

## 特集 I : 長寿革命に係る人口学的観点からの総合的研究

## 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—

石井太\*・林玲子・篠原恵美子<sup>†</sup>・別府志海

本研究は、近年、わが国においても利用可能性が広がってきた複合死因データに関する分析手法について概観するとともに、特に、複合死因間の関係分析にネットワーク分析を応用した Egidi et al. (2018) の方法論を日本のデータに適用することを試み、複合死因分析へのネットワーク分析の応用の可能性を検討することを目的とする。

海外では国際的な複合死因に関する研究ネットワークである MultiCause Network において、これまで様々な研究が蓄積されていることなどから、複合死因データに関する分析について多くの先行研究が存在している。それらの先行研究については、[1] 死因について原死因だけでなく複合死因も含めて分析を行ったもの、[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、[3] 1死亡当たりの死因数を分析するもの、[4] 複合死因間の関係を分析するもの、[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を行ったもの、[6] 死因コーディングの妥当性の検証などに応用するものなどの類型にまとめることが可能であるが、近年、この類型 [4] である、複合死因間の関係分析にネットワーク分析を応用した研究が行われている。その中でも、Egidi et al. (2018) は、イタリアの複合死因データにネットワーク分析の手法を適用して死因間の関連分析を行った先駆的な研究であり、本研究ではこの方法論をわが国の複合死因データに適用することを試みた。分析では、構築されたネットワーク構造に対して、ネットワーク中心性を測定する指標を用いて、他の死因と関係性の深い死因を抽出するとともに、Egidi et al. (2018) では行われていない、ネットワーク分析で用いられるコミュニティ検出アルゴリズムの適用を通じて、死因間の結びつきが強いコミュニティへの分類を行った。

本研究で行った複合死因データに関する分析手法の概観からは、複合死因データが有する豊富な可能性が理解されたとともに、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よう。

キーワード: 複合死因, ネットワーク分析, 死亡分析

## はじめに

現在、人口動態統計では、死亡票に記載されている複数の死因から、世界保健機関が勧告する「疾病及び関連保健問題の国際統計分類」に準拠し、直接に死亡を引き起こした一連の事象の起因となった疾病もしくは損傷等を表す単一の「原死因」を用いて死因統計の

\* 慶應義塾大学

<sup>†</sup> 東京大学

集計・分析を行っている。これは、一連の病的事象を起こす原因を防止するという公衆衛生的な観点に基づくものであるが、一方で、現在、わが国では生活習慣病が死因の上位を占めるに至り、一人が複数の疾患を抱えることも多くなってきていることから、原死因以外の死因に着目する必要性が高まっている。

このような問題意識の下、社会保障審議会統計分科会疾病、傷害及び死因分類部会は、平成26年11月に出した報告（「疾病、傷害及び死因に関する分類に係る部会審議の際に出された意見に基づく報告」）の中で、「中長期的には、基礎疾患の情報や介入の状況、合併症、予後等、死亡診断書・死体検案書から得られる複合的な要因を把握できるような分析がなされることが望ましい。」との方向性を打ち出している。しかしながら、人口動態統計死亡票の原死因以外の複合死因情報については、近年、はじめて二次利用が可能となったところであり、わが国ではこのような複合死因データを全人口ベースで取り扱った経験が多いとはいえない状況にあった。一方、諸外国においては、従来から複合死因のデータの活用事例が存在しており、例えば人口学領域においては、原死因と複合死因の関係性を記述するための人口学的指標の構築などの先行研究が行われてきており、特に、国際的な複合死因に関する研究ネットワークである MultiCause Network において様々な研究が蓄積されている。この中で、Egidi et al. (2018) は、イタリアの2011年の65歳以上の複合死因データを用いて、欧州死因分類をノード、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、これにネットワーク分析の手法を適用して死因間の関連分析を行った先駆的な研究である。

本研究は、近年、わが国においても利用可能性が広がってきた複合死因データに関する分析手法について概観するとともに、特に、複合死因間の関係分析にネットワーク分析を応用した Egidi et al. (2018) の方法論を日本のデータに適用することを試み、複合死因分析へのネットワーク分析の応用の可能性を検討することを目的とする。

## 1 先行研究

本節では複合死因データに関する分析手法に関する先行研究について概観する。先述の通り、海外では、国際的な複合死因に関する研究ネットワークである MultiCause Network において、特に様々な研究が蓄積されていることを踏まえ、まず1.1節において海外の先行研究について述べ、その後、1.2節でわが国における先行研究について述べる。また、本研究で適用を試みるネットワーク分析の応用については1.3節において述べることにする。

### 1.1 複合死因データ分析に関する海外の先行研究

ここではまず、海外における先行研究を概観する。複合死因データを用いた分析に関する先行研究の分析内容や手法は、必ずしも排他的ではないが、下記に示すようないくつかの類型にまとめることが可能であろう。

- [1] 死因について原死因だけでなく複合死因も含めて分析を行ったもの
- [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
- [3] 1死亡当たりの死因数を分析するもの
- [4] 複合死因間の関係を分析するもの
- [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を行ったもの
- [6] 死因コーディングの妥当性の検証などに応用するもの
- [7] その他

それぞれの類型ごとに研究の特徴を概観すると以下の通りである。

[1] は原死因による集計では多く計上されないものについてその死因の潜在的な貢献を示すため、また、[2] はさらにそのような原死因と全ての複合死因を含めた場合の比較を行う観点から分析が行われている。これらについては、方法論としては記述統計を用いて、単純集計やクロス集計を行った研究、さらに、年齢階級別死亡率や年齢調整死亡率を算定・比較した研究などが多く存在している。例えば、Dushoff et al. (2005) では、インフルエンザが死亡に与える影響を評価するため複合死因データを用いて分析を行っているが、インフルエンザが原死因となる場合と、複合死因にも挙げられている場合の両者を比較し、インフルエンザの潜在的な大きさを示した上で、複合死因で計上した死亡数を用いた回帰分析を行っている。また、糖尿病に関連する死亡を分析したものとして、Balkau and Papoz (1992) や Barreto et al. (2007) による分析が挙げられる。Balkau and Papoz (1992) は、フランスにおける糖尿病関連死亡のレベルを評価したものであり、原死因・複合死因のいずれかに糖尿病が出現する数を用いた SMR を求め、原死因のみによる SMR より高いことを指摘している。また、Barreto et al. (2007) はブラジルにおける糖尿病関連死亡の属性別状況を分析したものであり、糖尿病が原死因に出現する場合に対する、原死因以外の複合死因に出現するオッズ比を多重ロジスティック回帰により属性別に推定して分析を行った。

このように、複合死因を用いて明らかとなる重要な特性として、原死因のみの観察ではその影響が過小評価されてしまう死因の適切な評価があり、インフルエンザ、糖尿病、アルツハイマーなど様々な死因に着目した分析が行われている。

[3] と [4] は原死因では得られない、複合死因ならではの特徴を分析する研究といえる。White et al. (1989), Wilkins et al. (1997) はそれぞれアメリカ、カナダについて、一つの診断書あたりの複合死因数の分布やその平均値などの分析を行ったものである。また、Gorina and Lentzner (2008) は一診断書あたりの複合死因数の分布や原死因・複合死因の比較などに加え、死因間の関係を示す指標として、以下を提案している。

$$\frac{\text{Observed numbers of deaths with both causes A\&B}}{\text{Expected numbers of deaths with both causes A\&B}}$$

ここで、”Observed numbers of deaths with both causes A&B” とは実際の複合死因統計から得られる死亡数であり、”Expected numbers of deaths with both causes A&B” は死因間の独立性を仮定して、以下により得られた数である。

$$\frac{\text{Deaths with cause A} \times \text{Deaths with cause B}}{\text{Total deaths (all causes)}}$$

この指標が1よりも大きいほど、二つの死因が独立であると仮定した期待出現数よりも、実際の観察数が高い頻度であることを示すことから、両者の結びつきが強いことになる。Gorina and Lentzner (2008) ではこの指標を用いて、糖尿病との関係性が強い死因に関する分析を行っている。

一原死因あたりの死亡数や死因間の関係を示す指標は、複合死因の特徴を記述統計によって表現する上で優れた指標であると考えられるが、一方で、複数時点や異なる人口の間での比較を行う場合、年齢構成の違いによって適切な比較を行えない場合がある。そこで、MultiCause Network グループのメンバーによって、Désesquelles et al. (2012), Désesquelles et al. (2010) などの先行研究を中心に、より洗練された人口学的指標が提案されており、これが SRMU (Standardized Ratio of Multiple to Underlying Cause) と CDAI (Cause of Death Association Indicator) である。

SRMU は、原死因死亡に対する複合死因死亡の比を表すもので、以下で定義される。

$$SRMU_u = \frac{\sum d_{u,x,s}^{[M]}}{\sum d_{u,x,s}^{[U]}}$$

$d_{u,x,s}^{[M]}$  : 年齢  $x$ , 性別  $s$  の複合死因死亡数 (原死因  $u$ )

$d_{u,x,s}^{[U]}$  : 年齢  $x$ , 性別  $s$  の原死因死亡数 (原死因  $u$ )

この指標は死因分類に依存することから、複数時点や異なる人口間の SRMU を比較する場合には使用する死因分類を統一しておくことが必要である。また、 $d_{u,x,s}$  については、複合死因について、同じ分類のものについても重複して計上する方法 (重複計上) と、同じ分類については一回のみ計上する方法 (重複調整) の2種類が考えられる。また、さらに、死亡数に関しては、年齢調整を行わない場合と年齢調整を行う場合の2種類の指標が考えられるが、やはり複数時点や異なる人口間での比較については、年齢調整を行うことが望ましい。年齢調整を行う場合には、

$$d_{u,x,s}^{<S>} = d_{u,x,s} \frac{E_{x,s}^{<S>}}{E_{x,s}}$$

とする。ここで、

$d_{u,x,s}$  : 年齢調整前の死亡数

$d_{u,x,s}^{<S>}$  : 年齢調整した死亡数

$E_{x,s}$  : 対象とする人口の Exposure  
 $E_{x,s}^{<S>}$  : 基準となる人口の Exposure  
 である。

なお, Désesquelles et al. (2012), Désesquelles et al. (2010) では, 分類については重複調整を行い, 年齢調整を行った比を推定している。

CDAI は, 死因  $c$  の標準化出現率 ( $SP_c$ ) 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 ( $SP_{c/u}$ ) の比率であり,

$$CDAI_{u,c} = \frac{SP_{c/u}}{SP_c} = \frac{\sum_{x,s} \left( \frac{{}_u d_{c,x,s}}{{}_u d_{x,s}} \cdot \bar{d}_{x,s} \right) / \sum_{x,s} \bar{d}_{x,s}}{\sum_{x,s} \left( \frac{d_{c,x,s}}{d_{x,s}} \cdot \bar{d}_{x,s} \right) / \sum_{x,s} \bar{d}_{x,s}} \times 100 = \frac{\sum_{x,s} \frac{{}_u d_{c,x,s}}{{}_u d_{x,s}} \cdot \bar{d}_{x,s}}{\sum_{x,s} \frac{d_{c,x,s}}{d_{x,s}} \cdot \bar{d}_{x,s}} \times 100$$

${}_u d_{c,x,s}$  : 原死因  $u$  で, 死因  $c$  を含む  $x$  歳, 性別  $s$  の死亡数

${}_u d_{x,s}$  : 原死因  $u$  の  $x$  歳, 性別  $s$  の死亡数

$d_{c,x,s}$  : 死因  $c$  を含む  $x$  歳, 性別  $s$  の死亡数 (原死因によらない)

$d_{x,s}$  :  $x$  歳, 性別  $s$  の死亡数 (原死因によらない)

$\bar{d}_{x,s}$  : 基準時点における  $x$  歳, 性別  $s$  の標準的死亡数 (死亡率と Exposure の積)

これは, 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であり, CDAI が100を上回るほど原死因  $u$  との関連性が高いと考えることができる。Gorina and Lentzner (2008) による死因間の関係性指標を標準化したものと捉えることができよう。

[5] は, 競合リスクモデルや特定死因を除去した生命表分析への応用などを研究したものであり, 1969年アメリカの複合死因データを用い, 原死因を除去した生命表と複合死因を除去した生命表を作成して比較を行った Manton and Stallard (1982) や, 原死因以外の複合死因を考慮したウェイトを持つ状態への移行を表現できる生存時間分析の拡張モデルを提案し, これによる Cox 回帰モデル分析のシミュレーションを行った Moreno-Betancur et al. (2017) などが挙げられる。

[6] の類型は [1]~[5] の類型とはやや異なり, 原死因の選択の妥当性について, 複合死因データから検証などを行うことを目的としたものであり, イギリスの心疾患死亡による400の死亡診断書・死体検案書を用いて臨床医パネルに原死因選択を依頼し, 半数近くでの不一致を観察したという, 原死因選択の信頼性の限界について論じた Mant et al. (2006) などが挙げられる。

## 1.2 わが国における複合死因データ分析に関する先行研究

次に, わが国における, 人口動態調査の死亡票 (もしくは死亡診断書) を用いた複合死



因データ分析に関する先行研究について述べる。わが国においては、1950年代から上田による複合死因に関する研究がなされている。上田（1956）は、第6回国際死因統計分類修正会議の複合死因及び製表に関する勧告に基づいてWHOセンターに寄せられた6つの研究報告（WHOセンター、アメリカ、アイルランド、ドイツ、フランス、イタリア）の概要についてまとめたものであり、上田・加藤（1957）は、試験的調査として昭和30年の東京都・長野県の死亡票のうち、心臓疾患が記載されているもの（東京都4,355件、長野県1,973件）を対象として、複合死因の記載や、原死因と二次死因の関係について分析を行った。また、上田（1960）は昭和30年の東京都の死亡票（44,255件）を対象とし、手術解剖施行有無別、性別、年齢別の複合割合、原死因と二次死因（合併症・副因）の関係分析として、死因別二次死因頻度、原死因に対する二次死因割合、合併症の二次死因に対する割合、原死因と二次死因のクロス表分析などを詳細に行い、さらにこれらを諸外国の結果と比較して考察を行っている。

佐々木（1975）は、昭和42～46年の大阪府の人口動態死亡票から糖尿病の記載があるものを対象として（原死因1,893件、二次死因1,389件）、糖尿病を原死因とする二次死因、糖尿病を二次死因とする原死因に関する分析を行い、これを1955年の米国公衆衛生局の全国調査、1968～69年の米国ペンシルベニアの調査と比較して糖尿病の死因の日米間の相違に関する考察を行っている。方波見〔等〕（1982）は昭和52年人口動態調査死亡票から原死因・二次死因に糖尿病の記載のあるもの（原死因9,023件、原死因以外9,148件）について、原死因と二次死因に関する年齢別・都道府県別糖尿病数や、糖尿病死亡率に関する分析を行い、また、方波見〔等〕（1983）は同じ対象に対して、糖尿病死亡の直接死因、二次死因糖尿病の原死因、原死因糖尿病と二次死因糖尿病に記載された二次死因の相違、二次死因糖尿病、全死因中に占める原死因各疾患の割合に関する分析を行っている。

梅田（1988）は昭和60年8月の11都県における人口動態調査の死亡票を対象として分析を行ったものであり、高血圧性疾患や糖尿病は原死因として他の死因を伴いやすく、他の死因の二次死因となる場合が多いこと、いずれの原死因においても二次死因として心疾患が多かったことなどを明らかにしている。

鈴木まき〔等〕（2001）は、運輸業を主体とする労働組合の加入者本人（男性）の1988～1995年の死亡診断書から、外因誌を除いた20～59歳の1,138件を対象とし、死亡診断書記載パターン別死亡数、主要な原死因に関連する主要死因、年齢階級別・原死因及び原死因に対応する死因における関連の強固性について分析を行い、糖尿病、高血圧、腎炎・ネフローゼでは原死因として表出がされにくいことから、複合死因の観点からの解析の必要性を指摘している。

一方、近年、国立社会保障・人口問題研究所においても複合死因の人口学的分析が進められてきた。その一連の研究は、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）「死亡個票統計における循環器疾患関連死因の妥当性に関する検討（H27-統計-一般-006）」（研究代表者：橋本英樹）、「人口動態統計死亡票の複合死因情報を活用した集計・分析方法に関する調査研究（H29-統計-一般-001）」（研究代表者：石井太）に

よって始められ、これがその後、国立社会保障・人口問題研究所の「長寿革命に係る人口学的観点からの総合的研究」, 「超長寿社会における人口・経済・社会のモデリングと総合分析」における本格的な複合死因分析につながっている。その中では1.1節において述べた SRMU や CDAI などの指標をわが国のデータを用いて算出し、死因間の関連状況について研究した石井 [等] (2019), 死亡期間に注目した突然死に関する分析と認知症関連死の分析について分析を行った林 [等] (2019), 糖尿病に焦点を当て、原因因を特定する前の情報である死亡個票を用いて疾病間、死因間の関連を示した別府 [等] (2019) 及び別府 [等] (2020) などの研究成果が報告書に記載されている。

さらに、現在、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）JPMH20AB1001「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）においてさらなる研究が継続されているところである。

### 1.3 複合死因分析へのネットワーク分析の応用に関する先行研究

近年、1.1節の類型 [4] である複合死因間の関係分析にネットワーク分析を応用した研究が行われている。これは、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研究である。Egidi et al. (2018) は、イタリアの2011年の65歳以上の死亡診断書に記述のある死因（原因因と関連死因を区別しない）に対して、欧州死因分類をノード、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、これにネットワーク分析の手法を適用して死因間の関連分析を行った先駆的な研究であり、以下、この内容についてレビューする。

最初に、死因  $i$  と  $j$  が同じ死亡診断書に書かれているとき、行列の  $(i, j)$  成分に 1 を計上していくことによって粗隣接行列を作成する。この行列は周辺度数である死因の出現頻度に依存していることから、各死因の出現頻度が同一となるよう、反復比例フィッティング (iterative proportional fitting (IPF)) 法を用いて正規化を行う。さらに、要素が小さい部分についての影響を排除する観点から、正規化された頻度について、最も強い関係があるものとして95%タイルよりも大きいものだけを選択し、その成分が表す二つの死因の間に関係があると考えて隣接行列を作成し、これに基づくネットワーク構造を考える。

このネットワーク構造に対して、ネットワーク分析の手法を適用しつつ、分析が行われる。最初に、ネットワーク中心性 (centrality) を測定する指標である次数中心性 (degree centrality), 隣接中心性 (closeness centrality), 媒介中心性 (betweenness centrality) を示し、これらの指標に関する男女間比較を行っている。ここで、ネットワーク中心性を示す各指標は以下で定義される。

- 次数中心性  $C_D(n_i)$ : ノード  $n_i$  と他のノードを結ぶリンクの数で定義される。  $n_i$  の

次数  $d(n_i)$  と表し、ノード  $n_i$  と  $n_j$  の間に関係があるとき  $x_{ij} = 1$ 、関係がないとき  $x_{ij} = 0$  という変数を用いると、

$$C_D(n_i) = d(n_i) = \sum_j x_{ij} = \sum_j x_{ji}$$

と表される。

- 隣接中心性  $C_C(n_i)$ ：他のノードとの距離が小さいノードほどより中心的であると  
する指標であり、 $d(n_i; n_j)$  をノード  $n_i$  と  $n_j$  の距離（最短パスの長さ）として、

$$C_C(n_i) = \left[ \sum_{j=1}^N d(n_i; n_j) \right]^{-1}$$

で定義される。ただし、 $N$  は  $n_i$  が属する連結成分におけるノードの数である。

- 媒介中心性  $C_B(n_i)$ ：あるノードが隣接していないノードを結びつける役割の大きさを表す指標であり、 $g_{jk}(n_i)$  をノード  $n_j$  と  $n_k$  を結び  $n_i$  を含む最短パスの長さ、 $g_{jk}$  をノード  $n_j$  と  $n_k$  を結ぶ最短パスの長さとして、

$$C_B(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk}$$

で定義される。

Egidi et al. (2018) ではさらに、密度 (density) や平均次数などのネットワークの構造係数の男女間比較、また、ネットワーク構造をグラフにより視覚化し、ICD-10の分類と対応付けながら、ネットワークをサブグループに分ける検討を行っている。そして、これらのネットワーク指標やコミュニティ検出などから得られた結果を既存研究等と比較し、死因間の関係性に関する議論を行っている。

なお、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などの方法に加え、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している。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。

まず、矢久保 (2013) に基づき、コミュニティ構造の定義とモジュラリティという指標、また、モジュラリティを利用したコミュニティ検出法について述べる。あるネットワークがその内部で互いに密に繋がり合った部分グラフに分けられるとき、このネットワークはコミュニティ構造をとるといい、その部分分割グラフのことをコミュニティまたはモジュールと呼ぶ。しかしながら、普遍的に受け入れられているコミュニティの定義は未だに存在せず、与えられた問題に応じていくつもの定義が提案されているとされ、その中で、現在、最も広く受け入れられている方法がモジュラリティという量に基づいたものとされている。これは、ノード数  $N$  のネットワーク  $G$  を、互いに重なり合いのない  $n_c$  個の部分ノード集



合族（コミュニティ） $C_l$  ( $l = 1, 2, \dots, n_c$ ) に分割した際、

$$Q = \frac{1}{2M} \sum_{i,j}^N (a_{ij} - \langle t_{ij} \rangle) \delta[C(i), C(j)]$$

で定義される。ここで、 $M$  は  $G$  の全エッジ数、 $C(i)$  はノード  $i$  が属する分割ノード集合、 $\delta[C(i), C(j)]$  は分割集合  $C(i)$  と  $C(j)$  が同一集合であれば 1、そうでなければ 0 となる量である。また、 $t_{ij}$  は  $G$  の各ノード数の次数を保持したまますべてのエッジをランダムに繋ぎ直したグラフ  $G'$  の隣接行列要素であり、 $\langle \dots \rangle$  はエッジ配置に関する平均を表す。すなわち、 $\langle t_{ij} \rangle$  は、 $G'$  における  $(i, j)$  間のエッジ本数の期待値を表す。  $Q$  の定義から、 $Q$  が大きな値をとる時、ネットワーク  $G$  は分割  $\{C_1, C_2, \dots, C_{n_c}\}$  をモジュールとするコミュニティ構造をとっていると考えることができる。

矢久保（2013）では、モジュラリティの最大値  $Q_{\max}$  を与える分割集合族をコミュニティとする定義が最も広く受け入れられているものの、ノード数が大きくなると分割パターンの組み合わせが非常に大きくなることから、全てのパターンに対してモジュラリティを計算することは事実上不可能となるとされている。このため、 $Q_{\max}$  の近似値を与える分割集合族を見つける方法が多数提案されている。

その一つが、Newman（2004）による貪欲法（fast-greedy）と呼ばれる方法である。これは、矢久保（2013）によれば、ネットワーク  $G$  の各ノード一つ一つを分割ノード集合と考え、 $G$  のエッジを介して互いに隣接する分割集合対  $\{C_i, C_{i'}\}$  を合併させた場合のモジュラリティ  $Q'$  を計算し、合併させる前からの増加分  $\Delta Q = Q' - Q$  が最大となる分割集合対を実際に合併させることをネットワークが一つになるまで繰り返し、その過程で現れる最も大きい  $Q$  を  $Q_{\max}$  の近似値とする方法である。なお、この方法は Clauset et al.（2004）により、さらに計算量を削減できることが示されている。

貪欲法とは逆に、 $n_c = 1$  の分割から  $n_c = N$  の分割に向かう流れの中で  $Q$  の最大値を探す方法もあり、この場合、分割ノード集合を次々と二分していくことになるが、このためには二分された集合ができるだけコミュニティとしての性質を持つようにする必要が生じる。Newman and Girvan（2004）は媒介中心性を用いてこの分割を行う方法を提案したものである。具体的には、与えられた連結ネットワーク  $G$  の全てのエッジの媒介中心性を計算して最大となるエッジを削除し、削除されたネットワークの全てのエッジの媒介中心性を再計算し、また最大のエッジを削除する、という操作を繰り返し、その過程で現れるネットワークのモジュラリティを求めてその最大値を  $Q_{\max}$  の近似値とする方法である。ただし、この方法は貪欲法に比べて多くの計算量が必要となるとされている。

また、Barabási（2019）によれば、モジュラリティを最適化するアルゴリズムとして、最近、Blondel et al.（2008）らによる Louvain のアルゴリズムがよく使われるとしている。これは、 $N$  個のノードが別々のコミュニティに属しているとする初期状態から出発し、ステップ I として、個々のノード  $i$  をその近接ノード  $j$  の属するコミュニティに加えてみたときのモジュラリティの増加量を計算する。このモジュラリティ増加量が正の値と

なるもののうち、最大となる近接ノードコミュニティにノード $i$ を置き直す。もし、どのモジュラリティ増加量も正にならなければ、ノード $i$ の属するコミュニティは変化しない。このプロセスをどのノードの属するコミュニティも変化しなくなるまで全てのノードについて繰り返す。以上がステップIとなる。次に、ステップIIとして、各ノードがステップIで検出されたコミュニティをノードとする新しいネットワークを構成する。このステップIとIIを繰り返すことによりコミュニティは減っていくが、モジュラリティの増加がなくなり、最大モジュラリティが得られたところで終了する。

なお、ネットワーク分析においては、以上の他にも様々なコミュニティ検出法が提案されており、Luke (2015) では、コミュニティ検出アルゴリズムとして、先述の貪欲法 (fast-greedy)、Louvain のほか、Walktrap, Edge-betweenness, Spinglass, Label propagation, Leading eigenvector などのアルゴリズムが採り上げられている。

## 2 データと方法

次に、1.3節で述べた Egidi et al. (2018) の方法を基礎として、わが国の複合死因データにネットワーク分析の適用を試みる。本節ではこれに用いるデータと方法について述べる。

### 2.1 データ

本研究では、複合死因データとして、厚生労働省『人口動態統計』における調査票の死亡個票データを用いる<sup>1)</sup>。なお、死亡個票ではデータがコード化されておらず、死因名がそのまま記載されている。そこで本研究では記載された死因名を ICD-10 ベースに変換し、分析に用いた。死因のコード化についての詳細は篠原 [等] (2021) を参照されたい。本稿では、後に述べる MultiCause Network グループのガイドラインを考慮し、2009年のデータについて分析を行った。

2009年の日本における日本人の死亡総数は1,141,865件であるが、このうち、複合死因を含んでいる死亡個票とのマッチングが可能だったのは853,412件である。本研究のネットワーク分析においては、年齢を65歳以上に限定した分析も行うことから、この中で、年齢不詳がなく、かつ、分類不能等に属さない複合死因が2つ以上ある380,372件を分析対象とした。一方、ネットワーク分析によるコミュニティ検出結果については、死因関連指標 CDAI との比較を行うが、この CDAI の算出には、複合死因とのマッチングが可能だった853,412件から、複合死因全不詳や分類不能等を除いた782,590件を対象とした。

---

1) 厚生労働省『人口動態調査』における死亡票は、前述の死亡届および死亡診断書等をもとに作成される。この調査票情報（死亡個票）を厚生労働省へ報告するシステムは2003年以降、順次オンライン化が進められた。本研究では死亡票情報に加え、オンラインにより提出された死亡個票情報を統計法第32条の規定に基づき二次利用している。また、死亡個票情報の独自集計であるため、公表数値とは一致しない場合がある。

## 2.2 方法

本研究では、MultiCause Network グループが複合死因分析に係る標準的指標の作成方法に関してまとめたガイドラインに従って、死因を取り扱う。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、以下の条件に従って用意することを求めている。

- 年齢範囲については制限をしない。
- 地理的範囲については特別な死亡パターンをもつ地域については対象外とすることもできる。
- 年次については2009年について全ての指標を計算する。ただし、2009年が利用可能でない場合、最も近い年次を算出する。
- 死因について、星印(\*)のコードとZコードを削除する(わが国の死因統計にはこれらは使用されていない.)。
- 一つの死亡データの中に全く同じICD-10コードがあった場合には一つのみを残す。
- 関連死因が死因分類(後述)66,67,68に当たるものは、分類不能や死亡のメカニズムを表すコードであり、対象から除外する。

また、死因分類としては、表3が提案されており、これには2つのレベルである15分類(groupと表記)と72分類(causeと表記)が存在するが、本研究では72分類を用いている。

この死因分類に従ってコード化されたデータを用いて、1.3節において述べたEgidi et al. (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 et al. (2018)と同様、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、Egidi et al. (2018)では、ネットワーク構造のグラフをICD-10の分類と対応付けながらサブグループ(コミュニティ)の検討を行っているのに対し、本研究では、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 [cfg], 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。

## 3 結果と考察

### 3.1 ネットワーク構造

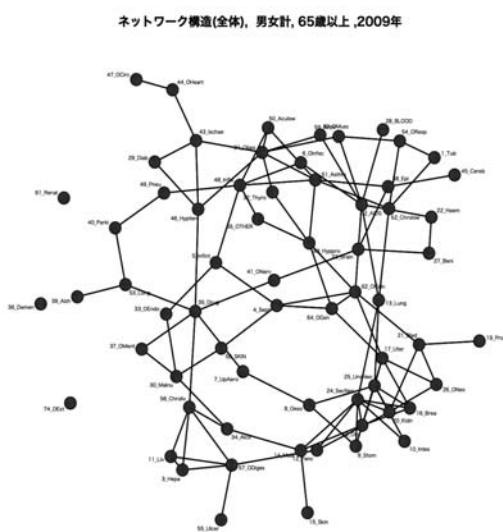
本研究におけるネットワーク分析の適用については、性・年齢に関し、男女計・全年齢、男女計・65歳以上、男性・65歳以上、女性・65歳以上の4通りについて分析を行った。それぞれのネットワーク構造をグラフで視覚化したものが図1~4である。

グラフからは、他のノードと結びついていない孤立点が存在することがわかる。これは、95%タイルよりも大きい関係だけを取り出したことによるものである。男女計・全年齢では、血液及び造血器の疾患並びに免疫機構の障害（28\_BLOOD）、認知症（アルツハイマー病を除く）（36\_Demen）、脳血管疾患（45\_Cereb）、腎不全（61\_Renal）、男女計・65歳以上では、認知症（アルツハイマー病を除く）（36\_Demen）、腎不全（61\_Renal）、その他の外因（74\_Oext）、男性の65歳以上では子宮の悪性新生物（17\_Uter）、血液及び造血器の疾患並びに免疫機構の障害（28\_BLOOD）、その他の外因（74\_Oext）、女性の65歳以上では、前立腺の悪性新生物（19\_Prost）、前立腺肥大（症）（63\_Hyppro）、その他の外因（74\_Oext）であり、さらに女性の65歳以上では、薬物依存（35\_Drug）と高血圧性疾患（46\_Hypten）が独立した連結成分を形成している。これらについては、連結成分でしか計算できない隣接中心性の評価と、コミュニティ検出の対象からは除外している。

図1 ネットワーク構造（男女計，全年齢）



図2 ネットワーク構造（男女計，65歳以上）



### 3.2 ネットワーク中心性指標

ネットワーク中心性指標の上位5位の死因を示したものが、表1である。まず、次数中心性を見ると、性別、年齢別に関わらず、上位5位以内に入っているのは、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）である。Egidi et al. (2018) でも、続発性悪性新生物は上位となっており、死因分類は異なるものの共通の傾向が見られると言えるが、肥満（症）及びその他の過栄養<過剰摂食>は上位に入っておらず、わが国とは異なる結果となっている。

次に、隣接中心性を見ると、次数中心性とは異なり、性別や年齢別で上位が異なっている傾向が見られる。肥満（症）及びその他の過栄養<過剰摂食>（31\_Obes）は男女計・全年齢、男性・65歳以上では1位であるが、男女計・65歳以上、女性・65歳以上では上位

図3 ネットワーク構造(男性, 65歳以上)

図4 ネットワーク構造(女性, 65歳以上)



表1 ネットワーク中心性指標の上位5位の死因

順位	次数中心性			隣接中心性			媒介中心性			
	略称	死因名	次数	略称	死因名	隣接性	略称	死因名	媒介性	
男女計全年齢	1	24_SecNeo	続発性悪性新生物	8	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	0.00541	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	462.6
	2	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	8	34_Alco	アルコール使用<飲酒>による精神及び行動の障害	0.00541	34_Alco	アルコール使用<飲酒>による精神及び行動の障害	367.8
	3	25_UnsNeo	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	7	50_Aculow	その他の急性下気道感染症	0.00521	48_Influ	インフルエンザ	354.1
	4	17_Uter	子宮の悪性新生物	6	35_Drug	薬物依存	0.00510	24_SecNeo	続発性悪性新生物	233.6
	5	48_Influ	インフルエンザ	6	14_Mela	皮膚の悪性黒色腫	0.00505	5_Infint	腸管感染症	204.2
男女計65歳以上	1	24_SecNeo	続発性悪性新生物	8	2_AIDS	ヒト免疫不全ウイルス [HIV] 病	0.00532	35_Drug	薬物依存	419.8
	2	25_UnsNeo	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	8	51_Asthm	喘息	0.00529	2_AIDS	ヒト免疫不全ウイルス [HIV] 病	298.8
	3	2_AIDS	ヒト免疫不全ウイルス [HIV] 病	7	62_OKidn	糸球体疾患及び腎尿管間質性疾患	0.00518	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	285.4
	4	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	7	13_Lung	喉頭, 気管, 気管支及び肺の悪性新生物	0.00518	51_Asthm	喘息	269.4
	5	17_Uter	子宮の悪性新生物	6	35_Drug	薬物依存	0.00510	13_Lung	喉頭, 気管, 気管支及び肺の悪性新生物	247.1
男性65歳以上	1	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	9	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	0.00592	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	500.1
	2	24_SecNeo	続発性悪性新生物	8	2_AIDS	ヒト免疫不全ウイルス [HIV] 病	0.00571	24_SecNeo	続発性悪性新生物	323.8
	3	2_AIDS	ヒト免疫不全ウイルス [HIV] 病	7	24_SecNeo	続発性悪性新生物	0.00556	13_Lung	喉頭, 気管, 気管支及び肺の悪性新生物	321.2
	4	56_Chroliv	慢性肝疾患	6	13_Lung	喉頭, 気管, 気管支及び肺の悪性新生物	0.00549	2_AIDS	ヒト免疫不全ウイルス [HIV] 病	311.9
	5	13_Lung	喉頭, 気管, 気管支及び肺の悪性新生物	5	58_SKIN	皮膚及び皮下組織の疾患	0.00538	58_SKIN	皮膚及び皮下組織の疾患	188.5
女性65歳以上	1	25_UnsNeo	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	10	34_Alco	アルコール使用<飲酒>による精神及び行動の障害	0.00568	34_Alco	アルコール使用<飲酒>による精神及び行動の障害	553.8
	2	24_SecNeo	続発性悪性新生物	8	52_Chrolow	その他の慢性下気道感染症	0.00526	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	411.0
	3	31_Obes	肥満(症)及びその他の過栄養<過剰摂食>	8	50_Aculow	その他の急性下気道感染症	0.00521	52_Chrolow	その他の慢性下気道感染症	371.8
	4	34_Alco	アルコール使用<飲酒>による精神及び行動の障害	7	37_OMent	その他の精神及び行動の障害	0.00515	50_Aculow	その他の急性下気道感染症	282.7
	5	2_AIDS	ヒト免疫不全ウイルス [HIV] 病	6	12_Panc	膵の悪性新生物	0.00510	2_AIDS	ヒト免疫不全ウイルス [HIV] 病	241.1

出所: 筆者作成



5位以内には入っていない。男女計・65歳以上で1位のヒト免疫不全ウイルス [HIV] 病 (2\_AIDS) は、男性・65歳以上でも2位である一方、男女計・全年齢、女性・65歳以上では上位5位以内には入っていない。

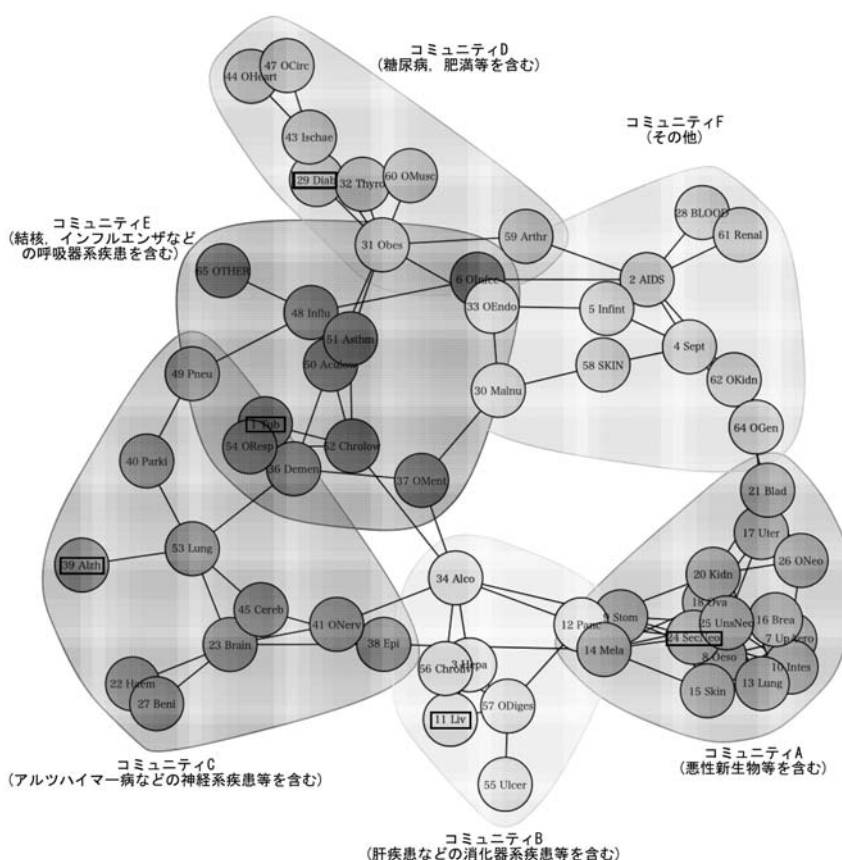
一方、媒介中心性を見ると、性別、年齢別に関わらず肥満 (症) 及びその他の過栄養 < 過剰摂食 > (31\_Obes) が1～3位に入っているという特徴が見られる。一方で、アルコール使用 < 飲酒 > による精神及び行動の障害 (34\_Alco) は女性・65歳以上で1位、男女計・全年齢で2位であるのに対し、男性・65歳以上、男女計・65歳以上では上位5位に入っておらず、性・年齢によって異なる特徴も観察された。

### 3.3 コミュニティ検出

次にコミュニティ検出の結果について述べる。女性・65歳以上について、Fast-greedy アルゴリズム (貪欲法) によってコミュニティ検出を行って、視覚化したものが図5である。

ここでは、死因が6つのコミュニティ (サブグループ) に分かれている。右下に位置し

図5 コミュニティ検出 (Fast-greedy), 女性, 65歳以上



ているのが悪性新生物等を含むコミュニティ A である。次に中央下側に位置しているのが肝疾患などの消化器系疾患等を含むコミュニティ B となっている。左下に位置しているのがアルツハイマー病などの神経系疾患等を含むコミュニティ C、中央上部に示されているのが糖尿病、肥満等を含むコミュニティ D、中ほどに位置しているのが結核、インフルエンザなどの呼吸器系疾患等が含まれるコミュニティ E であり、右上がそれ以外を示すコミュニティ F である。ここで、コミュニティ A~E を代表する死因として、A は続発性悪性新生物 (24\_SecNeo), B は肝, 肝内胆管, 胆のう及びその他の胆道の悪性新生物 (11\_Liv), C はアルツハイマー病 (39\_Alzh), D は糖尿病 (29\_Diab), E は結核 (1\_Tub) を選択し, Fast-greedy 以外のアルゴリズムによる結果についても, これらの代表する死因が含まれるコミュニティを同じ記号で表して比較したものが表 2 である。

これを見ると, どのアルゴリズムによる結果を見ても, 先の Fast-greedy アルゴリズムによる結果から得られた 5 つのコミュニティに類似したコミュニティが抽出されていることがわかる。しかしながら, その結果を詳細に観察すると, アルゴリズムによって異なる点があることがわかる。まず, 抽出されたコミュニティの数は, Fast-greedy [cfg], Walktrap [cw], Edge-betweenness [ceb], Louvain [cl] では 6, Spinglass [cs], Leading eigenvector [cle] では 7, Label propagation [clp] では 11 と異なっている。また, コミュニティに含まれる死因を観察すると, 肺炎 (49\_Pneu) は, Spinglass [cs], Fast-greedy [cfg], Label propagation [clp], Leading eigenvector [cle] ではアルツハイマー病 (39\_Alzh) を含むコミュニティ C に分類されるのに対して, それ以外のアルゴリズムでは結核 (1\_Tub) を含むコミュニティ E に分類されており, アルゴリズムの選択によってはどちらのグループと近いと判定されるかが大きく分かれる死因が存在することが理解される。

次に, このようにして検出されたコミュニティと, 1.1 節で述べた, 原死因と複合死因

表 2 コミュニティ検出結果の比較, 女性, 65 歳以上

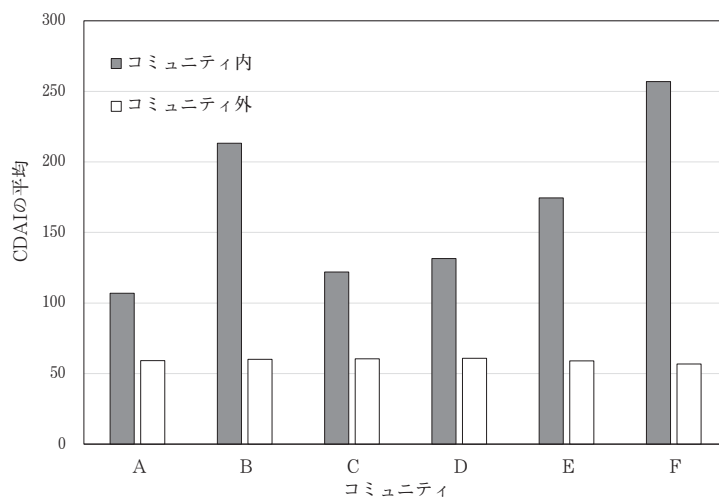
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20	21	22	23	24	25	26	27	28	29	30	31
	Tub	AIDS	Hepa	Sept	Infint	Ofnfc	UpAero	Oeso	Stom	Intes	Liv	Panc	Lung	Mela	Skin	Brea	Uter	Ova	Kidn	Blad	Haem	Brain	SecNeo	UnsNeo	ONeo	Beni	BLOOD	Diab	Malm	Obes
cfg	E	F	B	F	F	E	A	A	A	B	A	A	A	A	A	A	A	A	A	A	C	C	A	A	A	C	F	D	F	D
cw	E	F	B	F	F	E	A	A	A	B	A	A	A	A	A	A	A	A	A	A	C	C	A	A	A	C	F	D	F	E
ceb	E	F	B	F	F	E	A	A	A	B	A	A	A	A	A	A	A	A	A	A	C	C	A	A	A	C	F	D	F	E
cs	E	F	B	F	F	E	A	A	A	B	B	A	A	A	A	G	A	G	G	C	C	C	A	A	A	C	F	D	F	D
clp	E	F	B	G	G	H	A	A	A	B	B	A	A	A	A	A	A	A	A	A	I	I	A	A	A	I	F	D	J	D
cle	E	F	B	F	F	F	A	A	A	B	A	A	A	A	A	A	A	A	A	A	G	G	A	A	A	G	F	D	F	D
cl	E	F	B	F	F	E	A	A	A	B	B	A	A	A	A	A	A	A	A	A	C	C	A	A	A	C	F	D	F	D

	32	33	34	36	37	38	39	40	41	43	44	45	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	64	65
	Tybro	OEndo	Alo	Demn	OMent	Epi	Alzh	Parki	ONerv	Ischae	OHeart	Cereb	OCirc	Influ	Pneu	Aculow	Asthm	Chrolow	Lung	OResp	Uter	Chroliv	ODiges	SKIN	Arthr	OMuse	Renal	OKidn	OGen	OTHER
cfg	D	F	B	E	E	C	C	C	C	D	D	C	D	E	C	E	E	E	C	E	B	B	B	F	D	D	F	F	F	E
cw	E	F	E	E	E	C	C	E	C	D	D	C	D	E	E	E	E	E	C	E	B	B	B	F	F	E	F	F	F	E
ceb	E	F	B	C	C	C	C	E	C	D	D	C	D	E	E	E	E	E	C	E	B	B	B	F	F	E	F	F	F	E
cs	D	F	B	C	B	C	C	C	C	D	D	C	D	E	C	E	E	E	C	E	B	B	B	F	D	D	F	F	F	E
clp	D	G	B	B	B	I	C	C	I	D	K	I	K	H	C	H	H	H	I	E	B	B	B	G	D	D	F	G	G	H
cle	D	F	B	C	B	A	C	C	G	D	D	C	D	C	C	C	D	E	C	E	B	B	B	F	F	D	F	F	A	C
cl	D	F	B	C	F	C	C	E	C	D	D	C	D	E	E	E	E	E	C	E	B	B	B	F	F	D	F	F	F	E

出所: 筆者作成

図6 コミュニティ内外別の CDAI の平均値 (Fast-greedy), 女性, 65歳以上



出所: 筆者算定

との関連を表す指標である CDAI との整合性を検証するため、原死因と複合死因が両方ともある同一のコミュニティに含まれる場合の CDAI の平均値（ただし原死因と複合死因が同じ場合を除く）を「コミュニティ内 CDAI 平均値」、どちらかがそのコミュニティ以外に含まれる場合の平均値を「コミュニティ外 CDAI 平均値」とし、Fast-greedy アルゴリズムによるコミュニティについて、コミュニティ内外別 CDAI 平均値を示したものが図6である<sup>2)</sup>。これを見ると、どのコミュニティにおいても、コミュニティ内平均値はコミュニティ外平均値を大きく上回っていることが観察され、ネットワーク分析に基づくコミュニティは CDAI 指標と整合性を持っていることが理解できる。

おわりに

本研究では、複合死因データに関する分析手法について、海外・日本の先行研究を概観するとともに、特にネットワーク分析の応用についての先行研究をレビューした。そして、Egidi et al. (2018) によるネットワーク分析応用に関する方法論を日本の複合死因データに適用することにより、複合死因関連分析へのネットワーク分析応用の可能性を検討した。本研究で行った複合死因データに関する分析手法の概観からは、複合死因データが有する豊富な可能性が理解されたとともに、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よう。

しかしながら、残された課題も存在している。本研究では、MultiCause Network グ

2) 各原死因・複合死因別の CDAI の数値については、石井 [等] (2019) を参照。

ループのガイドラインに従い、2009年の日本の複合死因データについて、表3に示した死因分類を用いて分析を行った。しかしながら、このような分析は、対象とする年次や死因分類によって異なる結果が導かれる可能性も考えられる。したがって、2009年以外のデータへのネットワーク分析の適用や、死因簡単分類など MultiCause Network グループの分類以外の死因分類についても検討することが必要であろう。また、本研究では様々なコミュニティ検出アルゴリズムを適用し、複合死因ネットワークを、相互に関係の深いと考えられるいくつかのコミュニティへと分類することを試みた。しかしながら、詳細に観察すると、分類結果はアルゴリズムによって異なっており、どのアルゴリズムがわが国の複合死因分析に適しているのかということなどについても検討することが必要と考えられる。このような課題については、今後、さらに検討を行っていくこととしたい。

表 3 使用する死因分類

Group	Cause	略称	Disease or external cause	死因名	ICD-10 code	
1		INF	Infectious and parasitic diseases	感染症及び寄生虫症		
	1	Tub	Tuberculosis	結核	A15-A19, B90	
	2	AIDS	AIDS (HIV-disease)	ヒト免疫不全ウイルス [HIV] 病	B20-B24	
	3	Hepa	Viral hepatitis	ウイルス肝炎	B15-B19, B94.2	
	4	Sept	Septicaemia	敗血症	A40-A41	
	5	Infint	Intestinal infectious diseases	腸管感染症	A00-A09	
	6	OInfec	Other Infectious and parasitic diseases	その他の感染症及び寄生虫症	(A00-B99) + U01-U49 + U80-U89 - Supra 'A' and 'B' codes	
2		NEO	Neoplasms	新生物		
	7	UpAero	Malignant neoplasm of lip, oral cavity, pharynx	口唇, 口腔及び咽頭の悪性新生物	C00-C14	
	8	Oeso	Malignant neoplasm of oesophagus	食道の悪性新生物	C15	
	9	Stom	Malignant neoplasm of stomach	胃の悪性新生物	C16	
	10	Intes	Malignant neoplasm of small intestine, colon, rectum and anus, and other/ill-defined digestive organs	小腸, 結腸, 直腸 S 状結腸移行部, 直腸, 肛門, 肛門管, その他及び部位不明確の消化器の悪性新生物	C17, C18, C19, C20, C21, C26	
	11	Liv	Malignant neoplasm of liver, the intrahepatic bile ducts, gallbladder and other unspecified parts of biliary tract	肝, 肝内胆管, 胆のう及びその他の胆道の悪性新生物	C22, C23, C24	
	12	Panc	Malignant neoplasm of pancreas	膵の悪性新生物	C25	
	13	Lung	Malignant neoplasm of larynx and trachea/bronchus/lung	喉頭, 気管, 気管支及び肺の悪性新生物	C32-C34	
	14	Mela	Malignant melanoma of skin	皮膚の悪性黒色腫	C43	
	15	Skin	Malignant neoplasm of skin	皮膚のその他の悪性新生物	C44	
	16	Brea	Malignant neoplasm of breast	乳房の悪性新生物	C50	
	17	Uter	Malignant neoplasm of cervix uteri and other parts of uterus	子宮の悪性新生物	C53, C54, C55	
	18	Ova	Malignant neoplasm of ovary	卵巣の悪性新生物	C56	
	19	Prost	Malignant neoplasm of prostate	前立腺の悪性新生物	C61	
	20	Kidn	Malignant neoplasm of kidney	腎盂を除く腎の悪性新生物	C64	
	21	Blad	Malignant neoplasm of bladder	膀胱の悪性新生物	C67	
	22	Haem	Malignant neoplasm of lymph./haematopoietic tissue	リンパ組織, 造血組織及び関連組織の悪性新生物	C81-C96	
	23	Brain	Malignant neoplasm of eye, brain and other parts of central nervous system	眼, 脳及びその他の中枢神経系の部位の悪性新生物	C69-C72	
	24	SecNeo	Secondary malignant neoplasm	続発性悪性新生物	C77-C79	
	25	UnsNeo	Malignant neoplasm of ill-defined/unspecified/ independent (primary) multiple sites	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	C76, C80, C97	
	26	ONeo	Other malignant neoplasms	その他の悪性新生物	(C00-C99) - Supra 'C' codes	
	27	Beni	Benign neoplasms, In situ neoplasms and neoplasms of uncertain or unknown behaviour	その他の新生物	D00-D48	
	3	28	BLOOD	Diseases of the blood (-forming organs), immunol.disorders	血液及び造血器の疾患並びに免疫機構の障害	D50-D89
	4		ENDOC	Endocrine, nutritional and metabolic diseases	内分泌, 栄養及び代謝疾患	
		29	Diab	Diabetes mellitus	糖尿病	E10-E14
		30	Malnu	Malnutrition and other nutritional deficiencies	栄養失調 (症) 及びその他の栄養欠乏症	E40-E64
		31	Obes	Obesity	肥満 (症) 及びその他の過栄養 < 過剰摂食 >	E65-E68
32		Thyro	Disorders of thyroid gland	甲状腺障害	E00-E07	
33		OEndo	Other Endocrine, nutritional and metabolic diseases	その他の内分泌, 栄養及び代謝疾患	(E00-E90) - Supra 'E' codes	



Group	Cause	略称	Disease or external cause	死因名	ICD-10 code
5		MENT	Mental and behavioural disorders	精神及び行動の障害	
	34	Alco	Alcoholic psychosis/chronic alcohol abuse	アルコール使用<飲酒>による精神及び行動の障害	F10.1-F10.9, G31.2
	35	Drug	Drug dependence, toxicomania	薬物依存	F11-F16 except F11.0, F12.0, F13.0, F14.0, F15.0, F16.0, F18-F19 except F18.0, F19.0
	36	Demem	Dementias (excluding Alzheimer)	認知症 (アルツハイマー病を除く)	F01, F03, G31.0, G31.1, G31.8, G31.9
	37	OMent	Other Mental and behavioural disorders	その他の精神及び行動の障害	(F00-F99) except F10.0, F11.0, F12.0, F13.0, F14.0, F15.0, F16.0, F17.0, F18.0, F19.0 - Supra 'F' codes
6		NERV	Diseases of the nervous system	神経系の疾患	
	38	Epi	Epilepsia	てんかん及びびてんかん重積	G40-G41
	39	Alzh	Alzheimer's disease	アルツハイマー病	G30
	40	Parki	Parkinson's disease	パーキンソン病	G20, G21
	41	ONerv	Other Diseases of the nervous system	その他の神経系の疾患	(G00-G98) except (G31.0, G31.1, G31.2, G31.8, G31.9) - Supra 'G' codes
7		CIRC	Diseases of the circulatory system	循環器系の疾患	
	43	Ischae	Ischaemic heart diseases	虚血性心疾患	I20-I25
	44	OHeart	Other heart diseases	その他の心疾患	I00-I09, I30-I33, I39-I45, I47-I48, I49 except I49.0, I50-I52
	45	Cereb	Cerebrovascular diseases	脳血管疾患	I60-I69
	46	Hypten	Hypertensive diseases	高血圧性疾患	I10-I15
	47	OCirc	Other Diseases of the circulatory system	その他の循環器系の疾患	(I00-I99) except (I46, I49.0, I95.9, I99) - Supra 'I' codes
8		RESP	Diseases of the respiratory system	呼吸器系の疾患	
	48	Influ	Influenza	インフルエンザ	J10-J11
	49	Pneu	Pneumonia	肺炎	J12-J18
	50	Aculow	Other acute lower respiratory diseases	その他の急性下気道感染症	J00-J09, J19-J22
	51	Asthm	Asthma	喘息	J45-J46
	52	Chrolow	Other Chronic lower respiratory diseases	その他の慢性下気道感染症	J40-J44
	53	Lung	Lung diseases due to external agents	外的因子による肺疾患	J60-J70
	54	OResp	Other Diseases of the respiratory system	その他の呼吸器系の疾患	(J00-J99) except (J96.0, J96.9) - Supra 'J' codes
9		DIGES	Diseases of the digestive system	消化器系の疾患	
	55	Ulcer	Ulcer of stomach, duodenum and jejunum	胃潰瘍, 十二指腸潰瘍及び胃空腸潰瘍	K25-K28
	56	Chroliv	Chronic liver disease	慢性肝疾患	K70, K73-K74
	57	ODiges	Other Diseases of the digestive system	その他の消化器系の疾患	(K00-K93) - Supra 'K' codes
10	58	SKIN	Diseases of the skin and subcutaneous tissue	皮膚及び皮下組織の疾患	L00-L99
11		MUS	Diseases of the musculoskeletal system/connective tissue	筋骨格系及び結合組織の疾患	
	59	Arthr	Rheumatoid arthritis and osteoarthritis	関節リウマチ及び関節症	M05-M06, M15-M19
	60	OMusc	Other Diseases of the musculoskeletal system/connective tissue	その他の筋骨格系及び結合組織の疾患	(M00-M99) - Supra 'M' codes
12		GEN	Diseases of the genitourinary system	尿路性器系の疾患	
	61	Renal	Renal Failure	腎不全	N17-N19
	62	OKidn	Other Diseases of kidney and ureter	糸球体疾患及び腎尿管間質性疾患	N00-N16
	63	Hyppro	Hyperplasia of prostate	前立腺肥大(症)	N40
	64	OGen	Other Diseases of the genitourinary system	その他の尿路性器系の疾患	(N00-N99) - Supra 'N' codes
13	65	OTHER	Other diseases	その他の疾患	O00-O99, (P00-P96) except P28.5, Q00-Q99, H00-H95, R95

Group	Cause	略称	Disease or external cause	死因名	ICD-10 code
14		ILLDEF	Other Symptoms, signs, abnormal findings, ill-defined causes	症状、徴候及び異常臨床所見・異常検査所見で他に分類されないもの	
	66	Senil	Senility	老衰	R54
	67	Mecha	Mechanisms of the death	死亡のメカニズム	I46, I49.0, R09.2, R40.2, R57
	68	Oilldef	Other Symptoms, signs, abnormal findings, ill-defined causes	その他の症状、徴候及び異常臨床所見・異常検査所見で他に分類されないもの	(R00-R94), (R96-R99), I95.9, I99, J96.0, J96.9, P28.5, U00 - Supra "R" codes
15		EXT	External cause	傷病及び死亡の外因	S, T, V, W, X, Y codes + F10.0,F11.0, F12.0,F13.0, F14.0,F15.0, F16.0,F17.0, F18.0,F19.0
	70	Trans	Transport accidents	交通事故	V01-V99, Y85
	71	Fall	Accidental falls	転倒・転落	W00-W19
	72	Pois	Accidental poisoning	有害物質による不慮の中毒及び有害物質への曝露	X40-X49
	73	Suic	Suicide/intentional self-harm	自殺	X60-X84
	74	OExt	Other external causes	その他の外因	S, T, V, W, X, Y codes + F10.0,F11.0, F12.0,F13.0, F14.0,F15.0, F16.0,F17.0, F18.0,F19.0 - Supra codes

## 謝辞

本研究は、国立社会保障・人口問題研究所「長寿革命に係る人口学的観点からの総合的研究」、  
「超長寿社会における人口・経済・社会のモデリングと総合分析」の一環として行われており、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）JPMH20AB1001「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）による助成を受けた。

## 参考文献

- 石井太，林玲子，篠原恵美子，別府志海，是川夕（2019）「わが国の複合死因データによる死因間の関連分析」、『長寿革命に係る人口学的観点からの総合的研究：第2報告書』，pp.13-36.
- 上田フサ，加藤寛夫（1957）「複合死因に関する研究-心臓疾患を中心にして」、『厚生指標』，第4巻，第6号，pp.30-35.
- 上田フサ（1956）「複合死因の分類および製表に関する各国の研究報告」、『厚生指標』，第3巻，第7号，pp.45-54.
- （1960）「複合死因に関する研究」、『厚生指標』，第7巻，第7号，pp.32-68.
- 梅田珠実（1988）「複合死因からみた死因構造に関する研究」、『厚生指標』，第35巻，第13号，pp.11-17.
- 方波見重兵衛，臼井竹次郎，永井正規，箕輪真澄，福富和夫，進藤隆夫，金子功（1982）「糖尿病死亡の死因に関する研究-1-糖尿病死亡の年齢分布について」、『日本公衆衛生雑誌』，第29巻，第1号，pp.9-16.
- （1983）「糖尿病死亡の死因に関する研究-2-原因，二次死因糖尿病の相違について」、『日本公衆衛生雑誌』，第30巻，第2号，pp.59-68.
- 国立社会保障・人口問題研究所「日本版死亡データベース」  
<http://www.ipss.go.jp/p-toukei/JMD/index.asp>.
- 佐々木陽（1975）「死亡統計からみた糖尿病の死因とくに日米比較を中心に」、『糖尿病』，第18巻，第5号，pp.506-516.
- 篠原恵美子，別府志海，林玲子，石井太（2021）「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』，第41（Suppl.）巻，pp.733-737.
- 鈴木まき，横山英世，三宅健夫，野崎貞彦（2001）「複合死因からみた死因構造の解析」、『日大医学雑誌』，第60

- 卷, 第6号, pp.260-268.
- 林玲子, 石井太, 篠原恵美子, 別府志海, 是川夕 (2019) 「複合死因データの概況と突然死および認知症関連死亡の分析」, 『長寿革命に係る人口学的観点からの総合的研究: 第2報告書』, pp.37-54.
- 別府志海, 石井太, 林玲子, 篠原恵美子, 是川夕 (2019) 「複合死因データを用いた糖尿病関連死亡の分析」, 『長寿革命に係る人口学的観点からの総合的研究: 第2報告書』, pp.55-72.
- 別府志海, 石井太, 林玲子, 篠原恵美子 (2020) 「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」, 『長寿革命に係る人口学的観点からの総合的研究: 第3報告書』, pp.53-76.
- 矢久保孝介 (2013) 『複雑ネットワークとその構造』, 共立出版.
- Barabási, A.-L. (2019) 『ネットワーク科学: ひと・もの・ことの関係性をデータから解き明かす新しいアプローチ (池田裕一他監訳)』, 共立出版.
- Balkau, B. and L. Papoz (1992) "Certification of cause of death in French diabetic patients.", *Journal of Epidemiology & Community Health*, Vol. 46, No. 1, pp. 63-65.
- Barreto, S. M., V. M. A. Passos, S. K. F. Almeida, and T. D. Assis (2007) "The increase of diabetes mortality burden among Brazilian adults", *Revista Panamericana de Salud Pública*, Vol. 22, No. 4, pp. 239-245.
- Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008) "Fast unfolding of communities in large networks", *Journal of statistical mechanics: theory and experiment*, Vol. 2008, No. 10, p. P10008.
- Clauset, A., M. E. Newman, and C. Moore (2004) "Finding community structure in very large networks", *Physical review E*, Vol. 70, No. 6, p. 066111.
- Désesquelles, A., M. A. Salvatore, L. Frova, M. Pace, M. Pappagallo, F. Meslé, and V. Egidi (2010) "Revisiting the mortality of France and Italy with the multiple-cause-of-death approach", *Demographic research*, Vol. 23, No. 28, pp. 71-806.
- Désesquelles, A. F., M. A. Salvatore, M. Pappagallo, L. Frova, M. Pace, F. Meslé, and V. Egidi (2012) "Analysing multiple causes of death: Which methods for which data? An application to the cancer-related mortality in France and Italy", *European Journal of Population/Revue Européenne de Démographie*, Vol. 28, No. 4, pp. 467-498.
- Dushoff, J., J. B. Plotkin, C. Viboud, D. J. Earn, and L. Simonsen (2005) "Mortality due to influenza in the United States - an annualized regression approach using multiple-cause mortality data", *American journal of epidemiology*, Vol. 163, No. 2, pp. 181-187.
- Egidi, V., M. A. Salvatore, G. Rivellini, and S. D'Angelo (2018) "A network approach to studying cause-of-death interrelations", *Demographic Research*, Vol. 38, pp. 373-400.
- Gorina, Y. and H. Lentzner (2008) "Multiple Causes of Death in old age", *Aging Trends*, Vol. 9, pp. 1-9.
- Luke, D. A. (2015) *A User's Guide to Network Analysis in R*: Springer.
- Mant, J., S. Wilson, J. Parry, P. Bridge, R. Wilson, W. Murdoch, T. Quirke, M. Davies, M. Gammage, R. Harrison, and A. Warfield (2006) "Clinicians didn't reliably distinguish between different causes of cardiac death using case histories", *Journal of clinical epidemiology*, Vol. 59, No. 8, pp. 862-867.
- Manton, K. G. and E. Stallard (1982) "Temporal trends in US multiple cause of death mortality data: 1968 to 1977", *Demography*, Vol. 19, No. 4, pp. 527-547.
- Moreno-Betancur, M., H. Sadaoui, C. Piffaretti, and G. Rey (2017) "Survival analysis with multiple causes of death: extending the competing risks model", *Epidemiology*, Vol. 28, No. 1, pp. 12-19.
- Newman, M. E. and M. Girvan (2004) "Finding and evaluating community structure in networks", *Physical review E*, Vol. 69, No. 2, p. 026113.
- Newman, M. E. (2004) "Fast algorithm for detecting community structure in networks", *Physical review E*, Vol. 69, No. 6, p. 066133.
- White, M. C., S. Selvin, and D. W. Merrill (1989) "A study of multiple causes of death in California: 1955 and 1980", *Journal of Clinical Epidemiology*, Vol. 42, No. 4, pp. 355-365.
- Wilkins, K., M. Wysocki, C. Morin, and P. Wood (1997) "Multiple causes of death.", *HEALTH REPORTS/RAPPORTS SUR LA SANTE*, Vol. 9, No. 2, pp. 19-29.

# Methods and Their Application in the Analysis of Multiple Causes of Death Data: Application of Network Analysis to Japanese Data

ISHII Futoshi, HAYASHI Reiko, SHINOHARA Emiko and BEPPU Motomi

In this study, we review the methods of analysis for multiple causes of death (MCOD) data, which have become increasingly available in Japan in recent years. We use the method by Egidi et al. (2018), which applies network analysis to the Japanese MCODE data, and investigates its applicability.

There is a large number of studies on the MCODE data outside Japan, especially by researchers in the "MultiCause network," an international research network for MCODE. The studies are classified into the following groups: [1] Analysis of multiple causes as well as underlying causes; [2] Comparing multiple causes with underlying causes; [3] Analyzing the number of multiple causes per death; [4] Analyzing the relationship between multiple causes; [5] Application to life table analysis such as competing risk model and cause-deleted life tables; and [6] Validation of the coding of causes. In recent years, some studies in group [4] have been applying network analysis on MCODE. Amongst them, the study of Egidi et al. (2018) is a seminal one that applies network analysis to explore relationships between the causes of death; we used the same method for the Japanese data. We extract some causes that have a strong relationship with other causes, using centrality measures. Moreover, we apply community detection methods used in network analysis, for better understanding of the complex relationship between causes of death, which is not examined in Egidi et al. (2018).

We established various possibilities in MCODE data through a comprehensive review of the methodologies. From the results of this study, we can also confirm that network analysis is effective for analyzing the relationship between the multiple causes of death in Japan.

keywords: multiple causes of death, network analysis, mortality