

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。

また、複合死因の集計・分析に関しては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）がある。ここの研究交流を行っていくとともに、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や複合死因・疾病と原死因の関係について分析を行う。

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

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。

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

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、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の二次利用し分析することとした。

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となっており、データのクリーニング等はされていない。そのため、分析に先立ってデータのクリーニングおよびコード化を行うことが必須となる。本研究では、記載された各死因名を標準病名に変換した上で ICD-10 (2013 年版) ベースに変換・コード化し、これを用いて分析を行うこととした。なお、具体的なコード化の方法等については、分担報告書 (篠原担当)

を参照頂きたい。

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。

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った。提供されたデータのうち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いる。死亡個票データのうち、各死因欄に記載された死因名は 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を行った。データ量が非常に大きく、素朴な実装では実行時間が非常に長くなるため、形態素解析器の MeCab を利用し効率化を図った。

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。また死因欄別の記載状況は、特に直接死因である I 欄アでは空欄の件数が極めて低いことから、多くの死亡において少なくとも一つの死因は分かる状況といえる。

死亡票とマッチングさせた死亡個票における各死因欄のデータを見ると、死因欄に

は複数の死因が記載されているケースも見られた。これら記載されたすべての死因についてコード化を行ったが、1件以上コード化がされたものに占める2件以上の記載がある死亡数は、I欄アでは2~4%と低く、I欄イ~ウは5%前後、I欄エは10%前後であったほか、II欄では同割合が25%ほどであった。また死因欄に関係なく、コード化された死因が2つ以上ある死亡の割合を求めると、全年次とも47~56%であり、全体の約半数について何らかの複合死因分析が可能である。

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、間接死因となる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、高血圧症などと言った死因の記載が多くなっており、いわゆる生活習慣病が多く記載されている。

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。

本年度に計画していた「海外における複合死因ネットワークとの情報交換」は、世界的なコロナウィルスの蔓延にともない困難であったことから、これに代えて文献レビューを行った。複合死因に関する分析方法等について、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おり、これを参考としつつ、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った。レビューの結果、近年ではネットワーク分析を応用した複合死因間の関係分析研究が始められていることから、ネットワーク分析の概要についてもレビューを行った。

なお、本年度には複合死因に関連する国際会議等は開催されなかったが、これまでの研究成果に基づき、複合死因に関するネットワーク分析および糖尿病に関する報告を日本人口学会大会において行うことにより、国内研究者と複合死因の分析手法等について意見交換・討論を行った。

D. 考察

死因のコード化には病名マスターを利用しているが、カバーされない病名表記は独自の変換表を用意して対応した。この結果、全ての年のデータについて、95%以上の死亡個票についてI欄に少なくとも1つのICD-10コードが付与された。しかしこのマスターと変換表でもまだカバーできていない範囲があり、今後改善が必要である。

また、期間正規化の結果がマイナスの数値になっているケースが見られた。この原因としては死亡個票の内容の誤りとプログラムの誤りの可能性があり、調査が必要である。前者に対してはそのようなケースをどう扱うかの検討が、後者に対してはプログラムの修正を行う予定である。

さらに、原因と期間の両方について、入力文字数に制限があるため、備考欄に続きが記載されているケースの存在が確認されている。このようなケースの割合の調査、対応する原因・期間と接続し再構成する方法の確立が必要である。

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。

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、例えば悪性新生物と肝疾患の関係数は14,216にのぼっており、肝がんと肝疾患の関係をみれば、より強い関係がみられるかもしれない、部位別に見ていく必要があるだろう。原死因、複合死因それぞれの関係性を示すような適切な指標、分析枠組みを複数検討・考慮する必要がある。

複合死因に関する先行研究の内容・手法を類型化すると、大きく以下の7類型に分けることができる。

[1] 死因について原死因に加え複合死因も含めた分析 (51)

[2] 原死因による分析と複合死因による分析を比較対照するもの (29)

[3] 1死亡当たりの死因数の分析 (11)

[4] 複合死因間の関係を分析するもの (24)

[5] 競合リスクモデルや特定死因を除去した生命表など生命表分析への応用 (7)

[6] 死因選択の妥当性の検証など (5)

[7] その他 (10)

※ 括弧内は当該領域を扱った論文数。複数領域を扱った論文もあるため合計は

113を超える。

特に分析方法について、近年では死因間の関係についてネットワーク分析手法を応用した研究が行われており、各死因分類をネットワークにおけるノード(点)と考え、同一の死亡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(線)で結ぶ無向ネットワークを考えて分析を行う先行研究などが存在している。ネットワーク分析では、ネットワークに関する特性を表す指標を観察する方法に加え、ネットワークのグラフをいくつかのサブグループ(コミュニティ)に分けるコミュニティ検出という手法が存在する。これを複合死因分析に応用すると、原死因だけの観察からはわからない、互いに結びつきの強い死因のグループを抽出することが可能となる。これは、長寿化が進行し、複数の病態を同時に併発することが多いわが国の現在の死亡状況の分析に有効であると考えられることから、本研究ではこのコミュニティ検出手法に関してもレビューも行った。

このように複合死因研究は国際的にもまだ開発途上にあるといえ、本研究から得られる知見が貢献できる余地は小さくないといえよう。

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。しかしながら、複合死因が収録されている現状のデータは集計・分析を行うことが想定されているとはいいがたい。とは言え原死因以外の死因情報が得られるデータはこれ以外にはなく、本研究が分析手法の一例を提示することによって複合死因情報の利用促進が期待される。

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作成されるため、死因分類変更の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分類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。

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にのみ用いられていた。

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を試みた。入力されているデータが複雑であるが、これらのデータを効率的に活用できるような方法をさらに検討したい。

死因には、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あり、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。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る。

現段階では、死亡票の死因欄の順番、また、I欄とII欄との違いについては考慮していない。因果関係を同定できるのかどうかを含め、今後検討する必要がある。また記入テキストがICDコードに変換できないケース、外因の複合死因、経年的に複合死因の関連がどう変化したかの分析も今後の検討課題である。

また先行研究のレビューから、複合死因研究は国際的にも注目されており、死因間の関連の指標化やネットワーク分析といった手法の導入など新たな試みが進められていることが示された。

複合死因研究は国際的にもまだ開発途上であり、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。本研究を進めることにより、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られる。

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

石井太・林玲子・篠原恵美子・別府志海「複合死因間関連分析へのネットワーク応」
日本人口学会第72回大会、オンライン

開催、2020年11月15日

別府志海・石井太・林玲子・篠原恵美子「複
合死因データを用いた糖尿病と関連死因
の人口学的分析」日本人口学会第72回
大会、オンライン開催、2020年11月15
日

H. 知的財産権の出願・登録状況

なし