

令和4年政策科学総合研究事業(統計情報総合研究事業)
分担研究報告書

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

死亡個票における死亡の原因欄の記載文字列の分析

研究分担者 篠原恵美子 東京大学大学院医学系研究科 特任助教

研究要旨 死亡個票データにおいて死亡の原因やその期間は自由入力データであり、統計処理に用いるためにはコード等への正規化が必要である。本年度は昨年度用いた正規化プログラムの変換精度の向上を図り、平成15年から令和3年までのデータの正規化を行った。

A. 研究目的

死亡個票における死亡の原因欄には病名が記載されるが、自由記載であるため様々な表記ゆれが含まれており、例えば「虚血性心筋症」と「心筋虚血」のように表現が異なる場合や、「肺癌」と「左肺癌」のように側性の情報が付加される場合がある。このような表記ゆれ関係にあるものを同一の病名として扱うためにはコード化を行う必要がある。また、「肺癌、動脈硬化症」のように1つの欄に複数の病名が含まれる場合には、それぞれを別の病名として計数できなければならない。原因とペアで記録される期間も自由記載であり、正規化処理をしなければ統計処理ができない。しかし死亡調査票の数は年間100万件を超えており、全件を手で処理することは現実的ではない。そこで自然言語処理による自動正規化が有用と期待される。

昨年度は過去に開発した正規化プログラ

ムを用い、平成15年から令和2年までの全データについて、死亡の原因欄に記載された内容のICD-10コード化および日数形式への変換を行った。本年度は変換精度の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。

B. 研究方法

死亡個票データのコード化については令和3年度に用いた変換プログラムを一部改変して利用した。改変の内容は3点あり、病名とICD-10コードの対応表として用いる標準病名マスターの最新版の組み入れ、令和2年のデータの変換結果をもとに作成した病名-ICD-10コードの対応関係の追加、および転落などの外因・自然死などICD-10コードが付与されない内因のカテゴリの詳細化である。

上述の変換プログラムを、平成15年から令和3年までの全データ(オンライン分、

18,562,156件)に適用した。

C. 研究結果

外因のカテゴリーは10種から18種へ、内因は7種から9種へ増加した。以下に分類内容を挙げる。

● 外因分類

1. 他殺
2. 焼死
3. 転落
4. 交通事故
5. 転倒
6. 餓死
7. 服毒・服薬
8. 凍死
9. 電撃死
10. 刑死
11. 土石流
12. 水の吸引
13. 煙の吸引
14. ヒートショック
15. 誤嚥
16. 頭部打撃
17. 圧死
18. その他

● 内因分類

1. 失血
2. 自然死
3. 心臓関連
4. 呼吸器関連
5. 循環器関連
6. 病死
7. 精神疾患
8. 徘徊
9. その他

変更した変換プログラムを全データに適

用した結果、全年について何らかの記載がある欄の97%にICD-10または外因・内因コードが付与された。

D. 考察

変換プログラムの変更の結果、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。

E. 結論

昨年度用いた正規化プログラムの改修を行い、平成15年から令和3年までの死亡個票データについて正規化を実施し、全年について何らかの記載がある欄の97%にコードが付与された。

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
なし

H. 知的財産権の出現・登録状況

なし