて、国際的に標準化された複合死因に関する各種指標の算定、複数の疾病が併存しやすい生活習慣病や認知症、死亡期間情報を用いて特定した突然死などに関し、複合死因・疾病と原死因の関係について分析している。

こうした研究を進めて疾病の罹患から死亡までのプロセスおよび疾病間の関係を明らかにすることにより、疾病の罹患を遅延させる等といった介入が行われた場合のシミュレーション分析、死亡や健康寿命に与える効果についての分析が可能となり、医療・介護・福祉への効果的な介入方法の検討に資する基礎資料を得られる。

B. 研究方法

死因統計にまとめられている原死因を特

定する以前の、直接死因等が記載されている死亡診断書のデータは、2003 年以降のオンラインで厚生労働省へ登録されるようになった分について死亡個票データとして保存・二次利用が可能となっている。そこで本研究では、従来の死亡統計作成に使用されている死亡票に加え、この死亡個票データを二次利用し分析する。ただし、死亡票はすでにコード化されたデータとなっているものの、死亡個票データはテキストデータでありかつデータのクリーニング等はされていない。

そこで死亡個票に記載されているテキストデータのコード化について独自にクリーニングを施し、複合死因については ICD-10 対応標準病名マスターで定義されている病名交換用コードおよび ICD-10 コードに、また死因別の期間欄は日数に変換する。

初年度はこうして分析が可能となったデータを元に、コード化した複合死因データを死亡票とマッチングを行い、原死因を含めた複合死因を用いて隣接行列を作成した。また、複合死因に関する分析方法等については主にフランス国立人口研究所 (INED) 等を中心に結成されている複合死因研究の国際ネットワーク (MultiCause network) において、分析・研究論文の収集が行われている。そこで、これを参考にしつつ、海外における研究レビューの対象として 113 種類の研究を選定してレビューを行うとともに、わが国における複合死因関係研究に関してもレビューを行った。

2 年度目は死因のコード化について、外因や「心臓関連死」などあいまいな記述内容をカテゴリ化しコードに反映させるなど改変を加えた。その上で、利用出来るようになった 2020 年のデータには新型コロナウイルス感染症 (COVID-19) による死亡が含まれていることから、この死因が死亡診断書のどこに記載されるか等について分

析を試みた。

一方、複合死因に関する分析方法等については、主にフランス国立人口研究所 (INED) 等を中心に結成されている複合死因研究の国際ネットワーク (MultiCause network) における Egidi(2018)の方法論を日本のデータに適用し、複合死因のネットワーク分析を行った。

さらに、近年では老衰が死因第3位になるに至っており、原死因が老衰である死亡が、どのような複合死因であるのかについて、記述されている内容 (傷病名および期間、手術、解剖、その他特に付言すべきことがら、備考) について集計・分析した。

3年度目は、変換精度のさらなる向上を図ったうえで平成15年から令和3年までのデータの正規化を行った。その上で ICD-10 コードから死因簡単分類に振り分けて原死因別の原死因死亡割合、複合死因割合の推移と原死因別複合死因順位を算定・図示し、その傾向を分析した。

さらに個別の死因を深掘して分析した。生活習慣病の代表例として糖尿病を分析するとともに、現下のわが国において社会的な影響が大きい新型コロナウイルス感染症について、死亡診断書のどこに記載され、併記される疾病にはどういったものが多いかについて分析を行った。

死因間の関係を表現する SRMU (Standardized Ratio of Multiple to Underlying Cause) と CDAI (Cause of Death Association Indicator) については、MultiCause Network グループが作成している標準的指標の作成方法に関するガイドラインに基づいて作成を行った。ガイドラインでは、最初に、基礎となる死因データについて、いくつかの条件に従って用意することを求めており、本研究でもこの条件に従ってデータを準備して指標の作成を行った。ネットワーク分析については、

Egidi(2018)による、複合死因データを用いた、欧州死因分類をノードとし、同一診断書に二つの死因分類による死因が記述されているときに二つのノードをエッジで結ぶ無向ネットワークを考えてネットワーク分析手法を適用する方法論に基づいて、2003～2021年の日本のデータにこの手法を適用し、複合死因間の関係分析を行った。

また、2022年度には複合死因ネットワーク (Multicausal network) が開催され、各国の複合死因研究者と情報交換を行った。

(倫理面への配慮)

本研究では独自の調査等は行っていない。

C. 研究成果

死亡診断書に記載されている各死因などの死亡情報を得るため、統計法に基づき人口動態統計の死亡票および死亡個票 (死亡診断書を転記したデータ) について二次利用申請を行い、利用可能な全期間にあたる2003～2021年についてデータの提供を受けた。

① 死亡個票データのコード化

データのうち死亡票はクリーニングされコード化されたデータとなっているが、死亡個票データは厚生労働省にオンライン登録された時点におけるテキストデータとなっており、データを入手しただけでは分析を行うことができない。

死亡個票データのうち、分析上で処理が必要となる項目は「死亡の原因」欄と「発症から死亡までの期間」欄の2種類である。死亡の原因欄は ICD-10 対応標準病名マスターで定義されている病名交換用コードおよび ICD-10 コードに、期間欄は日数に変換する。変換方法の概要は以下のとおりである。なお、「痴呆」など現在では使用されなくなった病名もデータに含まれるため、コード化の際に利用する ICD-10 対応標準

病名マスターには公開されている過去のバージョンも全て統合し用いた。

I. 「死因」欄

死因欄の記載には表記ゆれが含まれていることが考えられるため、まずこれを以下の順で処理する。

- ・1.文字レベルの正規化 (例. 頸→頸)
- ・2.語レベルの正規化 (例. 鬱血→うっ血)
- ・3.誤り訂正 (例. 十二腸→十二指腸)

1と2の違いは、前後の内容に依らず置換可能かどうかである。頸はいかなる場合も頸に置換して問題ないが、鬱は後ろが血ならば「うっ」、病ならば「うつ」である。

次に正規化した文字列を、コードの紐づいた病名や病名修飾語の列に分解する。分解処理は形態素解析器の MeCab を利用し、解析用辞書 (分解結果を構成する要素) は ICD-10 対応標準病名マスターの索引用語およびそれに含まれない頻出語のリストを用いる。これらの要素は ICD-10 コードおよび病名交換用コードと紐づけられている。

最後に、箇条書き番号等の不要な情報を削除し、複数のコードが含まれる場合にはこれらを並列とみなして分割する。さらに「心臓関連死」などあいまいな記述内容や「火災」や「墜落」といった外因をカテゴリ化しコードに反映させた。

II. 「期間」欄

期間欄も死因欄と同様に表記ゆれを含む場合があるため、漢数字からアラビア数字への変換、「約」などの語の削除を行う。「数日」「数十年」のように「数」が付く場合は後続する期間の3倍とし (例. 数日=3日、数十年=30年)、「半年」など先頭に「半」が付く場合は後続する期間の0.5倍、「2年半」など後ろに付く場合には先行する時間単位の0.5とする (例. 半年=0.5年、2年半=2.5年)。単位が日でない場合には日数表現に変換した。記載内容が日付であった

場合は死亡日との差分から日数表現に変換する。期間欄についても複数の情報が列記されていることがあるため、死因欄と同様に分解する。

② 複合死因のマッチング・記載状況

コード化した複合死因データを死亡票とマッチングを行った結果、オンライン登録が始まった2003年では9%弱にとどまっていたが、オンライン登録の広がりとともに割合も上昇して2007年には50%を超え、2019年には死亡票の99%とマッチングを行えた。死因欄別に死因順位を出すと、I欄アでは心不全や肺炎など死亡の定義と密接な死因や老衰といった記載が多いが、I欄イ～エにおいては、脳梗塞、肝硬変、糖尿病、高血圧症、慢性腎臓病と言った死因が多く記載されていた。またII欄では脳梗塞や心不全のほか、糖尿病や慢性腎臓病、高血圧症などの、いわゆる生活習慣病が多く記載されていた。

③ 死因簡単分類による複合死因の分析

原死因を含めた複合死因を用いて隣接行列を作成した結果、他の死因と深く関係している死因と、そうではない死因がみられた。前者は腎不全、糖尿病や敗血症、高血圧性疾患であり、後者は悪性新生物や肝疾患などである。また、心疾患や老衰は、他の原死因の複合死因となることが多いことが示された。血管性認知症、アルツハイマー病、パーキンソン病は老衰、肺炎、誤嚥性肺炎を複合死因に持つことが多く、一つの複合死因関連パターンを形成していた。

死因簡単分類別の原死因別に平均複合死因数をみると、いずれの原死因でも近年記載数が減ってきているが、そのうち老衰の平均複合死因数が一番少なく2020年では1.04であった。死因簡単分類のうち、合計された分類ではなく「その他の～」ではない84分類について、原死因と複合死因順位をみると、原死因で一番多いのは老衰、

次いで心不全、肺炎であり、この順位は複合死因順位も同じであった。このうち老衰死の複合死因をみると、「老衰」等と1箇所には記載がない死亡が2020年では全体の93.8%を占め、この割合は2005年の87.4%から増加している。

さらに原死因割合と複合死因割合における2003年から2021年にかけての変化を直線近似し、その傾きの値を見ると、全134の単純分類のうち、複合死因割合、原死因割合いずれも、減少傾向の死因の方が多い。死因単純分類別に、複合死因割合、原死因割合をプロットすると、つまり増加の傾向が著しいのは老衰であり、次いで誤嚥性肺炎、神経系の疾患である。逆に減少が著しいのは脳血管疾患、次いで肺炎、脳梗塞、悪性新生物などである。複合死因割合と原死因割合の傾きの相関は $R^2=0.8766$ とかなり高く、複合死因割合と原死因割合の推移は似ているといえる。

④ 糖尿病の複合死因分析

糖尿病を死因に含む死亡例について詳細に複合死因分析を行うと、糖尿病が記載される死因欄はI欄イが15%程度、II欄が65%前後であり、他の死因欄への記載は少なかった。I欄アに糖尿病が記載されている場合の他の死因をみると、他の死因欄にも糖尿病が記載されるケースが多い。他方でII欄へ糖尿病の記載がある場合は、I欄アに呼吸器系疾患のほか心不全、脳梗塞などが多く、老衰の記載も目立ち、II欄では糖尿病のほか、高血圧や脳梗塞、慢性腎臓病の記載が多くなっていた。

⑤ 新型コロナウイルス感染症 (COVID-19) の複合死因分析

COVID-19が原死因である死亡例について複合死因分析を行ったところ、2020年、2021年ともCOVID-19はI欄アへの記載が最も多く、原死因の死亡に対しておよそ2/3を占めていた。原死因がCOVID-19で

ある場合の他の死因についてCOVID-19以外に記載されている死因をみると、I欄アは肺炎が2020年に同17%、2021年に10%と最も高く、当初「新型肺炎」と呼ばれていたことと符合した。II欄では糖尿病や慢性腎臓病が多く記載されており、心不全や脳梗塞といった循環器系疾患も上位にあった。また、I欄アにCOVID-19が記載されていた場合について死亡までの平均期間を見ると、2020年は平均17.9日、2021年は19.0日であり、わずかに延びていた。

⑥ 複合死因の人口学的分析手法

SRMUは、原死因死亡に対する複合死因死亡の比を表す指標であり、年齢構成の影響を排除して比較できるように年齢調整を行って原死因死亡数に対する複合死因死亡数の比を計算する。年齢調整の基準人口としては、ガイドラインに示されている2013 European populationを用いる。一方、CDAIは、死因cの標準化出現率(SPc)に対する、原死因uに限定した場合の死因cの標準化出現率(SPc/u)の比率として算定され、原死因を限定しない場合の死因cの出現度合に対して、原死因をuに限定した場合どの程度出現度合が変化するかを100を基準として指数化したものに相当する。ただし、標準的死亡数として2009年の高所得国に関するWHO生命表に基づく死亡分布を用いることがガイドラインで示されており、本研究でもこの方法に従って算定を行った。

ネットワーク分析については、Egidi(2018)の方法と同様に、粗隣接行列を作成し、IPF法を用いて正規化を行った後、95%タイルよりも大きい関係だけを取り出してネットワーク構造を作成した。次に、Egidi(2018)と同様に、ネットワーク中心性を示す次数中心性、隣接中心性、媒介中心性を算定した。また、コミュニティ検出に関して提案されている複数のアルゴリズム

ムを適用して、死因間の結びつきが強いコミュニティへの分類を行い、比較を行うこととした。適用するアルゴリズムについては、Luke(2015)で取り上げられているコミュニティ検出アルゴリズムから、Fast-greedy[cf], Walktrap[cw], Edge-betweenness[ceb], Spinglass[cs], Label propagation[clp], Leading eigenvector [cle], Louvain[cl]の7つのアルゴリズムを用いた。

D. 考察

① 死亡個票データのコード化

変換プログラムの改変の結果、新しく追加された令和3年のデータについてもこれまでと同じ程度にコードを付与することができた。一方で、死亡の原因欄以外にもその他付言すべき事柄・備考欄には病名が記載されることがあるが、今回のコード化では対象外としている。別途調査を行った結果、死亡の原因欄には含まれないコードがこれらの欄に含まれる個票が全体の1%程度存在することが分かった。今後複合死因の分析を行う際にはこの2つの欄も対象に入れることを検討すべきと考えられる。

② 複合死因のマッチング・記載状況

コード化された死亡個票データを死亡票とマッチングさせたところ、コード化された死因が2つ以上ある死亡の割合は5割ほどであり、したがって死亡全体の約半数について何らかの複合死因分析が可能となる。複合死因分析では死亡者の死因構造しか扱うことが出来ないものの、それでも近年では60万件を超える数を分析対象にすることが可能であるため、より詳細に分析することで生活習慣病をはじめとする傷病間の状態や関係が明らかに出来ると考える。

③ 死因単純分類による複合死因の分析

原死因と複合死因について隣接行列を作成した結果、原死因別に見た複合死因と、

複合死因間の隣接行列は、それぞれ異なる関係性が示された。中でも悪性新生物は原死因数でも、複合死因数でもその合計が一番多い死因であるが、部位別にみれば、より強い関係がみられるかもしれない。

近年は老衰による死亡が増加しているが、死亡診断書に記載されている死因が「老衰」の一言しかなければ、本来の死因が隠されている可能性も否定できないが、老衰としか書きようのない「真の」老衰死も確かに存在していると考えられる。真の老衰があることをきちんと示せるような、また老衰死の状況を適切に死因統計で把握できるように死亡診断書の記載方法を検討する必要があるのではないだろうか。

個別の死因に注目すれば、134の死因単純分類の中には、かなり死亡数の少ない死因がある一方で、ある程度死亡数が多いにもかかわらず「その他の～」とまとめられてしまい、その重要性が把握されていない死因も一定数ある。変化する死因構造に合わせて、増えていて対策が必要な疾病の動向を可視化するために、適切な単純死因分類の設定が求められよう。

④ 糖尿病の複合死因分析

糖尿病は、死因欄のI欄アに記載されるケースは少なく、過半数においてII欄に記載され、次いでI欄イに記載があった。ただし、II欄に記載された場合の原死因が糖尿病である割合は近年上昇しているものの2021年でも2割弱であり、多くの場合は原死因を用いた死因統計に現れない。しかしながら糖尿病は腎臓疾患や循環器系疾患を併発することが知られており、既存の情報を活用するという観点からも、複合死因を用いた分析は重要であると考えられる。

⑤ COVID-19の複合死因分析

COVID-19による死亡データはその多くが直接死因であるI欄アに記載されていた。その約半数にはII欄に何らかの疾病が記載

されており、中でも糖尿病は同時に慢性腎不全や高血圧、慢性心不全などが併記されていることが多かった。また死亡までの期間は平均 19.7 日であるが、I 欄アに限定すると 17.9 日、30 日以下が 92%にのぼることから、死亡者は感染から 2~3 週間で死亡するケースが多かったと考えられる。

⑥ 複合死因の人口学的分析手法

SRMU をわが国の複合死因データに適用し、2003~2021 年における年次推移を 15 死因による死因グループで観察すると、どの死因でも概ね SRMU は減少する傾向が観察された。ただし、ICD-10 の 2013 年版が適用された 2017 年についてはいくつかの死因で傾向に不連続性が発生しており、死因 5(Mental and behavioral disorders), 6(Diseases of the nervous system)では上昇が、死因 8(Diseases of the respiratory system)では減少が見られた。

CDAI(65 歳以上女性)の 2010~2021 年における年次推移について、原死因と複合死因が同一の分類となる場合を除き、CDAI が 800 以上となるのが 12 年の過半数である 7 年分以上である組み合わせを見ると、neoplasm(新生物)のグループに属する死因では、原死因・複合死因ともにこのグループ内部で強い関係性を多くの年次で示す組み合わせが多く、これに CDAI が 400 以上で過半数となる組み合わせまでを加えると、diseases of the respiratory system(呼吸器系の疾患)でも同様にグループ内部で強い関係性を多くの年次で示す組み合わせが多いことが観察された。一方で、endocrine, nutritional and metabolic diseases(内分泌、栄養及び代謝疾患)や mental and behavioral disorders(精神及び行動の障害)のように、他の原死因との複合死因として強い組み合わせで多く現れるグループも存在する。

他方で、特定のいくつかの年次だけ強い

関係が見られるが、必ずしも多くの年次で強く関係となっていない組み合わせも存在しており、このようなケースは、endocrine, nutritional and metabolic diseases(内分泌、栄養及び代謝疾患)、mental and behavioral disorders(精神及び行動の障害)、diseases of the respiratory system(呼吸器系の疾患)が複合死因となる場合に多く見られた。

このように、CDAI によって原死因と複合死因の間関係の強さが定量的に示された一方、その関係性は年次によっても変化するため、継続的な観察を行うことの重要性も明らかとなった。

ネットワーク中心性指標の年次推移を男女計・全年齢で観察すると、次数中心性で 2021 年に高い数値を示す続発性悪性新生物(24_SecNeo)、肥満(症)及びその他の過栄養<過剰摂食>(31_Obes)、ヒト免疫不全ウイルス[HIV]病(2_AIDS)、部位不明確、部位が明示されていない、独立した(原発性)多部位の悪性新生物(25_UnsNeo)の中で、24_SecNeo と 25_UnsNeo はこの期間内で比較的高い値を維持しているのに対して、31_Obes と 2_AIDS は年次による変動が大きいことが観察される。次に、隣接中心性で 2021 年に高い数値を示すヒト免疫不全ウイルス[HIV]病(2_AIDS)、その他の感染症及び寄生虫症(6_OInfec)、肥満(症)及びその他の過栄養<過剰摂食>(31_Obes)を見ると、2_AIDS と 6_OInfec は緩やかな上昇基調にある一方で、31_Obes は概ね横ばい傾向にあるという違いが観察された。一方、媒介中心性で 2021 年に高い数値を示すヒト免疫不全ウイルス[HIV]病(2_AIDS)、肥満(症)及びその他の過栄養<過剰摂食>(31_Obes)、続発性悪性新生物(24_SecNeo)を見ると、24_SecNeo は概ね 300 を中心に変動しながら推移している一方で、2_AIDS と 31_Obes は年次ごと

の変動が激しいことが観察された。

また、コミュニティ分類の結果について 2021 年で見ると、Leading eigenvector [cle]法では全体が 5 のコミュニティに分類されているのに対して、Label propagation[clp]法では 10 となっている。また、Fast-greedy[cfg]法による分類数の年次推移を観察すると、6~8 の値を推移してきており、分類されるコミュニティ数はアルゴリズムや年次によっても大きく異なることが観察された。

従来、原死因という単一の死因のみを用いて死因推移や死因構造を考えてきたが、生活習慣病等の広がりにもなって関連死因間の分析が必要となっている。本研究で示した SRMU や CDAI の指標や、ネットワーク分析の活用などによって、複合死因情報の利用促進が期待される。

E. 結論

本研究は、原死因を決定する以前の死亡個票データを用いることにより、死因間の関連を分析し、長寿化を進展する要因を死因構造から分析するものである。これらの情報は、従来は原死因を特定するためのみに用いられ、統計として利用されてこなかった。本研究では研究チームで独自にコード化したデータを用い、主要な死因別に詳細分析を行い、国際比較可能な複合死因指標を示し、ネットワーク分析を行うと同時に、死因簡単分類別に原死因と複合死因の関係行列データを作成・公表することで、広く一般に複合死因情報が利用されるような基盤を提供した。

死亡診断書に記載される死因数が減っている傾向は、老衰死亡が大きく増加していることが理由の一つであった。日本の老衰死亡の多さは国際的にも際立っているが、現状の死亡診断書のかきぶりからは、老衰死亡を何か別の病因に割り振ることは難し

い。死亡に至った病気を明らかにする、という現在の死因統計の在り方を見直す必要があると考えられる。

生活習慣病は多くの疾患を併発する可能性がある。その代表である糖尿病に対して複合死因分析を行ったところ、糖尿病はⅡ欄へ記載されることが多いために原死因に選択されにくく、かつ腎臓疾患などの死因と関連が深いことが示された。こうした死因間の関係は従来の原死因を用いた分析からは得ることができず、複合死因分析の有用性の一端が示された。

COVID-19に関する死亡の複合死因をみると、その多くが直接死因であるⅠ欄への記載であり、また他の死因が記載されることは少なかった。ただし、Ⅱ欄に糖尿病や慢性腎臓病の記載があるケースは 2020・2021 年ともにやや多く、この感染症の特徴の一つであることが示された。

また、本研究で行った、わが国の複合死因データへのネットワーク分析適用からは、複合死因間の関係を分析する上でのネットワーク分析の有効性が明らかになったといえる。

複合死因を分析するためのデータは、日本では死亡票はクリーニングされコード化されているが、死亡個票データは厚生労働省にオンライン登録された時点におけるテキストデータとなっており、データを入手しただけでは分析を行うことができない。しかし原死因以外の死因情報が得られるデータはこの死亡個票データ以外には存在せず、本研究が分析手法の一例を提示することによって未利用の死亡診断書データ、複合死因情報に関する利用・分析の促進が望まれる。

死亡個票データを用いて分析を行うにはデータのクリーニング・コード化を行う必要がある。本研究では、ICD-10 対応標準病名マスターで定義されている病名交換用

コードおよび ICD-10 コードに、期間欄は日数に変換する方法を提示している。今後、実際のデータを扱っていく中で方法の修正等も行い、精度向上・効率化を進める必要があるだろう。

また、近く死因分類が ICD-10 から ICD-11 へと変更される予定である。原死因は ICD が定める死因分類によって選択されるため、死因分類の変更のみならず選択方法の変更によっても前後で不連続となる。他方、死亡診断書のデータから分析を行う場合には死因選択方法の変更による影響を受けることがないことから、時系列比較を行う際には死因構造および死因間の関連変化について、より適切な情報を提供できる可能性がある。また、特徴的な複合死因構造や死亡までの期間等の情報を活用することにより、どのような疾病対策が必要となるか等といった臨床でのニーズに答えられる分析が必要であり、そのために複合死因情報を広く発信し、利用の拡大を図りたい。

複合死因研究は国際的にもまだ開発途上であり、また分析を行う上でデータに関するハードルは低くないものの、得られる情報は豊富である。本研究の進展が貢献できる余地は小さくないといえよう。本研究により、死亡診断書データのさらなる有効活用や人口動態統計の集計表の充実や分析の高度化など、将来的な公的統計に関する企画・立案に貢献できるものと考えている。

F. 研究発表

1. 論文発表

別府志海・篠原恵美子「新型コロナウイルス感染症による死亡動向と複合死因分析：2020年」『人口問題研究』, 第78巻第4号(2022), pp.477-492.

林玲子・別府志海・石井太・篠原恵美子「老衰死の統計分析」『人口問題研究』, 第78

巻第1号(2022), pp.1-18.

HAYASHI Reiko “COVID-19 and Mortality Decline in Asia in 2020”, 『人口問題研究』, 第78巻第4号(2022), pp.493-508.

Ishii, Futoshi, *Modeling Shifting Mortality, and Its Applications*, Springer, Singapore, 近日刊行.

石井太「日本版死亡データベースで用いる死因分類とその死亡分析への応用」『人口問題研究』, 第78巻第1号(2022), pp.32-55.

石井太・林玲子・篠原恵美子・別府志海「複合死因データに関する分析手法とその応用—わが国データへのネットワーク分析適用の試み—」『人口問題研究』, 第78巻第1号(2022), pp.56-77.

別府志海「III. 新型コロナウイルス感染拡大期における死亡・死因の状況」, 岩澤美帆、小池司朗、林玲子、別府志海、是川夕『新型コロナウイルス感染拡大と人口動態：何が分かり、何が起きるのか』国立社会保障・人口問題研究所 Working Paper Series, No.51 (2021), pp.12-19.

別府志海・石井太「第4章 わが国における震災と死亡」井上・和田編『自然災害と人口』, 原書房(2021), pp.75-98.

別府志海「第3章 日本の健康寿命」金子・石井編『長寿・健康の人口学』, 原書房(2021), pp.43-65.

林玲子・別府志海「新型コロナウイルス感染症流行に応じた主要国の死亡統計早期公開について」国立社会保障・人口問題研究所 Working Paper Series, No.50 (2021)

林玲子「II. 国際的・長期的視点からみた新型コロナウイルス感染症の人口への影響」, 岩澤美帆、小池司朗、林玲子、別府志海、是川夕『新型コロナウイルス感染拡大と人口動態：何が分かり、何が起きるのか』国立社会保障・人口問題研究所 Working Paper Series, No.51 (2021),

- pp.2-11.
- 林玲子「外国人の死因－日本人・本国人と
の比較」『人口問題研究』第 76 巻第 2 号
(2020), pp.218-239.
- 林玲子・別府志海・石井太・篠原恵美子「老
衰死の統計分析」『人口問題研究』第 78
巻第 1 号 (2022), pp.1-18.
- 石井太「第 2 章 死亡分析の方法 (生命表
と死亡モデル)」金子・石井編『長寿・健
康の人口学』, 原書房 (2021), pp.25-42.
- 石井太「第 9 章 わが国の寿命の将来」金
子・石井編『長寿・健康の人口学』, 原書
房 (2021), pp.171-189.
- 石井太・林玲子・篠原恵美子・別府志海「複
合死因データに関する分析手法とその応
用－わが国データへのネットワーク分
析適用の試み－」『人口問題研究』第 78
巻第 1 号 (2022), pp.56-77.
- Saito, Y., F. Ishii, J.-M. Robine, “Centenarians
and Supercentenarians in Japan,” Maier,
Heiner, B. Jeune, J. W. Vaupel (eds.)
Exceptional Lifespans, Springer, 2021,
pp.125-145.
- 篠原恵美子・別府志海・林玲子・石井太。「死
亡個票における『死亡の原因』欄の記載
文字列の分析」『医療情報学』 41(Suppl.),
2021, pp.733-7.
- Kawazoe Y, Shibata D, Shinohara E, Aramaki
E, Ohe K. “A clinical specific BERT
developed using a huge Japanese clinical
text corpus.” *PLoS One*;16(11):e0259763.
doi: 10.1371/journal.pone.0259763. PMID:
34752490. 2021 Nov 9.
2. 学会発表
- Ishii, Futoshi; Reiko Hayashi, Emiko
Shinohara, and Motomi Beppu
“Analysis of the Multiple Causes of
Death in Japan with Network
Analysis” Fifth meeting of the
MultiCause network, Federal Institute
of Drugs and Medical Devices (BfArM),
Germany and online, May 19, 2022.
- Hayashi, Reiko; Futoshi Ishii, Emiko
Shinohara, and Motomi Beppu
“Senility Deaths in Japan” Fifth
meeting of the MultiCause network,
Federal Institute of Drugs and Medical
Devices (BfArM), Germany and online,
May 20, 2022.
- 林玲子・別府志海・石井太・篠原恵美子「日
本における複合死因の概況と死因簡単分
類分析」日本健康学会第 87 回総会,
東京医療保健大学, 2022 年 11 月 4 日.
別府志海「死亡・死因と新型コロナ」日本
人口学会第 73 回大会報告, 2021 年 6 月
- 林玲子・別府志海・石井太・篠原恵美子「日
本における複合死因の分析」日本人口学
会第 73 回大会報告, 2021 年 6 月.
- 林玲子「人口高齢化と新型コロナ感染症死
亡統計－年齢構造調整の必要性－」第 85
回日本健康学会総会ポスター報告, 2020
年 12 月.
- 林玲子「老衰死亡の統計分析」第 73 回日
本老年医学会関東甲信越地方会, 2021
年 11 月.
- Ishii, Futoshi, Reiko Hayashi, Emiko
Shinohara, Motomi Beppu
“Application of Network Analysis to
Multiple Causes of Death Data in
Japan,” The 29th International
Population Conference (IPC2021),
International Union for the Scientific
Study of Population (IUSSP), Dec. 9,
2021.
- 石井太・別府志海・菅桂太「日本版死亡デ
ータベースの地域分析・死因分析への拡
張・応用」日本人口学会第 73 回大会報
告, 2021 年 6 月.
- Ishii, Futoshi “Decomposition Analysis

by Cause of Death for the Modal Age at Death Using the Linear Difference Model,” Poster presented at 2021 Annual Meeting, Population Association of America, May 6, 2021.

Futoshi Ishii “Japanese Longevity from Population Projection Perspective: Background, Prospects and Impacts”, International Actuarial Association, Council and Committee Meetings (Members Forum), Westin Tokyo, Nov. 21, 2019 (Invited speech).

篠原恵美子・別府志海・林玲子・石井太. 「死亡個票における『死亡の原因』欄の記載文字列の分析」第41回医療情報学連合大会, 2021年11月20日.

G. 知的財産権の出願・登録状況

なし