

一般口演 | 第40回医療情報学連合大会（第21回日本医療情報学会学術大会） | 一般口演

一般口演15

広域保健医療・連携医療支援

2020年11月20日(金) 16:40 ~ 18:20 D会場(コンgresセンター4階・43~44会議室)

[3-D-4-03] 死因統計の精度と効率性の向上に向けた我が国の原死因確定課題の抽出

*明神 大也^{1,2}、大井川 仁美⁴、香川 璃奈³、今村 知明¹、今井 健⁵（1. 奈良県立医科大学 公衆衛生学講座, 2. 奈良県立医科大学 病理診断学講座, 3. 筑波大学医学医療系 医療情報マネジメント学, 4. 奈良県立医科大学 MBT学講座, 5. 東京大学大学院医学系研究科 疾患生命工学センター）

*Tomoya Myojin^{1,2}, Hitomi Oigawa⁴, Rina Kagawa³, Tomoaki Imamura¹, Takeshi Imai⁵（1. 奈良県立医科大学 公衆衛生学講座, 2. 奈良県立医科大学 病理診断学講座, 3. 筑波大学医学医療系 医療情報マネジメント学, 4. 奈良県立医科大学 MBT学講座, 5. 東京大学大学院医学系研究科 疾患生命工学センター）

キーワード：ICD-11, Auto coding, Causes of death, Underlying Cause of Death, ICD-10

【背景】日本では益々人口減少・高齢化の進展が予想されるが、今後のICD-11導入に際し死因統計の正確性の担保・より一層の効率性向上を図るためには、現行の原死因確定プロセスの課題を抽出することが重要である。これまでのヒアリング調査で、オートコーディングツールを利用した現行の原死因確定ロジックでは、死亡票の約4割に対し目視確認が必要であることが判明しているが、実データに基づく詳細調査は未だ十分に行われていない。

【目的】死亡票・死亡個票実データを用いて、現行の死因確定プロセスにおける課題点についての調査を行う。

【方法】人口動態調査票情報の提供を受け、平成27年~30年の死亡票・死亡個票を対象とし、結合処理を行って基礎的な統計量を得た。また、オートコーディングツール処理の結果、目視確認が必要とされた死亡票の一部を抽出し、目視確認の原因や原死因の変更について調査を行い、原死因確定への影響について推計を行った。

【結果】死亡票と死亡個票の結合処理(結合率99%)により得られた約520万件のうち、何らかの付帯情報（傷病名以外の、手術や解剖所見・備考欄・外因死の追加事項など）が含まれていた割合は約3割であった。また目視確認に回る理由は付帯情報がある場合以外にも、死因をICD-10コードに変換できない、死因に病名以外の外因が含まれている、などの場合が存在していた。目視確認の対象となる死亡票は全体の約3分の1程度で、そのうち原死因が変更になる割合は約10分の1程度であると推定された。

【考察と結論】実データを用いて現行の原死因確定ロジックの課題を明らかにした。ICD-11導入にあわせ、より高精度で効率性の高いロジックの開発が必要であるが、このためには自然言語処理による病名正規化処理と、原死因変更の有無を高精度に予測する機械学習アルゴリズムが有効と考えられた。今後はこの開発を進めていく予定である。

死因統計の精度と効率性の向上に向けた我が国の原死因確定課題の抽出

明神 大也^{*1*}、大井川 仁美^{*3}、香川 璃奈^{*4}、今村 知明^{*1}、今井 健^{*5}

^{*1} 奈良県立医科大学 公衆衛生学講座、^{*2} 奈良県立医科大学 病理診断学講座、

^{*3} 奈良県立医科大学 MBT学講座、^{*4} 筑波大学医学医療系 医療情報マネジメント学、

^{*5} 東京大学大学院医学系研究科 疾患生命工学センター

Extraction of issues in determining the underlying cause of death in Japan to improve the accuracy and efficiency of cause-of-death statistics

Myojin Tomoya^{*1*}, Oigawa Hitomi^{*3}, Kagawa Rina^{*4}, Imamura Tomoaki^{*1}, Imai Takeshi^{*5}

^{*1} Department of Public Health, Health Management and Policy, Nara Medical University,

^{*2} Department of Diagnostic Pathology, Nara Medical University,

^{*3} Department of MBT, Nara Medical University,

^{*4} Department of Biomedical Informatics and Management, Faculty of Medicine, University of Tsukuba,

^{*5} Center for Disease Biology and Integrative Medicine, Faculty of Medicine, the University of Tokyo

[Background] In Japan, it is expected that the population is aging and declining rapidly. In order to ensure the accuracy of the cause of death statistics when introducing ICD-11 in the future, it is important to identify current issues of the confirmation process of "underlying cause of death". The present process utilizes an auto-coding tool. We revealed that it requires visual confirmation for about 40% of death votes through our hearing survey. However, little has been operated the detailed researches based on actual data.

[Purpose] It is aimed to investigate the problems in the current process of determining the underlying cause of death by using the questionnaire information of death votes and death cards.

[Method] Through the death votes and death cards for 2015 to 2018, we combined them and collected the basic statistics. After transacting of an auto-coding tool, we extracted a part of death votes which was checked visually and surveyed the cause of visual confirmation and the alternation of underlying cause of death.

[Results] There are any information other than the name of injury and illness (E.g. operation, anatomical finding, remarks column and additional information of extrinsic death) in about 30% of the approximately 5.2 million cases obtained by the combined processing of the death vote and the death card (combination rate 99%). The reason for visual confirmation was not only when there was incidental information in death vote, but also when the cause of death could not be converted to the ICD-10 code, or when the cause of death included an external cause other than the disease name. We estimated that about 33% of total death votes were checked visually and about 3% of them were modified the original underlying cause of death to a new one.

[Discussion and Conclusion] We clarified the problems of the current logic for determining the cause of underlying death using actual data. Along with the introduction of ICD-11, it is necessary to develop more accurate and highly efficient logic. For this purpose, it was considered that the disease name normalization processing by natural language processing and the machine learning algorithm that predicts the presence or absence of the original cause of death with high accuracy are effective. We plan to continue this development in the future.

Keywords: ICD-11, Auto coding, Causes of death, Underlying Cause of Death, ICD-10

1. 緒論

我が国において人口動態調査は国勢調査と並ぶ国の基幹統計であり、中でも死因統計は最も重要な情報の一つである。今後 ICD-11 を国内適用するにあたっては原死因データを適切に収集・分析し、国際比較可能なデータを提供することが求められている。レセプトや現在普及が進む電子カルテでは標準病名の採用が進められているが、人口動態調査の死因は自由入力病名が元となっており、完全な自動集計は困難である。また、我が国では高齢化が進み死亡者数の増加が見込まれることから、より正確で効率性の高いデータ収集の方法の検討が求められている。

我々は平成 30 年度 厚生労働統計協会 調査研究委託事業「原死因確定作業についての実態・問題点の把握、ならび

に正確・効率性向上に向けた機械学習の適用可能性と課題に関する調査研究」において、原死因確定の一連の流れとその課題¹⁾を明らかにした。医師または歯科医師によって記載された死亡診断書は死亡届とともに市区町村に提出される。そして市区町村の職員が死亡診断書に記載された内容を情報端末に入力し、「死亡個票」として、保健所・都道府県を経て厚生労働省に提出される。厚生労働省にて死亡個票はオートコーディングツールにかけられて原死因が割り振られ、その一部は目視確認に回り、必要に応じて原死因が変更され、確定原死因になって「死亡票」が作成される。

オートコーディングツールを利用した現行の原死因確定ロジックでは、死亡票の約4割に対し目視確認が必要であることが判明しているが、実データに基づく詳細調査は未だ十分に

行われていない。

本研究では、死亡票と死亡個票の調査票情報の提供を受け、基礎的調査を行った。加えて、目視確認に回った事案をサンプリング抽出することで、原因確定の全体像の把握を試みた。

2. 方法

2.1 死亡票と死亡個票の基礎的調査

統計法 33 条第 1 項に基づき²⁾、平成 27 年～平成 30 年の 4 年間の死亡票とオンライン申請された死亡個票の調査票情報の利用申請を行った。死亡票では、調査年、提出年月、届出地、事件簿番号、性別、出生年月日時分、死亡年月日時分、原因、外因符号、母側病態、単多胎別、妊娠週数、母の生年月日、前回の妊娠、子の数の項目の利用申請を行った。また死亡個票では、死亡届出地の都道府県と市町村、事件簿番号、処理年月、備考欄記述有無、死亡した人の都道府県と市町村、死亡したところの種別、「死亡の原因」に含まれる I 欄・II 欄の原因と期間、手術の部位及び所見、手術年月日、解剖の部位及び所見、死因の種類、障害が発生した年月日時分、傷害が発生したところの種別、傷害が発生したところ、その他の記述、傷害発生場所、手段及び状況、生後 1 年未満での病死の病態・異状の詳細、その他付言すべき事柄、備考欄の項目の利用申請を行った。

なお、死亡個票に含まれる、手術の部位及び所見、手術年月日、解剖の部位及び所見、死因の種類、障害が発生した年月日時分、傷害が発生したところの種別、傷害が発生したところ、その他の記述、傷害発生場所、手段及び状況、生後 1 年未満での病死の病態・異状の詳細、その他付言すべき事柄を総じて「付帯情報」と称した。

そして実際に調査票情報の提供を受けて、死亡票と死亡個票を SQL server に取り込み、粗集計を行った。併せて、データの信頼性を確認するため、公開データである人口動態統計の死亡数と比較した。

その後、死亡票と死亡個票を結合した。結合条件は、処理年および届出地の都道府県・保健所・市区町村・事件簿の番号が一致するもので内部結合とした。結合後、死亡の原因の各欄に入力のある件数の割合を求めるとともに、付帯情報の記載があった割合を求めた。また、自由入力欄の多岐性を確認した。

本研究では Microsoft SQL server 2018 Standard Edition を用いて調査を行った。改行は CRLF、文字コードは UTF-8 として処理した。

2.2 目視確認リストの基礎的調査

本研究では上記に加えて、統計法 22 条に基づき、目視確認に回ったリストの調査票情報の提供を受けた。死亡個票はオートコーディングツールにかけられて仮原因が確定する。その際の死亡個票に付帯情報が入っておらず、且つコーディングエラーが発生しなければ仮原因が確定原因になる。一方、付帯情報が入っていたりコーディングエラーが発生していたりすれば目視確認に回る。コーディングエラーは何らかの理由で死因に ICD-10 コードが振られなかった場合、または死因に外死因(病死以外)が含まれる場合に発生する。そこでオートコーディングツール実行後の死亡票のうち、ランダムサンプリングされた 100 件の目視確認に回ったリストの調査票情報の提供を受け、基礎的調査を行った。

2.3 倫理面への配慮

本研究では個人情報や動物愛護に関わる調査・実験は行わず、研究の遂行に当たっては、各種法令や「人を対象とする医学系研究に関する倫理指針」を含めた各種倫理指針等の遵守に努めた。

3. 結果

人口動態統計の死亡数³⁾は平成 27 年、28 年、29 年、30 年の順に、1,290,510 人(男性 666,728 人/女性 623,782 人)、1,308,158 人(男性 674,946 人/女性 633,212 人)、1,340,567 人(男性 690,770 人/女性 649,797 人)、1,362,470 人(男性 699,138 人/女性 663,332 人)であった。

提供を受けた死亡票のレコード数は平成 27 年、28 年、29 年、30 年の順に、1,301,379 件、1,319,030 件、1,351,944 件、1,374,469 件(合計 5,346,822 件)であった。死亡個票のレコード数は平成 27 年、28 年、29 年、30 年の順に、1,249,469 件、1,280,853 件、1,325,955 件、1,374,780 件(合計 5,231,057 件)であった。

その結果、死亡票と死亡個票が結合できたのは平成 27 年、28 年、29 年、30 年の順に、1,248,057 件(男性 645,711 件/女性 602,346 件)、1,280,808 件(男性 661,501 件/女性 619,307 件)、1,325,413 件(男性 683,577 件/女性 641,836 件)、1,373,589 件(男性 705,560 件/女性 668,029 件)であった。合計 5,227,867 件(男性 2,696,349 件、女性 2,531,518 件)であった。

このうち、「死亡の原因」- [I 欄]- [ア欄]の原因・期間に入力があったのはそれぞれ 5,219,044 件(99.8%)・5,143,096 件(98.4%)であった。[イ欄]の原因・期間に入力があったのはそれぞれ 1,861,677 件(35.6%)・1,445,585 件(27.7%)、[ウ欄]の原因・期間に入力があったのはそれぞれ 393,799 件(7.5%)・312,152 件(6.0%)、[エ欄]の原因・期間に入力があったのはそれぞれ 73,542 件(1.4%)・54,670 件(1.0%)、[II 欄]の原因・期間に入力があったのはそれぞれ 1,744,402 件(33.4%)・1,628,459 件(31.1%)であった。付帯情報に記載があったのは 1,716,961 件(32.8%)であった。

原因・期間ともに自由入力であったため、多数の類似表記が確認された。例えば原因が誤嚥性肺炎の場合、誤嚥性肺炎(90.4%)、嚥下性肺炎(4.4%)、誤えん性肺炎(1.7%)、誤嚥性肺炎(1.2%)、反復性誤嚥性肺炎(0.3%)、肺炎(誤嚥性)(0.2%)などが 40 種類以上の表記が存在していた。期間についても 1カ月の場合、1ヶ月、約1ヶ月、1カ月、約4週間、間もなく1ヶ月などが存在していた。

また提供を受けた目視確認に回ったリストでは、ある 1 カ月間の死亡票 約 113 千件のうち、目視確認に回ったのは 40 千件で、割合は 35.6%であった。そのうちランダムサンプリングした 100 件のうち、付帯情報が含まれていたのが 80 件、コーディングエラーが含まれていたのが 45 件、付帯情報とコーディングエラーの両方が含まれていたのが 24 件であった。目視確認に回った死亡票のうち、仮原因と確定原因が異なった(目視確認で変更になった)のは 10 件、外因に関わる符号または母側に関わる符号が追加されたのは 9 件、変更なしが 81 件であった。

4. 考察

人口動態の死亡数に比べて死亡票の件数は約 1%(4 年間で 45 千件)多かった。これは死亡票には外国籍の死亡等が

含まれていた⁴⁾からと考えられる。

死亡票に比べて死亡個票の件数が約2%(4年間で116万件)多かった。これは、死亡個票にはオンライン申請されたもののみ含まれるが、一部の自治体および海外での死亡は含まれない^{1,5)}ためと考えられる。

また、死亡票・死亡個票の調査票情報提供および目視確認リストの調査票情報提供により、仮原死因を付与された死亡票は図1の流れになると考えられた。目視確認に回る割合35.6%は実数ではなく、1ヶ月の抽出死亡票に対する値である。また、原死因の変更や外因の追加が行われる割合も、そこからランダム抽出された100件に対するものであり、統計的な信頼性は必ずしも高くはない。しかし、付帯情報があることにより目視確認に回る事例(32.8%)以外にも、付帯情報は無いがコーディングエラーが生じる事例(約3%)があり、合わせた結果として目視確認が35.6%に対し行われている、と考えられた。また、目視確認の結果原死因が変更される割合は少なく、チェックされたものの約10%、全体の約3~4%であると推定された。

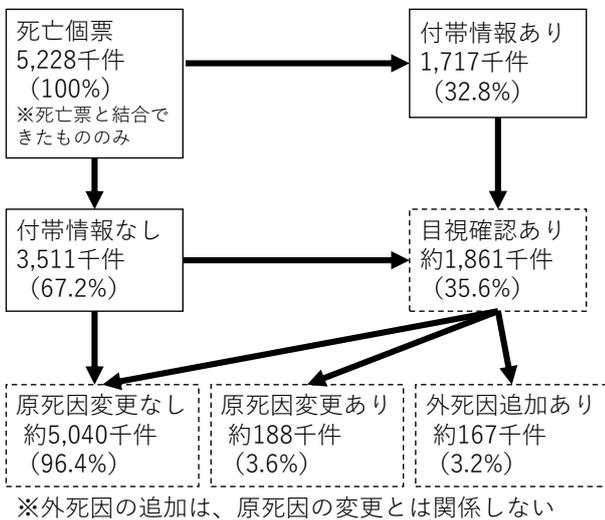


図1 原死因確定の流れ

なお、図1において外枠が実線の場合は死亡票・死亡個票から算出した実データで、点線の場合は抽出された目視確認リストから推察した予測値を示す。また図1中で、外因の追加は原死因の変更とは関係しない。

付帯情報やコーディングエラーの確認の結果、原死因が変更される割合は少なく、大多数は変更ないと考えられる。したがって原死因の変更の有無を高精度に予測する機械学習アルゴリズムの導入は、目視確認作業効率化のために有効と考えられた。

一方、死亡の原因・期間は自由記述になっており、前述のように、多様な表現が存在している。今後 ICD-11 導入に合わせ、高い精度で効率性の高い原死因確定ロジックへ円滑に移行していく必要があるが、仮に国際的に広く用いられているオートコーディングツールである Iris⁶⁾を活用する場合には、この自由記述に対する自然言語処理による高精度な ICD コード付与が重要となると考えられた。これは、現行の我が国でのオートコーディングシステムで、目視確認に回るコーディングエラーを減少させる上でも、重要である。

5. 結論

本研究では、実データを用いて現行の原死因確定ロジックの流れ、また課題を明らかにした。ICD-11 導入にあわせ、より高い精度で効率性の高いロジックの開発が必要であるが、このためには自然言語処理による病名正規化処理と、原死因変更の有無を高精度に予測する機械学習アルゴリズムが有効と考えられた。今後はこの開発を進めていく予定である。

謝辞

本研究は、厚生労働科学研究費補助金(政策科学総合研究事業(統計情報総合研究事業))「死因統計の精度及び効率性の向上に資する機械学習の検討に関する研究(H30-医療-一般-013)」および厚生労働統計協会平成30年度調査研究委託事業「原死因確定作業についての実態・問題点の把握、ならびに正確・効率性向上に向けた機械学習の適用可能性と課題に関する調査研究」の一環として実施したものである。

参考文献

- 1) 今井健、明神大也、大井川仁美、香川璃奈、今村知明. 原死因確定作業についての実態・問題点の把握、ならびに正確・効率性向上に向けた機械学習の適用可能性と課題に関する調査研究. 厚生労働省. 67(3), 17-24, 2020-03.
- 2) 統計法第33条第1項による調査票情報の提供について. 厚生労働省. [https://www.mhlw.go.jp/toukei/sonota/chousahyo.html (cited 2020-Aug-30)].
- 3) 平成30年(2018)人口動態統計(確定数)の概況, 第2表-1 人口動態総覧の年次推移. 厚生労働省. [https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei18/index.html (cited 2020-Aug-30)].
- 4) 平成30年(2018)人口動態統計(確定数)の概況, 調査の概要. 厚生労働省. [https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei18/dl/01_cho.pdf (cited 2019-Aug-30)].
- 5) 石井太、林玲子、篠原恵美子、別府志海、是川タ. わが国の複合死因データによる死因間の関連分析. 国立社会保障・人口問題研究所 2017~2019 年度一般会計プロジェクト「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. 2019. [http://www.ipss.go.jp/projects/j/Chouju/reports/FY2018_Ishii.pdf (cited 2019-Aug-30)].
- 6) Iris Institute. Federal Institute for Drugs and Medical Devices. [https://www.iris-institute.org (cited 2019-Aug-30)]