

ポスター

[PA1～PA19] ポスター

2018年6月22日(金) 15:00 ～ 16:00 ポスター会場 (3階・中会議室302)

[PA14] 死亡診断書における原死因の自動決定に関する試み

横田 賢一（長崎大学 原爆後障害研究所）

死亡診断書における原死因の自動決定に関する試み

横田 賢一, 中村 洋子, 三根 眞理子

長崎大学 原爆後障害医療研究所

Automated decision of underlying cause and ICD coding on a death certificate

Kenichi Yokota, Yoko Nakamura, Mariko Mine

Atomic Bomb Disease Institute, Nagasaki University

抄録:

原爆被爆者の疫学解析のため、1970 年以降、死亡診断書に基づき死因の ICD コード化を行ってきた。近年、人的資源の確保が困難となりコーディング作業の省力化のため、死因名テキストからの ICD コード化と原死因選択の辞書化を進めている。過去に人手でコーディングされた ICD コードデータと新たにデータ化した死因名テキストデータを用いて対応辞書を作成した。2008 年の 2,644 件の死因テキストデータから 713 件が辞書化できた。今後、2011 年までのデータにより辞書を充実させる。また、同様に ICD コード化された死因のパターンから選択される原死因の辞書を作成し、各辞書に学習機能を付加するための処理フローについて検討、評価を行っていく。

キーワード 死亡原因, ICD コード, 原死因, 自動化

1. はじめに

長崎大学は、原爆被爆者の健康影響研究のため、長崎県・市との協定に基づき被爆者の追跡を行っている。近年、被爆者の高齢化は進み死亡者数は漸増している。解析を行う上で重要な死因情報は 1970 年から 2016 年までは、死亡診断書の死因を人手により ICD-9 または ICD-10 によりコード化し、WHO の選択ルール[1]に準拠した方法で原死因を決定してきた。しかし、近年、熟練した診療情報管理士等の確保が困難となり、作業コストは膨大であるため作業の自動化、省力化が急務となっている。現在、年間 1,500 人程度をコーディングしなければならない。死亡診断書の記述はバリエーションに富んでおり、完全自動化にはかなりの工夫が必要であるが[2]、類型化は可能と考え、過去のデータに基づくパターン辞書を作成することにより省力化が可能と考えた。人手に回ったものも登録が可能なものは辞書に追加していく学習機能も不可欠である。これらの処理を確立するために、本研究では辞書適用の効率と処理の流れを検討することを目的とした。

2. 方法

1) 処理の流れ

毎年度まとめて持ち込まれる死亡診断書について、入力担当者が死因欄に記載された内容を

テキスト入力しデータを作成する。これを入力として、死因テキストに対する ICD コードの対応表である辞書1を適用し、自動的にコードに変換する。さらに(複数の)死因コードの組合せとそれらから選択される原死因コードとの対応表である辞書2を適用し、自動的に原死因コードを決定する(図 1)。各辞書になかったものは人手により ICD コーディング、原死因決定を行い、各辞書に追加しておき、次の処理に適用する。

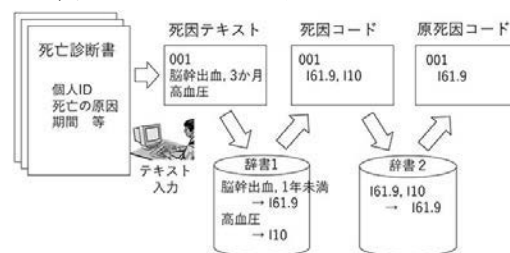


図 1. 処理の流れ

2) ICD コード化辞書の作成（辞書 1）

熟練した診療情報管理士によりコーディングされた死因コードと、その元となった死亡診断書から作成した死因テキストデータとを照合し、対応が一意でないものを除外して辞書1を作成した。コーディングに期間を必要とする死因の場合は期間データが含まれる。なお、死因テキスト、ICD コードの確認には医療情報システム開発センターの標準病名マスター[3]を利用した。

3) 原死因選択辞書の作成（辞書2）

データベースの死亡情報には、死亡診断書に記載されていた全ての死因とそれから選択された死亡原因のデータが保存されている。これらから各個人の死因(コード)の組合せとそれから選択された原死因(コード)との対応を抽出し、一意でないもの、誤り、情報不足等を除外して辞書2を作成する。

3. 結果

1) 1死亡あたりの死因数の平均

2008～2011年度の死亡数と記載された死因数は表1のとおりである。平均で1死亡あたり2.1個の死因が記載されている。各年の記載数のばらつきは少ない。4年間の最大値は14個、最大4個までで95%を占めていた。原死因を選択するパターンは多くはないと考えられる。

表1. 死亡あたりの死因記載数の平均

年度	死亡数	死因数	死因数/死亡数
2008	1,451	3,024	2.08
2009*	580	1,215	2.09
2010	1,556	3,279	2.11
2011	1,535	3,192	2.08
合計	5,122	10,710	2.09

* 2009年4-8月

2) 辞書の作成過程

2008年度の3,024件のテキストデータのうち、2,986件(98.7%)がテキストとコードとの照合ができた。そのうち2,644件(88.6%)が辞書化作業の対象となる。残りのうち192件(6.4%)は発病から死亡までの期間の情報を必要としていた。それらは心筋梗塞、脳梗塞、脳血栓症、脳幹部梗塞、脳挫傷、C型肝炎、くも膜下出血、脳出血であった。例を表2に示す。不明瞭な記述、事故や自殺などは除外した。整理後713件を辞書1に登録した。引き続き2009年度以降について進めている。

表2. 死亡までの期間とコードの例

記載テキスト	期 間	ICD-10コード	病 名
心筋梗塞	1か月未満	I21.9	急性心筋梗塞
	1か月以上	I25.2	陳旧性心筋梗塞
脳梗塞 / 脳塞栓	1年未満	I63.4	塞栓性脳塞栓
	1年以上	I69.3	陳旧性脳梗塞
くも膜下出血	1年未満	I60.9	くも膜下出血
	1年以上	I69.0	くも膜下出血後遺症

4. 考察

本研究に先立ち、概要をつかむために2008年から2011年までのテキストとコードを用いて機械的な対応表を作成し自動化の可能性を調べた。2014, 2015年の死亡診断書の記載全てが自動コード化できた割合は76.1%と78.7%であった。入力されたテキストの8割前後は自動コード化できる可能性がある。辞書2の効率は今後の検証が必要であるが、辞書1に比べ学習による効果向上が期待できるかもしれない。全体の自動化割合を仮に70%とした場合、死亡1,500件のうち、手作業でのコーディングは450件となり、十分な省力化が図れる。一方、バリエーションに富む記述の標準化のため入力担当者へのトレーニングが重要と考えられた。

5. 結語

2008年度の死亡診断書1,451件の記載内容のテキストデータと死因コードから死因名が適切で、一意にコードと対応していた713件を辞書化した。今後、原死因の自動選択辞書の作成および評価を進める。テキストデータ作成の際には入力者のトレーニングが重要である。

参 考 文 献

- [1] 厚生省大臣官房統計情報部編: 疾病、傷害および死因統計分類提要 ICD-10 準拠第1巻, 厚生統計協会, 東京都, 1995.
- [2] 今井 健, 荒牧英治, 梶野正幸, 美代賢吾, 大江和彦: 臨床医学分野における用語概念間の関係情報を用いた自動 ICD コーディングに関する研究, The 22nd Annual Conference of the Japanese Society for Artificial Intelligence, 2008.
- [3] 医療情報システム開発センター: ICD10 対応標準病名マスター, 2018. [<http://www2.medis.or.jp/stdcd/byomei/index.html> (cited 2018-Apr-04)].