

別添資料 2

インストラクションマニュアル事例を対象とした

Iris による原死因確定実験

概要

「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>」に記載されている原死因コーディング事例を対象に、Iris にて原死因確定を行い、結果を比較した。Iris では入力として病名の ICD-10 コーディング結果を用いるが、標準病名マスターを用いても、自由記載の病名に対し ICD-10 コーディングを完全に自動で行うことは難しい。そこで、原文記載まま (Step0)、文字列処理後 (Step1)、人手で判断を要する ICD-10 コーディング (Step2) の3段階に分け、それぞれ Iris での原死因確定精度を評価した。結果として Step0 で約 2 割、Step1 までで約 5 割について「全病名に ICD-10 コード付与可能」であり、Step0~2 の処理結果のいずれに対しても、Iris にて約 8 割以上の精度で原死因が確定できることが判明した。

はじめに

本研究全体では、原死因確定プロセスにおける人手チェックの効率化のための機械学習の適用を目的としている。我が国ではオートコーディングツールによって確定された仮原死因に対し、死亡票に何らかの付帯情報があった場合、人手チェックにより最終的な原死因の確定が行われている。ここで機械学習の援用により、原死因の変更の有無を予め高精度に予測することができれば、人手チェックを大幅に効率化することができる。

しかし、そのためには、WHO が定める原死因確定ルールに従い、仮の原死因を確定する必要がある。現状では厚生労働省内で実際に用いられているオートコーディングツールを用いることはできないため、この挙動を模すツールとして、世界的に利用されているオートコーディングツールである Iris をその代替として用いることを想定している。

本実験では、統計法 33 条により死亡票・死亡個票の調査票実データの提供を受ける前に行った、Iris が日本での原死因確定に適用可能か、についての予備調査実験について述べる。「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)中の原死因コーディング例(I 欄 II 欄中の病名+選択すべき原死因の傷病名)を対象に調査を行う。

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には複雑・イレギュラーなケースが多く含まれるが、ICD-10 コードのみを使用して原死因コーディングを行う Iris を使用するためには ICD-10 コードの付与の正確性が非常に重要である。ただし、全例の ICD-

10 コード付与の正確性を人手により担保するのではなく、部分的に機械による自動化を行って担保することで原死因コーディングの効率化を図るべきである。

以上より、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もとに、人手ではなく機械的な処理でICD-10コード付与を可能にする、機械実装を想定したルールを暫定的に示す。そして、そのルールをもとに実験を行い、ルールの有用性や改善点を調査する。

ICD-10コード付与ルール

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の内容を調査し、ルールを作成した(表1)。ICD-10コードを付与するために参考にした資料は以下の通りである。

- ・病名くん 2.0(ICD10-2013) Rel.2.2.3.0
(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)
- ・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)

ただし、Step1までは標準病名マスターのみを使用し、Step2で内容例示表と索引を使用する。

表1 ICD-10コード付与ルール

| | 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 |
| | b | 助詞、接続詞、空白などの区切り文字の除去と展開 |
| Step1 ～文字列処理～ | 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 |

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ものとする。

- 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（1巻 p11）
 - ◇ []：同義語、代替的語又は説明的な句を囲むため
- 記載に従い書き直し、再度 Step0 を適用する

例：高血圧（症） → 高血圧あるいは高血圧症で標準病名マスターに適用

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 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 リウマチ性 心疾患”の分割要素にしたのち分割要素の多い順からの組み合わせで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を人手で行う。

実験

ICD-10 コード付与ルールをもとに、再度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の傷病名に対して ICD-10 コード付与を行い、その結果をもとに Iris で原死因コーディングを行う。

方法

手順1. ICD-10 コード付与ルールを用いて ICD-10 コードを付与

ICD-10 コード付与ルールに基づき、手作業で ICD-10 コードを付与する。また、ICD-10 コード付与時にどのルールを適用したか記録を行う。

| 例 | ICD-10 | ルール | 例全体の使用ルール |
|--------------|--------|------|-----------|
| I 欄 a：肺の続発性癌 | C780 | b,c3 | b,c2,c3 |
| II 欄：腎原発 | C64 | c2 | |

手順2. Irisによる原死因コーディング

実験で用いる 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。

Iris による原死因コーディングを行うため、まず、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 欄に記載されている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 が対応していないため変換を行う。

- 糖尿病（3 件）：E14→E149
- 呼吸不全（4 件）：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。

表 2 I II 欄全ての傷病名に ICD-10 コード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% |

| ルール | | 件数 | 割合 (/156) |
|------|------------------------------------------------|----|-----------|
| a | 原死因コーディング例の記載をそのまま | 34 | 21.7% |
| 言語処理 | | 64 | 41.0% |
| b | 助詞、接続詞、空白などの区切り文字の除去と展開 | 60 | 38.5% |
| c | 言い換え | 10 | 6.4% |
| c1 | 悪性新生物→がん | 9 | 5.8% |
| c2 | <部位>原発・<部位> (における) 続発症 →前後の行にがんの記載があれば癌をつける | 1 | 0.6% |

| | | | |
|-------|---------|----|-------|
| c3 | 続発性→転移性 | 2 | 1.3% |
| d | 文字列処理 | 23 | 14.7% |
| 人為的判断 | | 80 | 51.3% |

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

手順3. Iris が出力した原死因と原死因コーディング例の原死因を比較
 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一致する
 か、精度を算出する。

結果

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、各ルール別の結果
 を表 3 に示す。また、I II 欄の各行の傷病名に適応したルールの結果を表 4 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 1 に示す。

表 4 I II 欄の各行の傷病名に適用した ICD-10 コード付与ルールの件数・割合

| | | | |
|------|-------------------------|-----|-----------|
| | | 件数 | 割合 (/315) |
| a | 原死因コーディング例の記載をそのまま | 135 | 42.9% |
| 言語処理 | | 75 | 23.8% |
| b | 助詞、接続詞、空白などの区切り文字の除去と展開 | 67 | 21.3% |

| End of Batch processing | | |
|-----------------------------|--------|----------------|
| 2020/07/11 12:30:51 | Number | Coding success |
| Certificates | 156 | |
| Processed | 156 | 100.0% |
| Rejected | 11 | 7.1% |
| 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 recoded' certificat | 0 | |
| Terms processed | 0 | - |
| Icd codes selected | 0 | - |
| Non-recognized | 0 | - |

図 1 Iris 結果画面

| | | | |
|-----------------------|----------------------------------------------|-----|-------|
| c | 言い換え | 11 | 3.5% |
| c1 | 悪性新生物→がん | 9 | 2.9% |
| c2 | <部位>原発・<部位>(における)続発症 →前後の行にがんの記載があれば癌をつける | 1 | 0.3% |
| c3 | 続発性→転移性 | 2 | 0.6% |
| d | 文字列処理 | 26 | 8.3% |
| 選択ルーチン候補・人手で意味を解釈する処理 | | 105 | 33.3% |

図1より、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%)
 - ① 原死因が付与されたデータ : 2件 (約1.3%)
 - ② 原死因が付与されていないデータ : 0件 (約0.0%)
- (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、(ア)に対しては、

- ・正解 117件 : (/145 ≒ 80.7%)
- ・不正解 28件 : (/145 ≒ 19.3%)

(イ)①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5に示す。

表 5 付与した ICD-10 コードを Iris に実装、原死因が一致した件数と割合

| | 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% |

考察

表 2 より、特別な処理をせず I II 欄に記載される全ての傷病名に ICD-10 コードが付与可能である原死因コーディング例は、約 2 割程度であり、Step1 の文字列処理を行うことで約 5 割程度に増加することがわかった。とくに表 3 と表 4 より、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) の処理が重要であると考えられる。ただし、今回の Step1 に挙げた文字列処理を機械実装するには、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変更や追加を行う必要がある。一方、Step2 の中でも、以下の場合については別途な工程 (モジュール) を追加することで、人的判断の数を減らすことができる可能性がある。

- 選択モジュール
 - ◇ 複数の候補から適切な ICD-10 コードを収集する
- 詳細不明モジュール
 - ◇ 5 桁目の分類が必要な場合に適切な ICD-10 コードを付与する

いずれにして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には、複雑・イレギュラーなケースが多く含まれるため、実際の死亡票に対しては各ルールで対応可能な場合が増えると予想され、実データで次回以降検証を行うことにする。

“Rejected Maybe”に関して、結果中の※より、精度が 9 割近いことがわかり、“Rejected Maybe”が付与されていても精度は保証されることが示唆される。

不正解のデータから、考察したことは以下の通りである。

◇ 悪性新生物に関する死因記載の解釈に問題がある (13/29 件)

今回の不正解データの半数は、悪性新生物に関するデータである。その原因は、以下に挙げるものだと考えられる。

i. 肝がん(C22.0)が転移好発部位リストに含まれていない

転移好発部位リストは、以下の通りである。

骨、脳、横隔膜、不明確な部位(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

Iris の MUSE にも転移好発部位リストは組み込まれており、原発性で付与された ICD-10 コードは続発性の ICD-10 コードに訂正される。しかし、肝臓に関しては全件訂正が入らないことから、その理由の詳細を調査する必要がある。ちなみに、今回該当した例を表 6 に示す。

ここで、肝がんの ICD-10 コードは検索により C22.0 を入力している。一方で、表 6 に挙げる例中の C22.0 に対して C22.9 (肝の悪性新生物<腫瘍>、詳細不明) を置換したところ Iris による訂正が入った上で、例 119 以外は正解の原死因に変化した。つまり、転移好発部位リストには C22.9 が登録されていると考えられ、肝がんに ICD-10 コードを付与する際は C22.9 するべきであると

表 7 不正解データ該当例（悪性新生物：肝がん関係）

| 番号 | 例 | ICD | Iris | マニュアル |
|-----|----------------------|-------------------|-------|-------|
| 88 | 例 26 : I (a) 肝がん | C 220 | C 220 | C 189 |
| 88 | (b) 結腸の悪性黒色腫 | C 189 | | |
| 98 | 例 36 : I (a) 肝がん | C 220 | C 220 | C 189 |
| 98 | (b) 結腸がん | C 189 | | |
| 98 | (c) 膀胱がん | C 679 | | |
| 99 | 例 37: I (a) 肝臓がん | C 220 | C 220 | C 189 |
| 99 | (b) 結腸の腺がん | C 189 | | |
| 99 | (c) 大腿の皮膚の悪性黒色腫 | C 437 | | |
| 102 | 例 40 : I (a) 肝臓のがん | C 220 | C 220 | C 349 |
| 102 | (b) 肺癌（腫） | C 349 | | |
| 119 | 例 57 : I (a) 肺の転移性がん | C 780 | C 220 | C 349 |
| 119 | II 胸膜、肝臓及び脳のがん | C 384,C 220,C 719 | | |

表 6 不正解データ該当例（悪性新生物：転移性関係）

| 番号 | 例 | ICD | Iris | マニュアル |
|-----|-------------------------|-------------------|-------|-------|
| 117 | 例 55 : I (a) 前立腺の転移性腺がん | C 798 | C 189 | C 61 |
| 117 | (b) 結腸の転移性腺がん | C 189 | | |
| 118 | 例 56 : I (a) 肺の転移癌（腫） | C 780 | C 809 | C 349 |
| 119 | 例 57 : I (a) 肺の転移性がん | C 780 | C 220 | C 349 |
| 119 | II 胸膜、肝臓及び脳のがん | C 384,C 220,C 719 | | |

考えられる。ただし、例 119 については ii .にてさらに考察を行う。

ii. 転移性で入力した ICD-10 コードが原発性の ICD-10 コードに再付与されない

転移性と書かれた病名に対して、インストラクションマニュアルでは、

- 転移好発部位リストにない部位が「転移性」又は「～の転移性」として特定されている場合、その部位を原発とみなし、その部位の原発性悪性新生物<腫瘍>にコードする。
- 記載されている唯一の悪性新生物<腫瘍>が肺の「転移性」新生物<腫瘍>である場合、肺の原発性新生物<腫瘍>にコードする。また、死亡診断書に記載されているその他の新生物<腫瘍>が、いずれも転移好発部位リストにある場合も、肺の「転移性」新生物<腫瘍>を原発性と考える。別の悪性新生物<腫瘍>が死亡診断書に記載され、その悪性新生物<腫瘍>の部位が転移好発部位リストにない場合、肺を続発と考える。
- 形態型が C40-C47、C49 又は C70-C72 に分類できるもので、死亡診断書に記載されている部位が同じ種類の組織を示す場合、その形態型の適した再分類項目にコードする。

表 8 不正解データ該当例（悪性新生物：リスクのある病態関係）

| 番号 | 例 | ICD | Iris | マニュアル |
|----|------------------------------|-----------|------|-------|
| 73 | 例 11： I (a) 肝臓及び肺のがん | C220,C349 | C220 | C229 |
| 73 | (b) 慢性肝炎 | K739 | | |
| 74 | 例 12： I (a) 肺がん | C349 | C220 | C229 |
| 74 | (b) 肝がん | C220 | | |
| 74 | (c) 塩化ビニルへの長期間の暴露 | Z581 | | |
| 77 | 例 15： I (a) 縦隔及び肝臓の悪性新生物<腫瘍> | C383,C220 | C220 | C383 |
| 77 | (b) アスベスト粉じん<塵>の長期間の吸引 | T578 | | |

と記載されている。これをふまえ、不正解データをまとめたものを表 7 に示す。

例 119 では、i. の訂正から、原死因が C80.9 に変化した。このように、転移性の ICD-10 コードをそのまま付与した場合、表 6 の大部分の原死因が原発部位不明(C80.9)になってしまう。

ここで、表 7 の転移性部分を原発性の ICD-10 コードに変え入力を行ったところ、全て正解になった。したがって、ICD-10 付与時に各行を読み取り、インストラクションマニュアルのルールを適応させ、必要に応じて原発性の ICD-10 コードを再付与すれば、ii. は解決する可能性がある。

iii. 悪性新生物<腫瘍>のリスクを高める病態との組み合わせができない

インストラクションマニュアルによると、

- 悪性新生物<腫瘍>が一般にその部位の悪性新生物<腫瘍>のリスクを高めると考えられる病態によって引き起こされたとの記載がある場合、その新生物<腫瘍>を原発としてコードする。その部位が転移好発部位リストにあるとしてもこれが適用される。

と記載されている。表 8 に不正解になった例を示す。

ここで、i. より肝がんを C22.9 に変更すると例 73 は原死因が C22.9 になった。ただし、例 74 は原死因が C34.9、例 77 は原死因が C80.9 と転移好発部位リストのルールに則った結果になった。考えられることとしては、現在付与している ICD-10 コード（塩化ビニル中毒、じん<塵>汚染）の誤りである。アスベスト粉じん<塵>の長期間の吸引に関しては、石綿肺（J61）で ICD-10 コードを付与すると正解の原死因の ICD-10 コードが付与された。このように、索引できなかった病態に対する ICD-10 コードの付与精度を高める必要がある。iii. の場合は、方法として、WHO のウェブサイトにも悪性新生物<腫瘍>のリスクを高める病態の情報があるとインストラクションマニュアルに記載されていたため、調査できれば精度が上がる可能性がある。

表 10 不正解データ該当例（悪性新生物：その他）

| 番号 | 例 | ICD | Iris | マニュアル |
|-----|-------------------------|-----------|------|-------|
| 68 | 例 6：I (a) 膀胱の移行上皮癌 | C679 | C679 | C64 |
| 68 | II 移行上皮癌、腎原発 | D099.C64 | | |
| 103 | 例 41：I (a) 気管支の癌（腫） | C349 | C509 | C349 |
| 103 | (b) 乳房の癌（腫） | C509 | | |
| 112 | 例 50：I (a) 肝臓から脳への転移性がん | C220.C793 | C220 | C229 |
| 112 | II 食道がん | C159 | | |

表 9 不正解データ該当例（“及び”と“又は”の区別）

| 番号 | 例 | ICD | Iris | マニュアル |
|-----|-----------------------|-----------|------|-------|
| 145 | 例 1：I (a) 腎臓又は膀胱のがん | C64.C679 | C64 | C689 |
| 146 | 例 2：I (a) 副腎又は腎臓のがん | C749.C64 | C749 | C809 |
| 147 | 例 3：I (a) 冠動脈硬化又は冠動脈瘤 | I251.I254 | I251 | I259 |
| 148 | 例 4：I (a) 心筋梗塞又は冠動脈瘤 | I219.I254 | I219 | I259 |
| 149 | 例 5：I (a) 肺の結核又はがん | A162.C349 | A162 | J984 |
| 150 | 例 6：I (a) 脳卒中又は心臓発作 | I64.I519 | I64 | I99 |
| 151 | 例 7：I (a) 胆嚢仙痛又は冠動脈血栓 | K802.I240 | K802 | R688 |
| 152 | 例 8：I (a) 冠状閉塞又は戦傷 | I240.Y369 | I219 | R99 |

iv. その他

癌に関する ICD-10 コードの付与に関する問題で、考察が困難で引き続き調査が必要な不正解データは表 9 の通りである。

例 68 に関しては、腎原発にも関わらず腎ではなく膀胱の移行上皮癌が原発扱いになってしまう。原発と書かれたものを原発であると認識させる必要がある。

例 103 に関しては、気管支の癌に対して ICD-10 コードの付与を行うと、肺がんと同様の ICD-10 コードになってしまい、転移好発部位リストに該当してしまう。そのため、気管支の癌の扱いが不明である。

例 112 に関しては、i. より C20.0 を C20.9 に置換して実施すると、食道がん (C15.9) が原因因として選択されてしまうため、改めて C20.0 と C20.9 の違いに関する考察が必要であると考えられる。

◇ “及び”と“又は”の区別ができない (8/29 件)

該当する不正解データの例を表 10 に示す。

複数の病名や病態が 1 行中に記載される場合、以下の場合が考えられる。

- ・ A 及び B が該当する (A and B)
- ・ A 又は B が該当する (A or B)

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

表 11 不正解データ該当例（その他）

| 番号 | 例 | ICD | Iris | マニュアル |
|-----|----------------------------------|-----------------|-------|-------|
| 1 | 例 6 : I (a) 肺膿瘍 | J852 | J851 | J181 |
| 1 | (b) 大葉性肺炎 | J181 | | |
| 11 | 例 18 : I (a) 脾のうく囊>胞性線維症 | E849 | E840 | E849 |
| 11 | (b) 気管支炎及び気管支拡張症 | J40,J47 | | |
| 34 | 例 43 : I (a) 心筋症 | I429 | I429 | I426 |
| 34 | (b) アルコール依存症 | F102 | | |
| 38 | 例 50 : I (a) 脳卒中 | I64 | I64 | I119 |
| 38 | (b) アテローム<じゅく<粥>状>硬化（症）及び高血圧性心疾患 | I709,I119 | | |
| 43 | 例 55 : I (a) 後頭骨骨折 | S0210 | W 18 | G 409 |
| 43 | (b) てんかん発作による転倒 | G 409,W 18 | | |
| 59 | 例 4 : I (a) 羊水塞栓症 | O 881 | O 829 | O 754 |
| 59 | (b) 帝王切開 | O 829 | | |
| 62 | 例 7 : I (a) 心不全及び呼吸不全 | I509,I699 | I509 | X599 |
| 62 | (b) 強心剤の投与停止 | R99 | | |
| 62 | (c) 鎖骨下カテーテルの不慮の抜去 | X599 | | |
| 62 | II 胆嚢の急性破裂に対する外科手術 | K822,R69 | | |
| 154 | 例 2 : I (a) ニューモシスチス・イロベチイ 6 ヶ月 | J173 (6M onths) | Y871 | B210 |
| 154 | (b) HIV 5 年 | B24 (5Yars) | | |
| 154 | (c) 脾破裂 7 年 | S360 (7Yars) | | |
| 154 | (d) 暴行一素手の殴り合い 7 年 | Y04 (7Yars) | | |

しても“and”の意味で“、”を使用していると考えられる。そのため、“又は”の記載に
関しては、注意して ICD-10 コードを付与する必要がある。

また、“又は”の表記でも、単純にどちらかを選択するような“or”を適用できるわけ
ではなく、例 145 のような腎臓がん(C64)と膀胱がん(C679)の組み合わせから尿路
の悪性新生物(C689)の ICD-10 コードを付与するような場合があることには、留意
しなくてはならない。

◇ その他 (8/29 件)

以上に挙げた以外の不正解データを表 11 に示す。

いずれに関しても、ICD-10 コードの付与はできているが、Iris との違いがあるもの
である。

まとめ

機械実装を想定したルールを暫定的に設定し、そのルールをもとに実験を行い、ルールの
有用性や改善点を調査した。現時点での ICD-10 コード付与ルールを適用した結果、Step1

までの自然言語処理によって機械的に ICD-10 コードが付与可能なのは全体の約 5 割程度であり、全付与した状態での Iris による原死因コーディングの精度は約 8 割であることがわかった。今後は、機械実装用にルールをさらに明確化することに取り組み、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コーディング過程のデータを収集する予定である。