

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-4-04] 原死因確定プロセスにおける IRISの国内導入可能性に関する基礎的な検討

*大井川 仁美¹、明神 大也^{2,3}、香川 璃奈⁴、今村 知明³、今井 健⁵（1. 奈良県立医科大学MBT学講座, 2. 奈良県立医科大学病理診断学講座, 3. 奈良県立医科大学公衆衛生学講座, 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）

*Hitomi Oigawa¹, Tomoya Myojin^{2,3}, Rina Kagawa⁴, Tomoaki Imamura³, Takeshi Imai⁵（1. 奈良県立医科大学 MBT学講座, 2. 奈良県立医科大学病理診断学講座, 3. 奈良県立医科大学公衆衛生学講座, 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）

キーワード：underlying cause of death, Vital Statistics, Iris, ICD-10

【背景】日本では、死亡統計に用いる原死因の確定を行うために、人口動態調査死亡票と独自開発した人口動態死因オートコーディングシステムを用いている。一方、世界ではフリーソフトウェア Irisの利用が広まっている。ICD-11の導入に伴い、人口動態死因オートコーディングシステムを大幅に修正する可能性がある中で、日本はどのような原死因確定プロセスを行うのか検討する必要がある。【目的】Irisが日本の原死因確定に適用可能か調査を行う。Irisは、病名ではなく病名に対応する ICD-10コードを読み込ませる必要があるため、ICD-10コードのみで決定される原死因の精度を算出する。【方法】人口動態死因オートコーディングシステムと Irisが準拠する「疾病、傷害および死因統計分類提要 ICD-10準拠第2巻インストラクションマニュアル<総論>」（以下、インストラクションマニュアル）中の病名と原死因が併記された原死因コーディング例(156件)を抽出し、標準病名マスター等を用いて病名に ICD-10コードを付与した。付与した ICD-10コードを Irisに適用して得た原死因と、コーディング例の原死因を比較し、Irisの原死因確定精度を算出した。【結果】Irisの原死因確定精度は約8割であった。残りの原死因に関しては、標準病名マスターとのポリシーの相違、根本的に ICD-10コードの付与が困難な例、Irisが未対応であるインストラクションマニュアルのルールに則った原死因の決定が必要な例があった。【考察】正しい ICD-10コードが付与できれば、Irisにより約8割以上の原死因確定が可能だと考えられる。今回は、特殊な例を多く含むインストラクションマニュアル中の原死因コーディング例を用いているが、実際の死亡票のデータを用いた場合はより精度が高くなる可能性も考えられるため、今後は実データを用いた実験を行う予定である。

原死因確定プロセスにおける IRIS の国内導入可能性に関する基礎的な検討

大井川 仁美^{*1}、明神 大也^{*2,*3}、香川 璃奈^{*4}、今村 知明^{*3}、今井 健^{*5}

*1 奈良県立医科大学 MBT 学講座、*2 奈良県立医科大学病理診断学講座、

*3 奈良県立医科大学公衆衛生学講座、*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

Feasibility of Domestic Use of Iris in the Determination Process of the Underlying Cause of Death

Hitomi Oigawa^{*1}, Tomoya Myojin^{*2,*3}, Rina Kagawa^{*4}, Tomoaki Imamura^{*3}, Takeshi Imai^{*5}

*1 Department of MBT, Nara Medical University,

*2 Department of Diagnostic Pathology, Nara Medical University,

*3 Department of Public Health, Health Management and Policy, Nara Medical University,

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

*5 Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo

[Background and Purpose] In Japan, to determine the underlying cause of death, the Ministry of Health, Labour and Welfare of Japan use their auto-coding system. On the other hand, the use of free software Iris is widespread in the world. In this study, we investigated whether Iris could replace the Japanese auto-coding system. [Method] We coded ICD-10 codes for “underlying causes of death coding examples” (156 codes) in the instruction manual of the International Statistical Classification of Diseases and Related Health Problems, 10th revision. Next, we used Iris to code the underlying causes of death and calculate the accuracy. [Results and Discussion] As a result, the accuracy of the underlying cause of death coding by Iris was 81.4%. The causes of incorrect answers were differences in policy from the standardized disease-code master, the fundamental difficulty in assigning an ICD-10 code, and the instruction manual rules that Iris did not support. Since we used the original cause-of-mortality coding example, which contains many exceptional cases, it is suggested that the accuracy would be higher if we were to use actual data. [Conclusion] It is suggested that Iris could be a replacement for the Japanese auto-coding system. In the future, we plan to conduct experiments using real data.

Keywords: underlying cause of death, Vital Statistics, Iris, ICD-10

1. 緒論

死亡統計は保健に関連する重要な資料であり、その製表には原死因が世界的に使用される。原死因は、死亡診断書中に記載される傷病名のうち「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又は損傷」または「致命傷を負わせた事故又は暴力の状況」を指す¹⁾。死亡診断書に複数の傷病名が記載されている場合は、WHO が定める原死因選択ルールに則り、適切に原死因を選択する必要がある。

現在、日本の死亡統計を担当する厚生労働省では、死亡届と死亡診断書からなる人口動態調査死亡票(以下、死亡票)から原死因確定を行っている。その際、年間 130 万件以上の死亡票を処理するため、独自開発した人口動態死因オートコーディングシステム(以下、オートコーディングシステム)が使用されている。オートコーディングシステムの構成は、死亡票の「死亡の原因」の I 欄と II 欄(以下、I 欄 II 欄)にある傷病名に対する ICD-10 コードの付与と、原死因選択ルールに則った原死因決定の二段階である。ただし、傷病名に対する ICD-10 コード付与が困難である場合や I 欄 II 欄以外の項目(手術欄や「その他特に付言すべきことがら」など。以下、付帯情報)に何かしら記載がある場合は、職員による原死因確定作業に切り替えるためにエラーを返す仕組みが存在する。この死亡票の取得から職員による原死因確定作業までが、現行の原死因確定プロセスであり、エラーがなかった死亡票、つまりオートコーディングシステムのみで原死因確定される死亡票は全体の約 6 割にあたる。一方、残りの 4 割は、主に付帯情報に記載される内容解釈が原死因確定に影響を及ぼし

ていることがこれまでの調査により明らかになっている²⁾。日本における年間死亡数は年々増加傾向にあり、職員の作業件数も今後増加することが予想され、より正確で効率のよい原死因確定が求められている。

本研究を含む「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」では、原死因確定プロセスにおける人手作業の効率化のための機械学習の適用を目指している。中でも、付帯情報による原死因の変更の有無を機械学習することにより、人手作業が大幅に効率化できると考える。そのため、①死亡票データ、②オートコーディングシステムが付与した I 欄 II 欄記載の傷病名に対する ICD-10 コード、③オートコーディングシステムが原死因選択ルールに則り決定した原死因の ICD-10 コード、④最終的に確定された原死因の ICD-10 コードを用いた実験を計画したが、オートコーディングシステムは非公開であり、②および③の利用ができない。③は、原死因変更の有無を評価する上でとくに重要な情報であり、研究を進めるためにも入手する必要がある。

そこで、本研究では、オートコーディングシステムを再構築して③を得るのではなく、世界でも利用が広がりつつあるオートコーディングシステムに類似したフリーソフトウェア Iris³⁾を用いることによって疑似ではあるが③のデータの確保を試みる。また、世界は ICD-11 の導入に向けて動いており、日本でもそれに伴いオートコーディングシステムに大幅な修正を加える可能性が出ているため、Iris がオートコーディングシステムの代替として国内導入可能かどうかを調査する。

2. 関連研究

2.1 国内における原死因に関する議論

原死因を調査することは、公衆衛生の目的でもある死亡の防止につながる。原死因を用いた研究として、要介護・要支援認定時データと死亡票データを突合し、死亡をエンドポイントすることで疑似的なコホート研究³⁾を行った例や、東日本大震災後の自殺の超過死亡についての調査⁵⁾した例などがある。一方で、原死因確定の原点となる死亡診断書について、医師の経験年数や書き方などにより原死因の正確性に影響が出ることを示唆する研究⁶⁾も存在しており、医師の教育や診療情報管理士の導入の重要性が議論されている。

また、原死因確定プロセスに関する研究について、平成 5 年から平成 7 年に厚生労働省が開発した原死因選択システム ACSEL⁸⁾についての報告がある。ACSEL は、ICD-9 から ICD-10 の導入に伴い開発され、その後、ICD-10 の改定毎にシステムの改修し、現在のオートコーディングシステムになったとされる。ICD-10 の改定や ICD-11 導入に向けてシステムに言及する審議会は存在するが、職員作業の効率化や機械学習の適用という点では、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」による研究²⁾が起点であると考えられる。本研究を含む「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」はその延長に位置づけられる。

2.2 Iris

Iris は、欧米の保健機関を中心とした IRIS INSTITUTE により開発され、現在 20 を超える国が利用しており、その多くは原死因確定プロセスの一部として本格的に導入している³⁾。日本も利用国の 1 つとされており、国立がん研究センターが全国がん登録データベース作成時に Iris 導入を検討していた記録⁹⁾と社会保障審議会統計分科会疾病、傷害及び死因分類専門委員会がインストラクションマニュアルの改定時に Iris を参照している記録¹⁰⁾が存在しているが、現在の利用は不明である。

Iris の導入に関する研究では、イングランドとウェールズが死因の基本分類で原死因の精度を調査した研究¹¹⁾があり、95%の精度が報告されている。同様に、スペインのナバラでも精度に関する研究¹²⁾が行われており、死因 ICD-10 コードが完全一致していたものは 79.1%、基本分類が一致していたものは 92.0%であることが述べられている。また、Iris に入力する自由記載の傷病名を ICD-10 コードに変換するための処理に関するドイツの研究¹³⁾では、作成した医療用語変換モジュールにより対象の傷病名全体の 92.92%の ICD-10 コード付与が可能になったと報告している。他にも、Iris による原死因選択により従来よりも増加する原死因について調査するオランダの研究¹⁴⁾や、Iris をポルトガル語で使用するための辞書の作成と評価を行うブラジルの研究¹⁵⁾、Iris を導入するためのシステム開発と Iris による原死因選択の適用可能性について述べたブルキナファソの研究¹⁶⁾がある。

すでに Iris を導入している国では、Iris により選択された原死因をもとに研究を行っている。例えば、イタリアでは ICD-10 のバージョンの違いによる原死因の変化について調査¹⁷⁾している。また、フランスでは人工知能により原死因選択を行う研究¹⁸⁾に取り組んでおり、Iris を比較対象としている。

以上のように海外では、Iris に関連する研究や Iris を用い

表 1 原死因付与ルール

ルール	内容
Step0 記載のまま標準病名マスターを適用	
a	標準病名マスターを参照し、 ・文字列が完全一致するもの ・ICD-10 コードが一つに決まるもの
Step1 文字列処理	
b	助詞、接続詞、空白などの区切り文字の除去と展開
c	言い換え
c1	“悪性新生物”を“がん”とする
c2	“<部位>原発”あるいは“<部位>(における)続発症”の記載と、その前後の行にがんの記載がある場合、“原発”あるいは“続発症”を“がん”とする
c3	がんの記載がある場合、“続発性”を“転移性”とする
d	形態素的な分割要素で標準病名マスターを参照し、 ・入れ替え・部分要素で完全一致したもの ・ICD-10 コードが一つに決まるもの
Step2 人為的判断	
x	Step1 までで ICD-10 コードが付与できなかったもの

た原死因の研究が進んでいるが、日本では詳細な研究が進んでおらず、本研究がその先駆けと考える。

3. 方法

死亡票の代替として「疾病、傷害および死因統計分類提要 ICD-10 準拠第 2 巻インストラクションマニュアル<総論>」¹⁾(以下、インストラクションマニュアル)中の、原死因コーディング例(I 欄 II 欄に記載される傷病名と原死因が明示されているもの)を対象に Iris を用いて原死因の付与とその精度を評価する。ここで、オートコーディングシステムはインストラクションマニュアルを参考に作られており、原死因コーディング例に記載される原死因がオートコーディングシステムの挙動による結果とみなすことができるとする。一方、Iris は日本語に非対応であり、傷病名ではなくそれに対応する ICD-10 コードを読み込ませる必要があるため、I 欄 II 欄に記載される傷病名に対する ICD-10 コードは 3.1 節に述べる方法で付与する。

3.1 ICD-10 コードの付与

インストラクションマニュアルの原死因コーディング例中の I 欄 II 欄に記載される傷病名に対して、以下を用いて人手で ICD-10 コードを付与する。

- 病名くん 2.0(ICD10-2013) Rel.2.2.3.0¹⁹⁾
(以下、標準病名マスター)
- 疾病、傷害および死因統計分類提要 ICD-10 準拠第 1 巻内容例示表 (以下、内容例示表)²⁰⁾
- 疾病、傷害および死因統計分類提要 ICD-10 準拠第 3 巻索引 (以下、索引)²¹⁾

また、今後の研究で Iris を使う場合、ICD-10 コードの付与作業が本研究同様必要である。そのため、大量の死亡票データに ICD-10 コードを自動付与可能にするルール作りや機械実装に今後取り組むことを想定し、どのようなルールが必

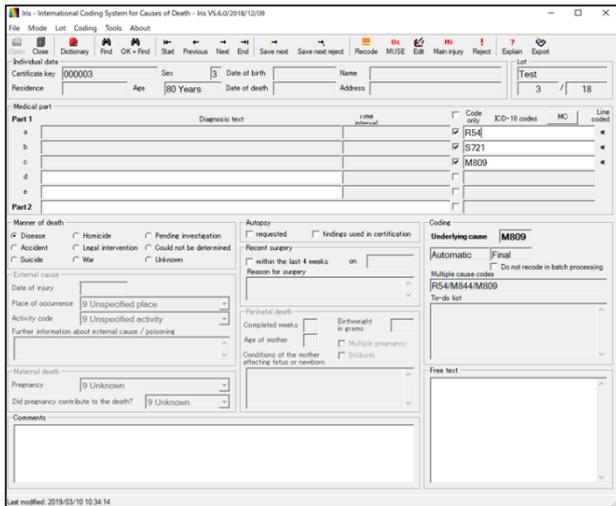


図 1 Iris のメイン画面

要であるが、表 1 に示す仮の ICD-10 コード付与ルールを設定し、適用したルールを記載しながら作業を行う。

表 1 に関して、Step0 では、傷病名を記載のまま標準病名マスターに適用する。ただし、インストラクションマニュアルには、通常の死亡票には載っていないと考えられる独自の記載があるため、以下の記載はあらかじめ処理する。

- 傷病名と期間の併記
→ 期間は傷病名欄では考慮しない
- ()、< >、[] の記載
 - ◇ () : 補足的な語句を囲む
 - ◇ < > : 異なった表現がある場合
 - ◇ [] : 同義語、代替的語又は説明的な句を囲む
→ 記載に従い書き直し、再度 Step0 を適用する

例えば、“高血圧(症)”と記載がある場合は、高血圧あるいは高血圧症として、標準病名マスターを参照し ICD-10 コードを付与する。

Step1 では、いくつかの文字列処理を行う。ただし、b~d までのルール適用の順番は考慮しないものとする。ルール b では、助詞、接続詞の除去と展開を行う。例として、除去では“肝臓のがん”を“肝臓がん”とし、展開では“肺と肝臓のがん”を“肺がん”と“肝臓がん”とすることで ICD-10 コードを付与する。ルール c では、記載内容の言い換えを行う。例として、ルール c1 では“結腸の悪性新生物”を“結腸がん”とし、ルール c2 では“腎原発”を“腎がん”とし、ルール c3 では“肺の続発性癌”を“肺の転移性癌”とすることで ICD-10 コードを付与する。ルール d では、傷病名を形態素的な分割要素に分けて標準病名マスターに適用する。例として、入れ替えでは“腎萎縮”に対して“腎”と“萎縮”の分割要素で標準病名マスターに適用し“萎縮腎”の ICD-10 コードを付与する。また、部分要素では“慢性リウマチ性心疾患”を“慢性”と“リウマチ性”と“心疾患”の分割要素にしたのち、分割要素の多い順から組み合わせを行い標準病名マスターに適用し“リウマチ性心疾患”の ICD-10 コードを付与する。

Step2 では、上記のルールを適用しても ICD-10 コードが付与できなかったものに関して、内容例示表や索引を用いて人為的判断で ICD-10 コードを付与する。

3.2 Iris による原死因コーディング

表 2 I 欄 II 欄全ての傷病名に対し ICD-10 コードが付与できた例の件数・割合 (Step 別)

	件数	割合 (/156)
step0	34	21.8%
step1	76	48.7%
step2	156	100.0%

表 3 I 欄 II 欄全ての傷病名に対し ICD-10 コードが付与できた例の件数・割合 (ルール別)

ルール	件数	割合 (/156)
a	34	21.7%
言語処理	64	41.0%
b	60	38.5%
c	10	6.4%
c1	9	5.8%
c2	1	0.6%
c3	2	1.3%
d	23	14.7%
x	80	51.3%

使用する Iris のバージョンは Iris Version 5.6.0-Y2019S1 であり、原死因コーディング部分を行う MUSE のバージョンは MUSE 2.7 である²²⁾。

Iris による原死因コーディングを行うため、まず、図 1 に示す Iris のメイン画面を起動し各例の ICD-10 コードを 1 件ずつ手作業で入力する。ここで、直接 ICD-10 コードを入力するためには対象者情報(年齢と性別)が記載されたロットを用意する必要がある。そのため、年齢・性別に関して、I 欄 II 欄に記載される傷病名との矛盾がないように入力を行う。また、以下の ICD-10 コードに関しては、Iris が非対応のため別途変換を行う。

- 糖尿病: E14→E149
- 呼吸不全: J9699→J969

全入力が終了したのち、Iris のバッチ(一括)処理機能を利用し原死因コーディングを行う。

3.3 Iris による原死因コーディング精度の評価

原死因コーディング例に記載される原死因を正解データとしたとき、Iris による原死因コーディングの結果が全件中何件正解データと完全一致するか、その精度を算出する。

3.4 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

Category	Number	Coding success
Certificates	156	
Processed	156	100.0%
Rejected	11	7.7%
Syntax	0	0.0%
Code	2	1.3%
Interval	0	0.0%
Multiple cause	0	0.0%
MayBe	9	5.8%
Coder	0	0.0%
MainInjury	0	0.0%
Final	145	92.9%
Not processed	0	
Empty	0	
Closed	0	
Manual	0	
Non initial	0	
'Do not recorded' certificat	0	
Terms processed	0	-
Icd codes selected	0	-
Non-recognized	0	-

図 2 Iris による原死因コーディング結果

等の遵守に努めた。

4. 結果

インストラクションマニュアル中の原死因コーディング例は 156 件であった。各例に適用した ICD-10 コード付与ルールの Step 別の結果を表 2 に示す。各ルール別の結果を表 3 に示す。

また、Iris による原死因コーディングを行った結果を図 2 に示す。図 2 より、全 156 件中、

(ア)原死因が付与されたデータ :145 件(約 92.9%)

(イ)“Rejected Code”と表示されたデータ :2 件(約 1.3%)

(ウ)“Rejected Maybe”と表示されたデータ :9 件(約 5.8%)

原死因コーディング例との比較を行った結果、原死因コーディング例の原死因を正解データとしたとき、(ア)に対しては、

正解 117 件:(/145 ≒ 80.7%)

不正解 28 件:(/145 ≒ 19.3%)

(イ)に対しては、

正解 2 件:(/2 ≒ 100.0%)

不正解 0 件:(/2 ≒ 0.0%)

また、(ウ)に対して確認を行ったうえで比較を行うと、

正解 8 件:(/9 ≒ 88.9%)

不正解 1 件:(/9 ≒ 11.1%)

であった。したがって、全 156 件に対しては、

正解 127 件:(/156 ≒ 81.4%)

不正解 29 件:(/156 ≒ 18.6%)

であった。また、ICD-10 コード付与ルールの Step 別に正解をまとめたものを表 4 に示す。また、不正解データの内訳は、

- 悪性新生物に関する記載の解釈問題 (13/29 件)
- 及び と 又は の区別の不備 (8/29 件)
- その他 (8/29 件)

であった。

5. 考察

表 2 より、ICD-10 コードの付与に関して、特別な処理をせず原死因コーディング例中の I 欄 II 欄に記載される傷病名

表 4 Iris による原死因の正解件数と割合 (Step 別)

	件数	一致数	割合 (/各件数)	割合 (/156)
step0	34	30	88.2%	19.2%
step1	76	62	81.6%	39.7%
step2	156	127	81.4%	81.4%

すべてに ICD-10 コードが可能なものは、約 2 割程度であり、Step1 の文字列処理を行うことで ICD-10 コードが可能なものは、約 5 割程度に増加することがわかった。オートコーディングシステムでは、全体の約 98%の傷病名に対して ICD-10 コードが自動付与されており²⁾、4 章の実験と比較して非常に高い付与率である。要因としては、オートコーディングシステムには記載内容に対する詳細な辞書やルールベースが存在していること、インストラクションマニュアル中の原死因コーディング例よりも実際の死亡票データが単純な可能性があることが考えられる。

表 3 より、step1、中でもルール b(助詞、接続詞、空白などの区切り文字の除去と展開)の処理が最も重要であることがわかった。ただし、今回の Step1 に挙げた文字列処理を機械実装するためには、いくつか高度な処理が含まれる。例えば、ルール b では、記載内容に対して展開が必要かどうかの判断をする処理や記載の意図に合った展開処理が必要である。ルール d では、形態素数が多いほど多くの組み合わせが存在するため、適切な ICD-10 コードを選択する判断処理部分の設計が重要となってくる。また、ルール c に関しては、インストラクションマニュアル固有のパターンがある場合もあり、実データに即して適宜ルールの変更や追加を行う必要がある。一方、Step2 の中でも、複数候補から適切な ICD-10 コードを選択するもの、ICD-10 コードの 5 桁目分類を適切に選択するものに関しては、自動化可能であると考えられる。

図 2 より、“Rejected Code”に関しては、シャーガス病 (B57.2) のような稀な傷病名に対して表示されるということがわかった。このほかにも、対象者の年齢や性別と傷病名の矛盾がある場合にも表示がされる。また、“Rejected Maybe”に関しては、付与された原死因に対して何かしらの疑義がある場合に表示されるということがわかった。いずれにしても、原死因のコーディングはされているため、対象者と傷病名の矛盾なく ICD-10 コードを与えれば、Iris はすべてのデータに対して原死因を選択可能であることが明らかになった。

Iris による原死因コーディング精度に関しては、約 8 割の原死因が完全一致で選択できることがわかった。これは、スペインのナバラでの研究結果(79.1%)¹¹⁾と比較しても、妥当と考えられる。また、“Rejected Code”や“Rejected Maybe”に関して、これらの表示がある場合でも原死因選択の精度が高いことが示された。

表 4 より、Step0 における一致率ももっとも高いことが分かった。これは、Step0 に該当する原死因コーディング例の I 欄 II 欄の記載が単純であり、選択する原死因も単純になっている場合が多いためだと考えられる。

一方、29 件の不正解のデータの内訳より、悪性新生物に関する死因記載の解釈に問題がある場合が最も多いことがわかった。実際、インストラクションマニュアルには悪性新生物に関するルールが多く存在し、解釈が複雑である。また、

全不正解データの詳細を考察した結果、主に以下あるいはその組み合わせにより生じる原因があることが示唆された。

- I. 標準病名マスターとのポリシーの相違
- II. 根本的に傷病名に対し ICD-10 コードの付与が困難
- III. Iris 未対応のインストラクションマニュアルのルールに則った原死因の決定が必要

Iは、肝臓がんの ICD-10 コード付与に関する問題である。インストラクションマニュアルでは、悪性新生物に関連するルールが詳細に設定されている。その中に転移好発部位リストと呼ばれるリストが存在し、原発部位が記載されていないがんの場合、リストに該当する部位のがんは転移性とみなすルールがある。具体的には、「骨、脳、横隔膜、不明な部位(C76に分類可能な部位)、肝臓、肺(特殊な指示も参照)、リンパ節(特殊な指示も参照)、縦隔、髄膜、腹膜、胸膜、後腹膜、脊髄」の部位が該当する。しかし、実験の結果、肝臓がんに関しては原発とみなされ不正扱いになる事例が複数あった。実験では、肝臓がんの ICD-10 コードは標準病名マスターを参照し C22.0 を入力としていた。Iris では C22.0 に対する処理はなく、C22.9(肝の悪性新生物<腫瘍>、詳細不明)に対して転移好発部位リストの処理を行っており、このことから不正解が生じたと考えられる。このように、標準病名マスターとのポリシーの相違が原死因確定に影響を与える場合があることがわかった。

II. に関して、「塩化ビニルへの長期間の暴露」など、自由記述の傷病名に対し ICD-10 コードの付与を行うあるいはほかの欄との関係から傷病名を推察して ICD-10 コードを付与するが場合があった。これに関しては、オートコーディングシステムの中でも処理が困難であると考えられる。また、「気管支の癌(腫)」と「乳房の癌(腫)」が入力された場合、インストラクションマニュアルでは乳房の癌を選択するよう指示があるが、気管支の癌の ICD-10 コードが肺がんと同様の ICD-10 コード(C34.9)のため、Iris では転移好発部位として扱われ原死因として選択されないという例があった。このように、ICD-10 コードの粒度の問題で原死因が異なる場合も存在した。

III. に関して、対応が必要だと考えられるルールは 2 つある。1 つは、「及び」と「又は」の区別を行うことである。複数の病名や病態が 1 行中に記載される場合、その意味合いは「A 及び B が該当する(A and B)」あるいは「A 又は B が該当する(A or B)」である。しかし、Iris ではシステム上“and”または“or”を区別することができず、どちらに対しても“and”の意味で解釈する。例えば、「腎臓又は膀胱のがん」の記載に対し、インストラクションマニュアルでは腎臓がん(C64)と膀胱がん(C67.9)の組み合わせから尿路の悪性新生物(C68.9)の ICD-10 コードを付与するよう指示があるが、Iris にはこの機能がついていない。もう 1 つは、転移好発部位リストにない部位に「転移性」が記載されている場合、その部位を原発とみなすことである。Iris では、一度転移性で ICD-10 コード付与したがんは原死因選択時に原発として再コードする仕組みが存在しない。そのため、ICD-10 コードを付与する際には転移性の ICD-10 コードは用いないなどの対策が必要であると考えられる。

その他に、Iris が参照するインストラクションマニュアルのバージョンが最新版であるため、2013 年版のインストラクションマニュアルを参照するオートコーディングシステムと異なる原死因を付与していると考えられる場合も存在するが、核心的な原因は不明である。しかし、不正解データ全体を通して、Iris の挙動を考慮した ICD-10 コード付与が必要であり、対策が可能になればさらに精度は上がると考えられる。

また今回は、特殊な例を多く含むインストラクションマニ

ュアルを対象に原死因の精度の評価を行っているが、表 2 の考察同様、実際の死亡票データでは記載内容がもっと単純な場合も想定できる。したがって、死亡票の実データを用いたさらなる検証が必要である。ただしその結果、原死因精度が著しく低下する可能性は低く、Iris をオートコーディングシステムの代替として用いても問題はないと考えられる。

6. 結論

本研究では、日本の原死因確定プロセスにおけるオートコーディングシステムの代替としての、Iris の導入可能性について調査を行った。その結果、傷病名に ICD-10 コードが付与できれば、81.4%の精度で原死因確定が可能であり、Iris の挙動を考慮した ICD-10 コードを付与するとさらに精度が上がるということが示唆された。また、インストラクションマニュアル中の原死因例には特殊な例が含まれていたため、死亡票の実データを用いた場合は、より精度が高くなる可能性も述べた。

今後は、実データを用いてさらに本研究の検証を進めること、傷病名に対する ICD-10 コード付与を行う機械実装用の ICD-10 コード付与ルールをさらに明確化することに取り組む。そのうえで、実データにルールを適用し、より効率的に原死因データを収集する。

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)」の一環として実施したものである。

利益相反: 開示すべき利益相反はない。

参考文献

- 1) 厚生労働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 2 巻 Instruction manual(総論). 厚生労働統計協会, 2016.
- 2) 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. 厚生指 2020 ; 67(3) :17-24.
- 3) Iris Institute. Federal Institute for Drugs and Medical Devices. [https://www.dimdi.de/dynamic/en/classifications/iris-institute/index.html (cited 2020-Aug-28)]
- 4) 三徳和子, 伊藤弘人, 後藤忠雄, 尾形由起子, 眞崎直子. 要介護高齢者の 10 年転帰と医療機関以外での死亡に関するコホート研究. 日本医療・病院管理学会誌 2018 ; 55(2) :59-69.
- 5) 眞崎直子, 橋本修二, 川戸美由紀. 人口動態統計に基づく東日本大震災後の自殺死亡数: 岩手県・宮城県・福島県の沿岸部と沿岸部以外の推移. 日本公衆衛生雑誌 2018 ; 65(4) :164-169.
- 6) 高橋文, 塚本哲, 阿部昌彦. 死亡診断書に基づく原死因の精度に連動する因子の解析. 診療情報管理 2013 ; 25 :50-55.
- 7) 柏倉夏枝, 片桐茂. 死亡診断書における原死因の精度および記載内容に関する検討. 診療情報管理 2014 ; 26(1) :70-73.
- 8) 竹村和子, 上田響, 中田正. 人口動態統計における新しい原死因選択システム(ACSEL)について. 厚生指 1996 ; 43(10) :9-16.
- 9) WHO 国際統計分類協力センター. 年次報告書(2014). WHO 国際統計分類協力センター, 2014. [http://who-fic-japan.jp/img/events_attended/pdf/report2014_j

- .pdf (cited 2020-Aug-28)].
- 10) 社会保障審議会統計分科会疾病、傷害及び死因分類専門委員会死因選択検討ワーキンググループ. 死因選択検討ワーキンググループにおける検討状況の報告. 厚生労働省, 2016. [https://www.mhlw.go.jp/file/05-Shingikai-12601000-Seisakuto-ukatsukan-Sanjikanshitsu_Shakaihoshoutantou/0000112550.pdf (cited 2020-Aug-28)].
 - 11) Wells C. Impact of the Implementation of IRIS Software for ICD-10 Cause of Death Coding on Mortality Statistics, England and Wales. UK: Office for National Statistics, 2014. [<https://www.ons.gov.uk/peoplepopulationandcommunity/births-deathsandmarriages/deaths/bulletins/impactoftheimplementationofirissoftwareforicd10causeofdeathcodingonmortalitystatisticsenglandandwales/2014-08-08> (cited 2020-Aug-28)].
 - 12) Yugo Floristán Floristán, Josu Delfrade Osinaga, Jesus Carrillo Prieto, Jesus Aguirre Perez, Conchi Moreno-Iribas. Coding Causes of Death with IRIS Software. Impact in Navarre Mortality Statistic. *Rev Esp Salud Publica* 2016 ; 90 : e1-9.
 - 13) Olaf Eckert. Electronic coding of death certificates. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2019 ; 62(12) : 796-805.
 - 14) P P M Harteloh. Changes in cause-of-death statistics as a consequence of automated coding. *Ned Tijdschr Geneesk* 2017 ; 161 : D1767.
 - 15) Renata Cristófani Martins, Cassia Maria Buchalla. Automatic coding and selection of causes of death: an adaptation of Iris software for using in Brazil. *Rev Bras Epidemiol* 2015 ; 18(4) : 883-893.
 - 16) Seydou Golo Barro, Grégoire Rey, Pascal Staccin. Study of the Usability of an Automated Coding Software for Causes of Death in an African Context. *Stud Health Technol Inform* 2019 ; 264 : 1978-1979.
 - 17) Orsi C, Navarra S, Frova L, et al. Impact of the implementation of ICD-10 2016 version and Iris software on mortality statistics in Italy. *Epidemiologia e Prevenzione* 2019 ; 43(2-3) : 161-170.
 - 18) Louis Falissard, Claire Morgand, Sylvie Roussel. A Deep Artificial Neural Network-Based Model for Prediction of Underlying Cause of Death From Death Certificates: Algorithm Development and Validation. *JMIR Med Inform* 2020 ; 8(4) : e17125.
 - 19) 標準病名マスター作業班. 病名くん 2.0. 標準病名マスター作業班. [<http://www.byomei.org/wg/index.html> (cited 2020-Aug-28)].
 - 20) 厚生労働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 Tabular list (内容例示表). 厚生労働統計協会, 2016.
 - 21) 厚生労働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 3 巻 Alphabetical index (索引表). 厚生労働統計協会, 2016.