

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

令和2年度 総括・分担研究報告書

研究代表者 今井 健
(東京大学 大学院医学系研究科)

令和3 (2021) 年3月

目 次

I. 総括研究報告書

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	1
今井 健	
(別添資料1) 原死因確定プロセス調査(昨年度調査のアップデート)	7
(別添資料2) Irisによる仮原死因付与処理	14
(別添資料3) 機械学習を用いた原死因変更有無判定	24

II. 分担研究報告書

死亡に関わる調査票情報提供に基づいたICD10コード自動付与ツールの作成	30
香川 璃奈	
(別添資料4) 記載変換ルール	34
(別添資料5) 未コード化傷病名変換ルール	41

III. 研究成果の刊行に関する一覧表

別添資料

Iris入力データ生成プログラムソース	44
機械学習による原死因コード変更有無予測結果(一部抜粋)	55

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

研究代表者 今井 健 (東京大学大学院医学系研究科 准教授)

研究要旨

人口動態調査は国勢調査と並ぶ国の主要統計で公衆衛生施策の中心的資料である。本研究は原死因確定に関する調査を行い、我が国での原死因データ収集における課題を抽出し、ICD-11における死亡診断書や死亡統計ルールの変遷を調査すると共に、原死因確定作業に対する機械学習の適用可能性について調査・検討を行うことを目的とする。本年度は、死亡票の実データを対象に、文字列処理と自動 ICD-10 コード付与を行った上で、オートコーディングツール IRIS を適用し、約 65% の死亡票に対し、仮原死因を確定した。またこの仮原死因の変更の有無、外因や母側病態のコード追加の有無の割合について明らかにすると共に、機械学習用データセットを作成した。さらに、「何らかの付帯情報に影響され、仮原死因が変更に至るか否か」を予測する2値分類モデルによる機械学習の結果、基礎的な情報しか用いないベースライン手法でも、非常に高い精度(Accuracy90%) で変更の有無を判別可能と判明した。本手法を発展させることで、これまで人手確認によって行われてきた原死因確定作業の大幅な効率化、負荷軽減が図れると期待される。

研究分担者

香川璃奈

筑波大学医学医療系 講師

研究協力者

大井川仁美

奈良県立医科大学 MBT 学講座博士課程

大江和彦

東京大学大学院医学系研究科 教授

今村知明

奈良県立医科大学公衆衛生学講座 教授

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、より正確で効率の高いデータ収集の方法の検討が求められている。

そこで、本研究は、原死因確定に関する調査を行い、我が国での原死因データ収集における課題を抽出し、ICD-11における死亡診断書や死亡統計ルールの変遷を調査すると共に、原死因確定作業に対する機械学習の適用可能性について調査・検討を行うことを目的とする。

A. 研究目的

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

B. 研究方法

B-1) 原死因確定プロセスにおける課題の抽出

すでに昨年までで明らかになっている原死因確定プロセスの詳細割合について、平成27年～平成30年の死亡票・死亡個票実データの分析を追加し、アップデートを行った。

B-2) 機械学習の適用可能性調査

死亡票・死亡個票実データ(平成27～30年、約520万件)に対し、各種の文字列処理と標準病名マスターを利用して自動ICD-10コーディングを行うシステムを開発した。

また、全病名にICD-10コードが振られたものについては、IRISに入力し、仮原死因コードを決定すると共に、確定原死因コードと比較を行った。その際に、IRISと国内のコーディングールの差異を吸収する処理も行った。以上の一連の処理は自動化し、Docker並びに仮想マシンにおいて実行可能なシステムとして実装した。

またIRIS処理の結果、何らかの修正処理が必要なケース、不要なケースに分類し、その内容について集計を行うと共に、機械学習のための教師データの作成を行った。

最後に、このデータを用い、何らかの付帯情報が存在するケース50万件を対象とし、付帯情報によって影響を受けて「IRISが付与した仮原死因」が変更されるか否か、を2値分類する機械学習を行った。本年度はまずはベースライン手法とし、① I欄 II欄各病名のICD10コード、② 付帯情報の各項目の有無、③ IRISが付与した仮原死因、を入力データとし、分類器学習モデルとして汎用的な勾配ブースティング決定木の一つであるXGBoostを用いて、仮原死因が変更されるか否かを予測するモデルを構築した。

B-3) ICD-11における死亡診断書や死因統計ルールの動向調査

我が国の現行の死亡統計ではICD-10を元に

したWHOによる原死因選択ルールが適用されている。しかし2018年6月にWHOがICD-11をリリースした今、ICD-11における死亡統計の動向は今後の我が国へのICD-11適用に際し重要である。本年度は昨年度に引き続きWHO並びに日本WHO-FIC協力センターの関係者へのヒアリングによってこの動向調査を行った。

尚、本研究では倫理面への配慮は必要としない。

C. 研究結果

C-1) 原死因確定プロセスのアップデート

昨年度までに判明している原死因確定プロセスについて、本年度は追加で提供を受けた平成30年度のデータを加え、総数約517万件に対して改めて原死因確定の流れを計算し直し、アップデートを行った。結果を図1に示す。昨年度までと大きな変更はないが、(1) 何らかの付帯情報があるもの(32.4%)の内訳はオートコーディングツールによるコーディングエラーあり(11.5%)、なし(20.9%)となっており、(2) 付帯情報がない67.6%のものから、コーディングエラーのある3.2%を加えた35.6%について目視確認を行っていると改めて推計された。これまでのヒアリング結果とも齟齬は無い。

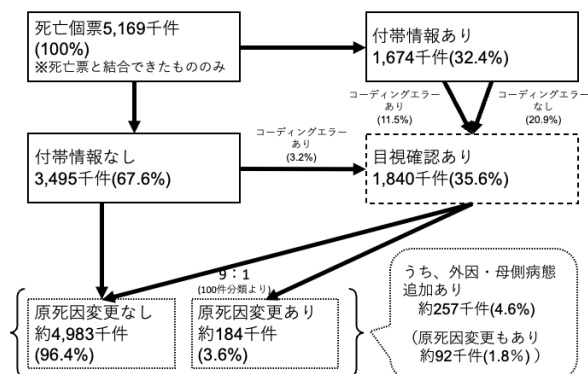


図1：原死因確定プロセス（改訂版）

詳細については、「別添資料1 原死因確定プロセス調査(昨年度調査のアップデート)」を参照されたい。

C-2) 機械学習の適用可能性

本年度は、平成 27～30 年の 4 年間の死亡票・死亡個票突合データ（約 520 万件）に対し、全 I 欄 II 欄病名に対する自動 ICD コーディングを行い、IRIS による仮原死因確定処理を行った。さらに、何らかの付帯情報が存在するケースを対象とし、付帯情報によって影響を受けて「IRIS が付与した仮原死因」が変更されるか否か、を 2 値分類する機械学習を行った。

(1) まず、4 年間の死亡票・死亡個票突合データについては、「届出地と事件簿番号」の組み合わせを各死亡案件のキーとして用いたが、複数回存在するものが 29,410 件存在したため、これを除いた約 517 万件のものを用いた（C-1, 図 1 も同様）。この複数存在するキーはヒアリングの結果、早期提出に起因するものとのことであった。

(2) 次に、全ての病名に対し、読点の削除、複数病名列挙の展開、文字の正規化などの文字列処理を施した上で標準病名マスターを用いて ICD-10 コーディングを行った。現状では、74%の病名に対して自動 ICD-10 コーディングが可能で、全ての病名が ICD-10 コード化された死亡票は 65%であった。この「65%」という割合は十分とは言えず、死亡票の特徴による偏りを可能な限り排除するために、この割合をなるべく向上させる必要がある。そこで、本年度は (A) この ICD-10 コーディング可能割合を向上させる処理、(B) ICD-10 コードが付与されたとして、その後 IRIS による仮原死因確定、変更の有無予測の機械学習、と進む処

理、とを分け、並行して研究を行った。

(A) については、研究分担者香川璃奈が中心に担当し、様々なルールベースの前処理を施すことで、最終的に「全ての病名記載に対し ICD-10 コーディング可能である死亡票の割合」は 80%にまで増加した。詳細については、香川璃奈分担研究報告書「死亡に関わる調査票情報提供に基づいた ICD10 コード自動付与ツールの作成」を参照されたい。

一方、これと並行し、(B) ICD10 コードが付与されて以降の処理を行ったため、以下は本年度当初の 65%の死亡票（全ての記載病名が ICD10 コーディングできたもの）を対象とした結果について記す。

(3) IRIS による仮原死因確定処理は全件に対して行うと約 10 日間程度かかるため、その後の調査は、約 10%をランダムサンプリングした 50 万件に対して行った。50 万件中、全病名が ICD-10 コーディングされ、IRIS への入力として利用できたものは約 32 万件(65%)である。

また、IRIS は ICD-10 の国際版に準拠しており、日本国内で適用されている独自コードは実装されていない。このような事例については IRIS が出力する仮原死因コード、死亡票における確定原死因コードのそれぞれについて修正処理を行った。

表 1 に、この調査対象とした 320,112 件（全 50 万件の調査対象セットのうち、全病名に ICD-10 コードが付与でき、IRIS に入力できたもの）の内訳を示す。IRIS が付与した仮原死因と、死亡票の確定原死因とを比較した結果、変更があるものは全体の約 10%、「何らかの付帯情報がある」81,725 件に限定すると、この中の約 15%（11,987 件）に

対し原死因の変更を行うべきであることが判明した。IRIS の処理時間は約 20 時間程度であった。

仮原死因	付帯情報 あり	付帯情報 なし	計
変更あり	11,987 (3.7%)	17,337 (5.4%)	29,324 (9.2%)
変更なし	69,738 (21.8%)	221,050 (69.1%)	290,788 (90.8%)

表 1：調査対象セット中の仮原死因変更の割合

詳細は、別添資料 2「Iris による仮原死因付与処理」を参照されたい。

また、原死因の変更以外にも、人手によるチェックで修正が行われるものがある。これが損傷や外因の影響（ICD-10 第 19 章）に対する「外因コード」（V01-Y98）の追加、周産期における母側病態コードの追加などの「コード追加」である。つまり原死因には変更がなくてもこのような補足コードが追加されることがあり得る。これについても、人手によるチェックが行われていることから、AI システムによる支援の対象と捉え、細分化した結果が以下の表 2 である。

仮原死因	コード追加 必要	コード追加 不要
変更あり	1.9%	12.8%
変更なし	4.3%	81.0%

表 2：付帯情報が存在するものに対するコード変更・追加の割合

表 2 は「何らかの付帯情報」が存在するもの（つまり必ず人手のチェックに回るもの）のうち、仮原死因の変更の有無、何らかのコード追加（外因、母側病態）の発生の有無の割合を示している。これによると、仮原死因の変更もなく、コード追加の必要もないのは約 8 割であり、残りの 2 割は何らかの修正処理が必要であることになる。

つまり、

- 1) 最初にこの 2 割と 8 割（1:4）の 2 値分類を行い、「対処の必要がない 8 割」をまず削減する
- 2) 次に 2 割のものについて、仮原死因の変更の有無、コード追加の必要性の有無を判別し、可能であればその変更先や追加コード自体の推薦を行う、

という 2 段階の支援システムが必要であることが判明した。また、これらの結果は同時に、今後支援システムに必要な分類器の学習に必要な教師データとして用いられるものである。

(4) 次に、以上の処理で得られた学習用データを元に、何らかの付帯情報が存在するケース 50 万件を対象とし、付帯情報によって影響を受けて「IRIS が付与した仮原死因」が変更されるか否かを 2 値分類する機械学習を行った。本年度はまずはベースライン手法とし、① I 欄 II 欄各病名の ICD10 コード、② 付帯情報の各項目の有無、③ IRIS が付与した仮原死因、を入力データとし、分類器学習モデルとして汎用的な勾配ブースティング決定木の一つである XGBoost を用いて、仮原死因が変更されるか否かを予測するモデルを構築した。学習に使用された特徴 (Feature) は 1577 次元であり、出力は「仮原死因の変更あり・なし」の 2 次元である。今後の精度比較のための最も単純なベースライン手法であるため、付帯情報については内容を考慮せず、単に各項目に記載があるかないかだけ (0/1) を用いている。

学習の結果、**Accuracy90.3%**で仮原死因の変更の有無が予測できることが判明した。結果を表 3（四分表）ならびに表 4（各種指標）に示す。

		正解		
		なし	あり	
予測	なし	13545	1161	14706
	あり	423	1209	1632
		13968	2370	16338
accuracy:		90.3%		

表 3：仮原死因変更の有無予測結果（4分表）

	なし	あり
recall	97.0%	51.0%
precision	92.1%	74.1%
F	0.94	0.60
specitivity	51.0%	97.0%

表 4：仮原死因変更の有無予測結果（各種指標）

また、分類に寄与した重要な特徴量について下記図 2 に示す。結果として年齢、備考欄の記載の有無、I501 コードの存在、その他付言すべき事柄の記載の有無、手術年月日の有無、性別、という順になった。I501, J188, A415 といったコードは、I 欄 II 欄病名にこれらのコードになる病名が存在することを意味し、付帯情報については内容を加味していないため、単にどの項目に記載があるかを意味している。

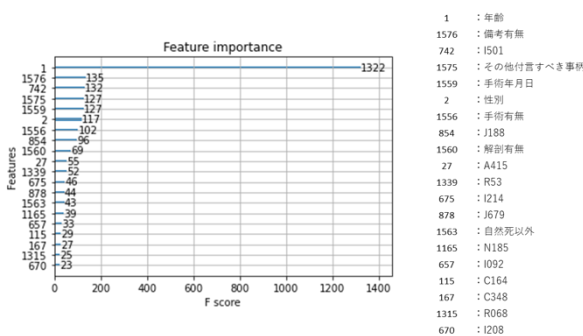


図 2：分類に寄与した重要な特徴量

詳細については「別添資料 3 機械学習を用いた原死因変更有無判定」を参照されたい。また本機械学習手法による変更有無予測結果と正誤については、一部を本統括・分担報告書全体の末尾にあ

る「別添資料」中に示したので合わせて参照されたい。

C-3) ICD-11 における死亡診断書や死亡統計ルールの動向

本年度も昨年度に引き続き関係者へのヒアリングの結果、現段階では WHO は ICD-11 における死因統計ルールについて公表しておらず、また Iris の ICD-11 対応も作業が開始されているもののリリースまでは当分時間がかかるということであった。原死因選択のルールについては基本的な考え方は踏襲されるものと思われる。しかし ICD-10 に比べて大幅に粒度が細かい疾病分類体系となった ICD-11 では Iris における原死因選択ルールテーブルが大幅に変更になり、これに合わせて我が国でのこれまでのオートコーディングシステムでのルールベースも大幅な変更を余儀なくされると予想される。次年度以降引き続き動向を注視することが必要である。

D. 考察

本年度の成果で、死亡票実データの約 65%を対象として、IRIS による仮原死因確定処理を行い、確定原死因と比較することで、原死因コードの変更の割合、また外因や母側病態コードの追加割合が明らかになった。またこの結果により今後の分類器学習に必要な教師データセットが得られ、本年度の目標を達成した。

何らかの対処が必要なもの（原死因コードの変更、外因や母側病態コードの追加）は両者を合わせて 2 割であり、最初に 1:4 の分類タスクにより「対処の必要がない 8 割」を除去した後に、細かな対処内容の分類を行う 2 段階処理が適していると考えられた。最初の分類タスクが高精度に行えるだけでまず 8 割の人手処理を削減できることが

期待でき、また2段階目の処理によって残りの2割に対する人手作業の効率化が図れると期待できる。

本年度行った仮原死因の変更の有無に関する2値分類学習では Accuracy90%と非常に高い精度での判別が可能であった。ベースライン手法であるため、各付帯情報については「記載の有無」しか用いておらず、その内容については一切考慮していない。その段階でこのような高い精度が実現できたことは驚くべきことであり、今後、付帯情報の内容を TF・IDF や BERT で学習された内部表現ベクトルなどを用いてモデルに組み込むことで一層の精度向上が見込まれ、非常に有望な手法と考えられた。特に、変更や追加の有無だけでなく、変更後・追加コードの提案まで行う場合は手術や解剖所見の文章、備考欄や付言欄の自由記述文章を対象とした BERT による汎用言語モデルの獲得と利用が必要と考えられ、現在実験中である。既にベースラインの手法で高精度を実現しているが、前述の2段階処理の第1段階目の分類精度としてはまだ向上の余地があり、次年度の課題である。

また並行して行った自由入力病名の ICD10 コーディング(分担研究)では 80%の死亡票につき全記載病名がコーディング可能と大幅に能力が向上しており、次年度ではこのデータを元に IRIS 処理、ならびに機械学習処理を行う予定である。

E. 結論

本年度研究では、死亡票の実データに対して IRIS を適用し、約 65%に対し仮原死因を決定した上で今後の分類器学習のための教師データが得られた。また原死因コードの変更、コードの追加割合についても明らかにした。付帯情報の影響による仮原死因変更の有無についての2値分類では

ベースライン手法にて Accuracy90%での判別が可能と判明した。

F. 健康危険情報

なし

G. 研究発表

1. 明神 大也, 大井川 仁美, 香川 璃奈, 今村 知明, 今井 健. 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. 医療情報学 40(Suppl.):674-676, 2020.
2. 大井川 仁美, 明神 大也, 香川 璃奈, 今村 知明, 今井 健. 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. 医療情報学 40(Suppl.):677-682, 2020.
3. 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 の 指 標 , 2020;67(3): 17-24, 2020.

H. 知的財産権の出願・登録状況

なし

別添資料1 原死因確定プロセス調査(昨年度調査のアップデート)

はじめに

本年度は昨年度に引き続き、平成27～30年度のデータに対し、原死因確定プロセスの整理を行い、詳細化・精緻化を行った。昨年度との違いは、①死亡票と死亡個票データの突合条件の変更、②「付帯情報あり」の再定義、③コーディングエラーの考慮、である。以下では、原死因確定プロセスの整理の方法と本年度の変更点、最終的に改定された原死因確定プロセスの図を示す。

方法と変更点

手順I. 死亡票と死亡個票データの突合

変更点① 死亡票と死亡個票データの突合条件の変更

昨年度は、死亡票と死亡個票それぞれの「調査年（死亡個票の場合は処理年）および届出地・事件簿の番号」を条件として内部結合で突合を行った。その結果、計5,227,867件の突合データが生成された。

しかし、厚生労働省へのヒアリングの結果から、死亡票の早期提出が行われていることが明らかになり、同様の条件で複数の突合データが内部結合により生成されてしまうことがわかった。

本年度は、昨年度の条件に加え、同様の「調査年（死亡個票の場合は処理年）および届出地・事件簿の番号」が複数回ある突合データを全て削除（どの突合データが正しいか判定できないため）する。また、今後の研究では、以上により生成された突合データを使用する。

手順II. 死亡票と死亡個票データの突合

変更点② 「付帯情報あり」の再定義

付帯情報とは、死亡票中の基本情報（生年月日や性別、住所など）と死亡の原因以外の表記を指す。付帯情報には、手術の有無や死因の種類（自然死、不慮の外因死など）などチェックを入れる場合と、所見などの自由記載を行う場合がある。

昨年度までは、付帯情報欄に記載があるという定義を行っていたが、本年度は、付帯情報をさらに正確に把握するためにバイト数や表現方法等も明らかにする。そして、どの条件を「付帯情報あり」とするか再定義を行い、①で得た突合データから、付帯情報有無別の件数と割合を算出した。

手順III. 死亡票と死亡個票データの突合

変更点③ コーディングエラーの考慮

厚生労働省へのヒアリングの結果から、現行の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）の処理過程でコーディングエラーにより、付帯情報の有無に関係なく目視確認が行われることが明らかになった。コーディングエラーは、本研究で取り扱うことができない（オートコーディングシステムが非公開）ため、機械学習への影響を考えるうえでその割合を知ることは重要である。

したがって本年度は、原死因確定プロセスの課題設定を明確にするため、昨年度に提供された、「目視確認に回ったリストの調査票情報」中の1カ月間の死亡票情報から、コーディングエラーの有無の件数と割合を算出する。

目視確認後の原死因変更の有無や外因死の追加等に関しては、昨年同様、「目視確認に回ったリストの調査票情報」中のランダムサンプリングで得られた100件の死亡票情報をもとに算出を行う。

結果

①により、重複するデータの組は29,410（2回重複：29,396、3回重複：12、4回重複：4）であり、それら（58,836件）を削除し、約5,169千（5,169,031）件の死亡票と死亡個票の突合データが生成された。

この突合データに対し、②で再定義した「付帯情報あり」の情報があったものは約1,674千（1,673,524）件（5,169,031≒32.4%）であった。ここで、再定義の仕様を本資料末尾「<参考資料> 付帯情報まとめ」に示す。各列の構成は表1の通りである。

表1 <参考資料> 付帯情報まとめ（本資料末尾）の列名と内容

列名	内容
1 列番号	昨年度作成した突合データの列番号
2 new_shibo_join 列番号	今年度作成した突合データの列番号
3 データベース項目名	突合データ各列の項目名
4 「付帯あり」の判定基準	「付帯情報あり」の定義
5 Cリスト付帯あり基準	目視確認で実際使われる項目
6 由来	死亡票情報であるか死亡個票であるか
7 頻度	出現頻度を色で示したもの（高いほど色が濃い）
8 死亡票記載	死亡票・死亡個票の情報項目
9 件数	①で得た突合データ中の該当数

10	機械学習採用	機械学習を適用するさいに必要なと思われるもの
11	追加処理	突合データに何らかの処理が必要なものとその処理内容
12	意味	項目の意味・内容

図1における付帯情報有無の判別は4列目「付帯あり」の判定基準をもとに行った。

また、「目視確認に回ったリストの調査票情報」中の1カ月間の死亡票情報から、目視確認が行われたものは40,223 (/112,939≒35.6%)であり、コーディングエラーがあったものは16,631件であった。つまり、残りの23,592件はコーディングエラーがなく目視確認していることになる。ここで、②で得た付帯情報有の32.4%を「目視確認に回ったリストの調査票情報」中の1カ月間の死亡票112,939件に外挿すると36,565件となる。以上の情報を用いてコーディングエラー数を算出した表は表2のようになった。

表2 コーディングエラー数

	コーディングエラーあり	コーディングエラーなし
付帯情報あり	12,973 (11.5%)	23,592 (20.9%)
付帯情報なし	3,658 (3.2%)	72,716 (64.4%)

目視確認後の原死因変更の有無や外因死の追加に関しては、昨年同様、「目視確認に回ったリストの調査票情報」中のランダムサンプリングで得られた100件の死亡票情報をもとに算出を行い、最終的に得られた原死因プロセスの図を図1に示す。

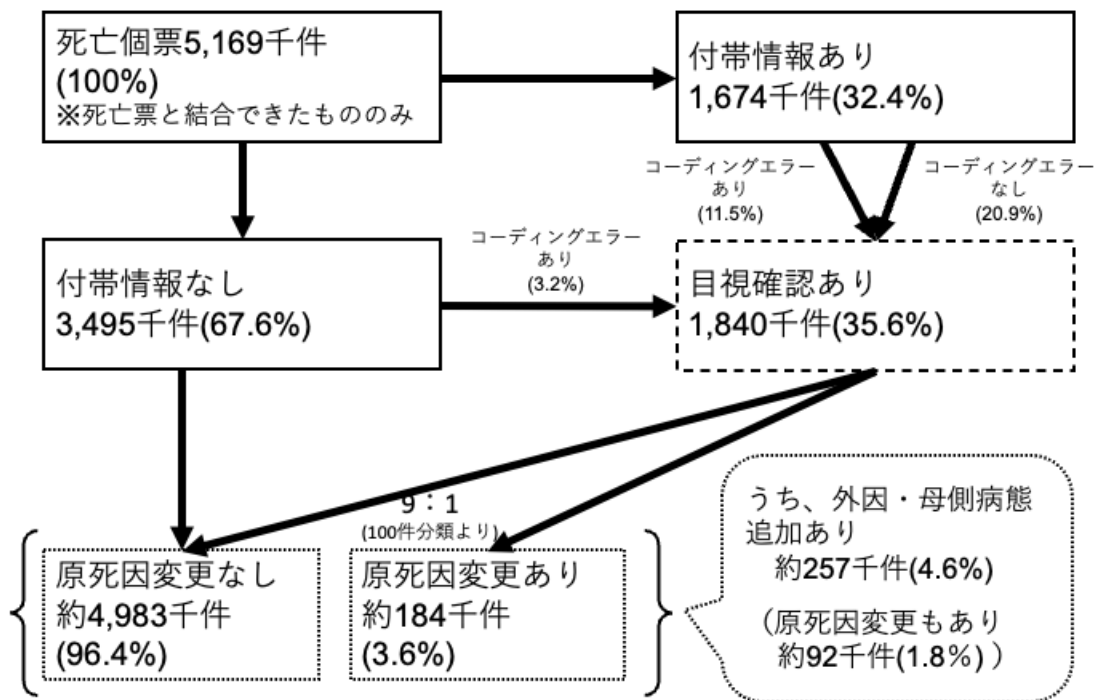


図1 原死因決定プロセス

図中の実線で囲まれたものは、突合データから得られた確定数を示すデータであり、破線は「目視確認に回ったリストの調査票情報」中の1カ月間の死亡票情報から推計された件数、点線は「目視確認に回ったリストの調査票情報」中のランダムサンプリングで得られた100件の死亡票情報から推計された件数である。

考察

①によって、外的要因（死亡票の早期提出）と内部結合によって重複が出てしまったデータが削除され、確定原死因が不明確になった部分が解消され、より正確なデータが生成されたと考えられる。重複データ数は、全体の中でも小さい割合だということもわかった。

②によって、付帯情報の内容の把握が可能になった。また、各付帯情報の頻度を算出したことから、「備考欄」「手術欄」「その他付言すべき事柄」に関する記述が多いこともわかった。

③より、コーディングエラーの影響がわかった。また、付帯情報なしでコーディングエラーが出る割合は、厚生労働省へのヒアリングの結果とも合致しており、信頼できると考えられた。

また、外死因（外因符号）の追加に関して、「目視確認に回ったリストの調査票情報」中のランダムサンプリングで得られた100件の死亡票情報より、母側符号の考慮も必要であることがわかった。本年度以降、この外因符号と母側符号を総じて追加コードとすることとした。そして、「目視確認に回ったリストの調査票情報」中のランダムサンプリングで得られた100件の死亡票情報より、原死因の修正と追加コードの有無を算出した表を表3に示す。

表3 コーディングエラー数

	原死因修正あり	原死因修正なし
追加コードあり	3	7（うち母側符号1）
追加コードなし	4	

表より、目視確認で原死因の変更と同様に外因符号の追加も原死因プロセスの課題として考えられたため、いずれ対策が必要である。

まとめ

平成 27～30 年度のデータに対し、原死因確定プロセスの整理を行った。昨年度と大きな差はなかったが、より正確な原死因プロセスが可能になった。

昨年度までは、外因追加にのみ考慮してきたが母側符号の追加も原死因プロセスの中では必要な作業であった。そのため、2つを追加コードと新たに定義し、今後「追加コードの付与」も原死因プロセスの課題として検討していく必要がある。

<参考資料> 付帯情報まとめ

列番号	new_shiba_join 列番号	データベース項 目名	「付帯あり」の判 定基準	CUIスト 付帯あり基準	由来	頻度	死亡票記載	件数	機軸字番 採用	追加処理	意味
0		id									irisに入れる用の連し番号
1		sub_area					届出地				届け出が出された場所の番号
2		jken_nm					事件簿番号				事件簿番号
3		shori_ym					処理年月				処理年月
4		bikou_yn					備考欄記載有無				備考欄に記載があるかないか
5		death_kbn					「死亡した人の住所」のうち届出地区分				届出地の区分
6		d_tdfk					都道府県				死亡したときの都道府県名
7		d_city					市区町村				死亡したときの市、郡、東京都の区
8		d_town					市区町村				死亡したときの町、村、指定都市の区
9		d_add					コード				死亡したときのコード
10		la_c					「死亡の原因」に含まれるI欄・A欄の原因	5160414			A欄記載の傷病名
11		la_p					期間	5085556			期間
12		lb_c					「死亡の原因」に含まれるI欄のI欄の原因	1842189			I欄記載の傷病名
13		lb_p					期間	1430645			期間
14		lc_c					「死亡の原因」に含まれるI欄のウ欄の原因	389976			ウ欄記載の傷病名
15		lc_p					期間	309219			期間
16		ld_c					「死亡の原因」に含まれるI欄の工欄の原因	72893			工欄記載の傷病名
17		ld_p					期間	54175			期間
18		ll_c					「死亡の原因」に含まれるII欄の原因	1726024			II欄記載の傷病名
19		ll_p					期間	1611341			期間
20		ope_flg		2			手術フラグ	739210	1		手術の有無
21		ope_detail	記述あり	記述あり		死亡届票	手術の部位及び所見	739003	1		手術の詳細記述
22		ope_bk_flg	記述あり				(手術)備考欄への記載	8	2		手術に関して備考がある場合はフラグ
23		ope_ymd	記述あり				手術日	656468	0	死亡日から引き算	手術年月日
24		ap_flg		2			解剖フラグ	119624	1		解剖の有無
25		ap_detail	記述あり	記述あり			解剖の部位及び所見	119422	1		解剖の詳細記述
26		ap_bk_flg	記述あり				(解剖)備考欄への記載	5	2		解剖に関して備考がある場合はフラグ
27		dc_type	空と0と1以外				死因の種類	233251			自然死・不慮の外因死・その他の不詳の死などの種類を表す数字
28		ini_ymd	記述あり				傷害が発生したとき	193719			障害発生年月日
29		ini_hm	記述あり				傷害が発生したとき	150902			障害発生時分
30		ini_p_type	記述あり	記述あり			傷害が発生したところの種類	194125			住居・工場・道路などの障害が発生した場所の種類を表す数字
31		ini_detail	記述あり				傷害が発生したところその他の記述	54432			傷害が発生したところの種類にない場合の記述(その他)
32		ini_tdfk	記述あり				傷害発生場所	187845			傷害発生都道府県
33		ini_city	記述あり				傷害発生場所	186349			傷害発生市郡
34		ini_town	記述あり				傷害発生場所	78935			傷害発生区町村
35		ini_method	記述あり	記述あり			手段及び状況	191053	1		障害発生時の手段及び状況に関する詳細記述
36		ini_biko	記述あり				(傷害)備考欄への記載	2	2		傷害発生に関して備考がある場合はフラグ
37		inf_detail	記述あり	記述あり			「生後1年未満での病死」の病態・異状の詳細	2427	1		妊娠・分娩時における母体の病態または以上に關する詳細記述
38		inf_biko_yn	記述あり				備考欄への記載	1	2		「生後1年未満での病死」に関して備考がある場合はフラグ
39		sonota	記述あり	記述あり			その他付帯すべき事柄	339827	1		その他に付帯すべきことに関する詳細記述
40		biko_yn	記述あり				備考欄外字有無	881572			備考欄に外字がある場合はフラグ
41		biko_detail	記述あり	記述あり			備考欄	881570	1		備考に関する詳細記述
42		ym					調査年				調査年
43		sub_ym					提出年月				死亡票提出年月日
44		jken_sb					事件簿番号サブ				事件簿番号の補助番号
45		sex					性別				性別
46		birth_yn					出生年月日フラグ				出生年月日あるか
47		birth_ymd					出生年月日時分				出生年月日
48		birth_hm					出生年月日時分				出生時分
49		death_yn					死亡年月日フラグ				死亡年月日あるか
50		death_ymd					死亡年月日時分				死亡年月日
51		death_hm					死亡年月日時分				死亡時分
52		ori_dc					原死因				原死因コード
53		galln					外因符号	196071			外因コード
54		mom_sy					母親病態	6978			母親病態コード
55		twin_yn					単胎多胎	6978			単胎か多胎か数字
56		preg_yn					妊娠適否	6978			妊娠の有無
57		preg_w					妊娠週数	6978			妊娠週数
58		mon_b_ymd					母の生年月日	6978			母の生年月日
59		p_preg_bir					前回の妊娠	6978			前回までの出生児数
60		p_preg_dea					前回の妊娠	6978			前回までの死産児数
61		child_1					子の数	6978			今回の出生子含む出生子
62		child_2					子の数	6978			出生児

別添資料2 Irisによる仮原死因付与処理

はじめに

本年度新たに作成した死亡票と死亡個票の突合データから、ランダムサンプリングで抽出した50万件のデータを対象に、今後の300～500万件規模の研究のための実験を行った。具体的には、現行の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）に替わりIrisを用いて原死因を付与し、その結果（以下、仮原死因）とオートコーディングシステムが導出した確定原死因との比較を行った。また、付帯情報の有無に関する情報も抽出し、付帯情報と追加コードの影響について考察を行った。

実験 方法

手順1. 対象データを抽出

本年度新たに作成した死亡票と死亡個票の突合データから、ランダムサンプリングで抽出した50万件に対し、各突合データ中の傷病名に全てICD-10コードが付与できたもの（320,112件）を対象とする。

手順2. 対象データの確定原死因、追加コード、付帯情報有無の情報抽出

手順1.で抽出したデータに関して、確定原死因と追加コードを抽出する。追加コードとは、死亡票における外因符号と母側符号のことを指す。また、付帯情報の有無を0（付帯情報なし）と1（付帯情報あり）で表現し抽出する。「付帯情報あり」の定義は、別添資料1「原死因確定プロセス調査（昨年度のアップデート）」の末尾 **参考資料「付帯情報まとめ」**の4列目に示す通りである。

手順3. 対象データをIris用のフォーマットに整形

突合データからIrisに必要な以下の情報を抽出し、Irisのフォーマットに則り整形を行う。

- 死亡票識別番号
- 生年月日
- 死亡年月日
- 性別
- I・II欄の傷病名

IrisはMicrosoft Access形式であり、基本情報が含まれるテーブル“Ident”とI・II欄に記載される傷病名の情報が記載されるテーブル“MedCod”の

2つから構成される。そのため、Access で読み取り可能な tsv 形式ではじめにファイルの整形を行い、自動的に Access ファイルを生成するバッチファイルを使用することで Iris に適用可能な Access ファイルを生成する。ファイルの生成具体的な Iris のフォーマットは、本資料末尾「<参考資料> Iris のフォーマット」に示す。

また、これまでの研究により、表 1 左から 1 列目に示す ICD-10 コードは日本（標準病名マスター）独自のコードであるため、Iris がコーディングエラーを引き起こす。そのため、左から 2 列目のように、事前に ICD-10 コードの変更を行う。

表 1 日本独自コードと Iris の対応

対象 ICD-10 コード	変更後 ICD-10 コード	傷病名
J9699	J969	呼吸不全
J9609	J960	急性呼吸不全
E14	E149	糖尿病
J9619	J961	慢性呼吸不全
G903	G239	多系統萎縮症
E11	E119	2 型糖尿病
B59	J173	ニューモシスチス肺炎
I7020	I702	末梢動脈硬化症
J9691	J969	2 型呼吸不全
E10	E109	1 型糖尿病
M7265	M726	外陰部壊死性筋膜炎
R17	R198	黄疸
J9611	J961	慢性 2 型呼吸不全
E13	E139	B 型インスリン受容体異常症
J9601	J961	急性 2 型呼吸不全

手順4. Iris による仮原死因付与

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。仮原死因の付与では、Iris のバッチ処理機能を用いて一括で原死因付与を行う。

手順5. 仮原死因と確定原死因の比較

Iris による仮原死因と、オートコーディングシステムによる確定原死因

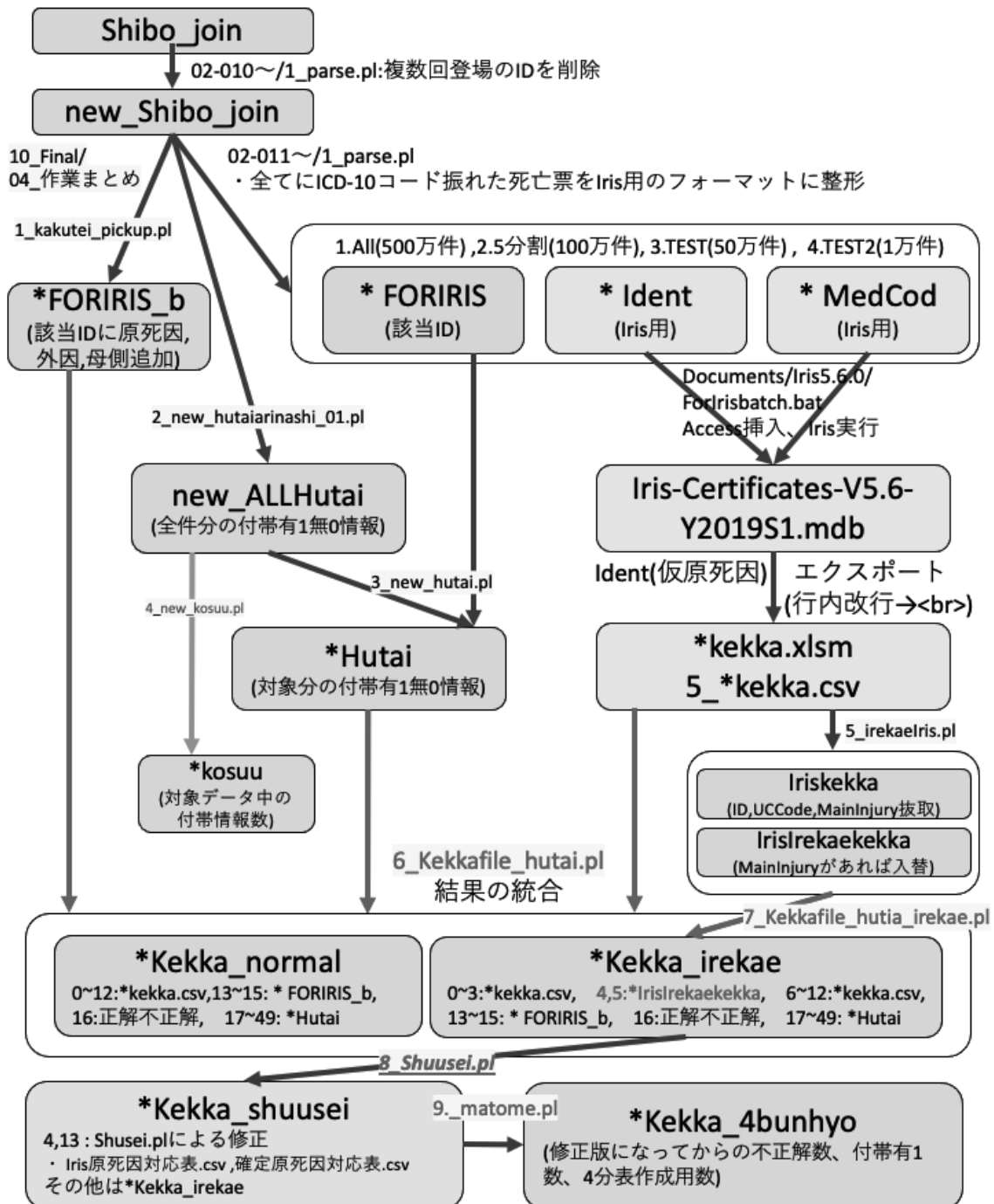
との比較を行い、一致率を算出する。ただし、比較を行う前に以下ことを行う。

- Iris の外因（複合）コードを仮原死因に変更
 - 厚生労働省へのヒアリングの結果、日本のオートコーディングシステムでは、S と T から始まるコードが原死因になり得るとし、死因統計を行う際、目視確認で付与する外因符号と合わせて集計を行っている。一方、WHO には、S と T に関するコードは原死因に使用しないという指針があり、Iris もそのルールに則っていることがわかった。そのため、Iris の仮原死因欄には、日本における外因コード、外因（複合）コード欄には日本における仮原死因が選ばれており、そのままの比較では一致しない。そこで、Iris で外因（複合）コードに記載がある場合は、仮原死因と交換する作業を行う。
- 標準病名マスターと Iris どちらにも存在しないが確定原死因に用いられる日本独自コードの対策
 - 厚生労働省へのヒアリングの結果、日本独自の詳細分類や、準拠する ICD-10 コードの違い(平成 28 年度以前は ICD-10 2003 年版、平成 29 年以降は 2013 年版) から、仮原死因と確定原死因の不一致が生じる場合があることがわかった。そこで、それぞれに関して、修正するか比較対象としない選択を行う。その基準は以下の通りである。
 - ◇ オートコーディングシステムによる確定原死因側
 - 1 桁多いため、仮原死因と一致しない
 - 桁を落とす
 - Iris の辞書に存在しない
 - 比較対象としない
 - 準拠する ICD-10 コードが変化した
 - 比較対象としない
 - ◇ Iris による仮原死因側
 - 1 桁多いため、確定原死因と一致しない
 - 桁を落とす

また、仮原死因と確定原死因が異なる場合は、仮原死因の変更があったとみなし、付帯情報と追加コードの有無を考慮して表を作成する。

以上の手順を自動化可能にするプログラムを作成した（次ページ Figure1 参照）。

(Figure. 1) 生成ファイルと処理プログラム概要



結果

50 万件のランダムサンプリングした突合データから、全ての病名に対して ICD-10 コードが付与できた 320,112 件に対し、Iris の処理時間は約 19 時間であった。確定原死因と仮原死因が比較可能であったのは、320,008 件であった。

仮原死因と確定原死因の一致率は、91.0%であった。また、仮原死因の変更有無と追加コードの変更有無、付帯情報の有無の表を表 2 に示す。

表 2 仮原死因変更・付帯情報・追加コードの有無別の件数

	付帯情報あり	付帯情報なし
仮原死因変更あり	11,857(/320,008 ≒ 3.7%)	16,957(/320,008 ≒ 5.3%)
追加コードあり	1,554(/81,695 ≒ 1.9%)	389(/238,313 ≒ 0.2%)
追加コードなし	10,303(/81,695 ≒ 12.6%)	16,568(/238,313 ≒ 7.0%)
仮原死因変更なし	69,838(/320,008 ≒ 21.8%)	221,356(/320,008 ≒ 69.2%)
追加コードあり	3,547(/81,695 ≒ 4.3%)	715(/238,313 ≒ 0.3%)
追加コードなし	66,291(/81,695 ≒ 81.1%)	220,641(/238,313 ≒ 92.6%)

考察

処理時間に関して、これまでの実験より、3,267 件に対する処理時間は約 4 分、6,484 件のデータに対する Iris の処理時間は約 12 分であった。6,484 件の場合の処理時間を基準に処理時間は件数の増加と等倍になると推測していたが、今回の 320,112 件に対する処理時間は推測の約 2 倍の時間を要していることがわかった。このことから、Iris では死亡票の件数が多いほど処理時間が爆発的に増えると考えられる。今後の研究では、対象データを 300 万～500 万件に拡張するが、全件を一括処理させるのではなく、多数に分割して一括処理を行うべきであり、分割の数を検討する必要がある。

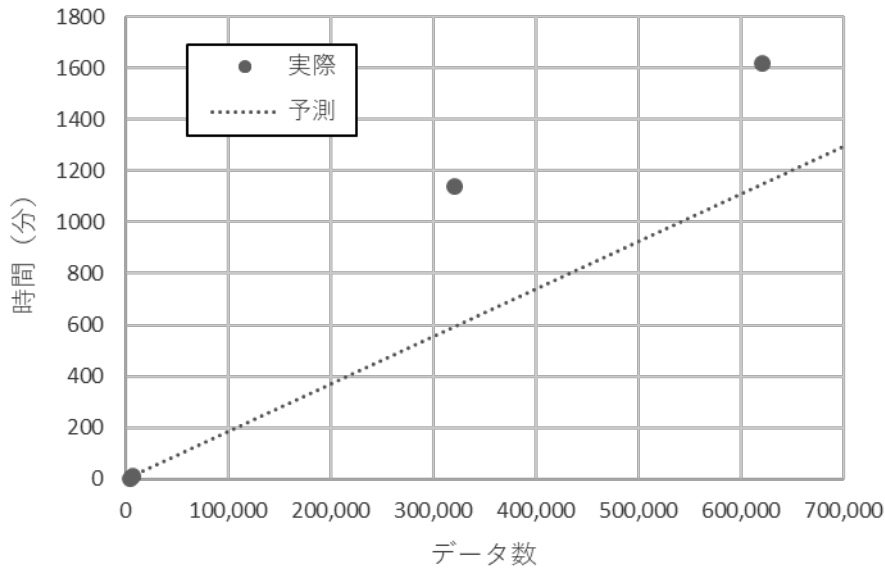


Figure 2 Irisの処理時間

確定原死因と仮原死因が比較に関して、手順5.の操作により、機械学習に用いることのできるデータ数の減少が懸念されたが、比較対象とならないものはほぼ存在しないことがわかった。そのため、手順5.の操作は、今後の研究でも用いていく。

仮原死因と確定原死因の一致率より、仮原死因の変更があるとみなされるものは全体の約10%であることがわかった。ここで、研究班の研究対象は「付帯情報あり」のものであるため、表2の黄色部分(81,695件)に限定すると、原死因の変更が行われるのは約15%であることがわかった。また、原死因の変更以外にも、目視確認時に追加コードが付与されていることは、本年度行った原死因プロセスに関する研究で明らかになっている(別添資料1参照)。表2の追加コードに関する情報から、仮原死因の変更と追加コードの付与どちらも必要もないのは約8割であり、残りの2割は何らかの処理(原死因コードの変更・追加コードの付与)が必要であることになる。以上のことから、今後の研究として、付帯情報のある死亡票を対象に機械学習を用いて解くべき課題は、以下の2段階であると考えられる。

- ①対処(原死因コードの変更、外因や母側病態コードの追加)の必要がない8割の除去
- ②細かな対処内容の分類

これらの効果に関して、①が達成されることにより、8割の人手処理の削減が可能になると考えられる。できることが期待でき、2段階目の処理によって残りの2割に対する人手作業の効率化が図ることができると期待できる。また、2段階の先としては、具体的な原死因変更先のICD-10コードや追加コードの提示が考えられる。

まとめ

本年度新たに作成した死亡票と死亡個票の突合データから、ランダムサンプリングで抽出した 50 万件のデータを対象に、Iris による仮原死因とオートコーディングシステムによる確定原死因の比較を行った。その結果、何らかの対処が必要なもの（原死因コードの変更、外因や母側病態コードの追加）は約 2 割であることがわかった。そして、今後の研究課題として、以下の 2 段階を挙げた。

- ① 対処（原死因コードの変更、外因や母側病態コードの追加）の必要がない 8 割の除去
- ② 細かな対処内容の分類

今後 300～500 万件の規模の研究のために留意すべき点として、Iris による仮原死因の付与の作業において、一括処理を行う死亡票の量を調整する必要がある。

<参考資料> Iris フォーマット

Irisに与えるファイル

- (理想) Access形式：2つのテーブルが必要
 - 末尾に"Ident"がつくテーブル
 - Ex. TestIdent
 - 死亡票の基本情報
 - 末尾に"MedCod"がつくテーブル
 - Ex. TestMedCod
 - 死亡票の I II 欄情報
- (方法) 2種類のtsvをIrisのAccessファイルにインポート
 - ○○Ident.tsv
 - ○○MedCod.tsv

インポートだけで済むフォーマットをお願いしたい

列番号	項目	説明	対応	列番号	項目	説明	対応
0	CertificateKey	識別番号 (主キー)	数え上げ? 市区町村コード等?	23	DiagnosisModified	手動で修正	空で良い
1	LastChange	最終更新	空で良い (後にIris更新)	24	Residence	国籍	空で良い
2	DateBirth	生年月日	new_birthそのまま	25	Name	氏名	空で良い
3	DateDeath	死亡年月日	new_deathそのまま	26	Address	住所	空で良い
4	Age	年齢	空で良い	27	AutopsyRequested	解剖有無	空で良い
5	Sex	性別	sex	28	AutopsyUsed	解剖主要所見	空で良い
6	MannerOfDeath	死亡の種類	0を記載	29	RecentSurgery	手術有無	空で良い
7	UCCode	原死因コード	空で良い (後にIris更新)	30	DateOfSurgery	手術年月日	空で良い
8	MainInjury	主傷病名	空で良い (後にIris更新)	31	ReasonSurgery	手術主要所見	空で良い
9	Status	Irisの処理状況	Initialを記載 (後にIris更新)	32	DateOfInjury	傷害発生年月日	空で良い
10	Reject	エラー理由	Noを記載 (後にIris更新)	33	PlaceOfOccurrence	傷害発生場所	空で良い
11	Coding	コーディング方法	Automaticを記載	34	ActivityCode	傷害発生場所コード	空で良い
12	CodingVersion	Irisバージョン	空で良い (後にIris更新)	35	ExternalFreeText	手段及び状況	空で良い
13	CodingFlags	複数の候補がある	空で良い (後にIris更新)	36	Pregnancy	妊娠状況	空で良い
14	SelectedCodes	I II 欄読み込み結果	空で良い (後にIris更新)	37	PregnancyContributeDeath	死亡と妊娠の関与	空で良い
15	SubstitutedCodes	I II 欄Iris用の正規表現	空で良い (後にIris更新)	38	Stillbirth	死産有無	空で良い
16	ErrnCodes	現バージョンは関係なし	空で良い	39	MultiplePregnancy	単胎・多胎	空で良い
17	AcmeCodes	MUSEへの読み込み	空で良い (後にIris更新)	40	CompletedWeeks	妊娠週数	空で良い
18	MultipleCodes	MUSEの分析コード	空で良い (後にIris更新)	41	BirthWeight	出生体重	空で良い
19	Comments	付言すべきことから	空で良い	42	AgeOfMother	母の生年月日	空で良い
20	FreeText	備考	空で良い	43	ConditionsMother	母体の状況	空で良い
21	ToDoList	Irisのリジェクト内容表示	空で良い (後にIris更新)	44	CertImage	画像ファイル有無	空で良い
22	CoderReject	手動でリジェクト	空で良い				

TestIdent.tsv変換規則

準備：0行目0列目～44列目にCertificateKey～CertImage（前スライド参照）を記載

- ① 0列目(CertificateKey)： 識別番号
 - 一番上を0000001として数え上げ（桁数を揃える）
- ② 2列目(DateBirth)： 46(birth_ymd)
- ③ 3列目(DateDeath)： 49(death_ymd)
- ④ 5列目(Sex)： 44(sex)
- ⑤ 6列目(MannerOfDeath)： 0
- ⑥ 9列目(Status)： Initial
- ⑦ 10列目(Reject)： No
- ⑧ 11列目(Coding)： Automatic
- ⑨ 他の欄は空

1死亡票につき
1行

行列番号は
0からカウント
しています

MedCodテーブル用tsvフォーマット

列番号	項目	説明	対応
0	CertificateKey	識別番号（主キー）	数え上げ？ 市区町村コード等？
1	LineNb	傷病名記載箇所（主キー）	0-3がI欄, 5がII欄
2	TextLine	傷病名	空で良い
3	CodeLine	ICD-10コード	ICD-10コード(期間)
4	IntervalLine	期間	空で良い（3列目に記載）
5	CodeOnly	ICD-10コードのみでのコーディング	1を記載
6	LineCoded	テキスト入力がある	1を記載（バッチ処理で結局1に変更される）

CertificateKey	LineNb	TextLine	CodeLine	IntervalLine	CodeOnly	LineCoded
000001	0		G903		1	1
000001	1		J181		1	1
000002	0		K829		1	1
000002	1		K831		1	1
000002	2		C250		1	1
000003	0		I619		1	1
000003	1		I10		1	1
000003	2		N119		1	1
000003	3		D291		1	1

TestMedCod.tsv変換規則

準備：0行目0列目～6列目にCertificateKey～LineCodedを記載

※ List.txt の9(la_c), 11(lb_c), 13(lc_c), 15(ld_c), 17(lI_c)に記載があれば, その都度, 以下を生成. なければ次の死亡票へ

① 0列目(CertificateKey): 識別番号・・・TestIdent.tsvと同じ

② 1列目(LineNb) : I II欄の該当する番号

I欄ア(la_c) : 0
I欄イ(lb_c) : 1
I欄ウ(lc_c) : 2
I欄エ(ld_c) : 3
II欄オ(lI_c) : 5

③ 2列目(TextLine) : 空

④ 3列目(CodeLine) : ③に対応するICD-10コード

- 香川先生により振られたICD-10コード
- 複数病名も記載可能 (区切り文字任意)
 - Ex. C169,C189 とか C169 C189
- 期間がある場合は(数字期間)書き
 - Ex. C169(3Years),Months,Days,Hours

20201226ver仕様書より
I欄ア : 63 (64列目)
I欄イ : 65 (66列目)
I欄ウ : 67 (68列目)
I欄エ : 69 (70列目)
II欄オ : 71 (72列目)

⑤ 4列目(IntervalLine) : 空

- ⑤にて()書きする

⑥ 5列目(CodeOnly) : 1

⑦ 6列目(LineCoded) : 1

1死亡票につき
1行～5行

行列番号は
0からカウント
しています

別添資料3 機械学習を用いた原死因変更有無判定

はじめに

Irisによる仮原死因が付与できた320,008件中、付帯情報があるデータを対象に、今後の300～500万件規模の研究のための実験を行った。実験では、教師あり機械学習手法であるXGBoostを用いて、Irisが付与した仮原死因が変更されるか否かの二値分類を行った。

実験1

方法

手順1. 機械学習用データの生成

「3-2 別添資料2 Figure.1」で示す8_Shuusei.plを適用して生成されるデータと320,008件で使用された仮原死因のICD-10コードのデータ(1553種類)を組み合わせて、機械学習用のデータを生成する。その仕様は以下の表の通りである。

表1 機械学習用データの仕様

列番号	列名	内容
1	certificateKey	死亡票の識別番号
2	age	年齢(小数点第1位まで)
3	sex	性別(男:1、女:2)
4～1556	各ICD-10	存在しなければ0。存在する場合、 仮原死因に該当:1 I II欄記載有:優先順位に基づき、0.85,0.7,0.55,0.4,0.25
1557～1578	各付帯情報	別添資料3-1<参考資料>の付帯ありに該当すれば1、しなければ0
1579	変更有無	仮原死因が確定原死因と一致すれば1、しなければ0

1579列目に関して、仮原死因はIrisが算出したものであり、確定原死因は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)である。そのため、変更有無は、Irisが算出した仮原死因はオートコーディングシステムが算出する仮原死因と同様であるという前提のもと算出される。したがって、以降の学習結果においても何らかの影響を与える可能性があることに留意しなくてはならない。

手順2. 機械学習の実行

手順1.で生成したデータを用いて機械学習を行う。機械学習では、表1に示す2～1578列目を用いて1579列目の変更有無を予測する。

機械学習の手法として、勾配ブースティング決定木(Gradient Boosting

Decision Tree: GBDT)の一種である XGBoost を使用する。XGBoost は、アンサンブル学習 (複数の弱学習器を組み合わせる手法) のブースティング (弱学習器を逐次的に構築) と決定木 (木構造を用いて分類や回帰を行う手法) を組み合わせたものであり、比較的精度の高い結果を算出することができる。ただし、テストデータの大きさやパラメータの設定が精度に影響を与える可能性が高く、調査が必要である。そこで、実験は以下のように進める。

① テストデータの大きさを以下のように変更し実行

(ア)0.1 (全体の 10%)

(イ)0.15 (全体の 15%)

(ウ)0.2 (全体の 20%)

(エ)0.25 (全体の 25%)

(オ)0.3 (全体の 30%)

ここで、テストデータに用いられなかったデータは学習用の (訓練) データとして用いられ、

② 交差検証とグリッドサーチを用いて実行

交差検証では、データをいくつかのセットに分割したのち、1つをテストデータのセット、残りを学習データとして精度の評価を行う。そして、分割したデータのセットすべてが1回ずつテストデータになるように学習を行なって出てきた精度の平均を算出する。今回は、分割数を5とする。

グリッドサーチとは、学習モデルに用いられるハイパーパラメータ (調整の必要があるパラメータ) を調整する手法であり、指定したハイパーパラメータの全組合せに対して学習を行い、最適なパラメータを選択する。ただし、組み合わせ数が多いほど時間を要するため、今回は決定木の深さにのみ着目し、深さを4,6,8と変えていく。

手順3. 予測と正解の比較

手順 2.から予測した結果をもとに、精度の算出と予測に効果的だとされたパラメータを算出する。

結果

① テストデータの大きさ別の結果

各テストデータの大きさ別の結果を図1~図5に示す。各図に関して、左の表は四分表にまとめた結果であり、一番下の行には精度が示されている。また、緑のセ

ルはテストデータの総数、青いセルは各行あるいは各列の総数である。一方、右の表は「変更なし」あるいは「変更あり」に着目した際の感度 (recall)・適合度 (precision)・F 値・特異度 (specificity) を示している。

		正解					
		なし	あり		なし	あり	
予測	変更なし	6792	618	7410	recall	97.3%	47.8%
	変更あり	192	567	759	precision	91.7%	74.7%
		6984	1185	8169	F	0.94	0.58
accuracy:		90.1%			specificity	47.8%	97.3%

図 1 (ア) 0.1 (全体の 10%)

		正解					
		なし	あり		なし	あり	
予測	変更なし	10193	890	11083	recall	97.3%	49.9%
	変更あり	283	888	1171	precision	92.0%	75.8%
		10476	1778	12254	F	0.95	0.60
accuracy:		90.4%			specificity	49.9%	97.3%

図 2 (イ) 0.15 (全体の 15%)

		正解					
		なし	あり		なし	あり	
予測	変更なし	13604	1220	14824	recall	97.4%	48.5%
	変更あり	364	1150	1514	precision	91.8%	76.0%
		13968	2370	16338	F	0.94	0.59
accuracy:		90.3%			specificity	48.5%	97.4%

図 3 (ウ) 0.2 (全体の 20%)

		正解					
		なし	あり		なし	あり	
予測	変更なし	16961	1520	18481	recall	97.1%	48.7%
	変更あり	499	1442	1941	precision	91.8%	74.3%
		17460	2962	20422	F	0.94	0.59
accuracy:		90.1%			specificity	48.7%	97.1%

図 4 (エ) 0.25 (全体の 25%)

		正解					
		なし	あり		なし	あり	
予測	変更なし	20371	1837	22208	recall	97.2%	48.3%
	変更あり	581	1718	2299	precision	91.7%	74.7%
		20952	3555	24507	F	0.94	0.59
accuracy:		90.1%			specificity	48.3%	97.2%

図 5 (オ) 0.3 (全体の 30%)

以上より、得られた精度をテストデータの大きさ別に示すと図 6 のようになる。

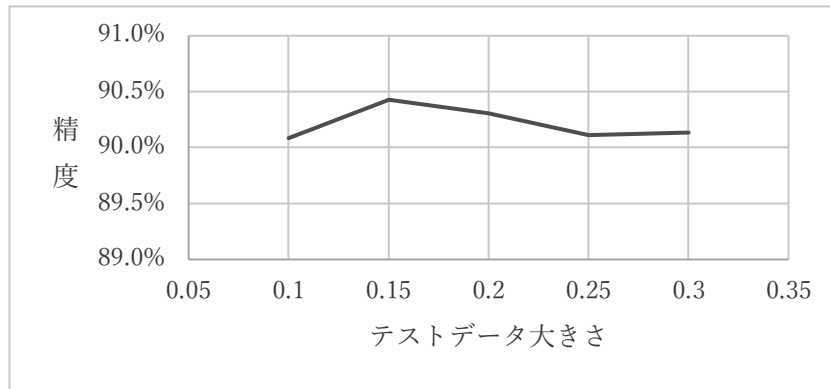


図 6 テストデータの大きさ別の精度推移

② 5 分割交差検証とグリッドサーチを用いた結果

5 分割交差検証とグリッドサーチをした結果、最適な木の深さは 4 であった。結果を図 7 に示す。

		正解		
		なし	あり	
予測	なし	13545	1161	14706
	あり	423	1209	1632
		13968	2370	16338
accuracy:		90.3%		

	なし	あり
recall	97.0%	51.0%
precision	92.1%	74.1%
F	0.94	0.60
specitivity	51.0%	97.0%

図 7 5 分割交差検証 (木の深さ 4) の結果

一方、テストデータの大きさが 0.2 のとき、学習に有効だと考えられた項目を図 8 に示す。

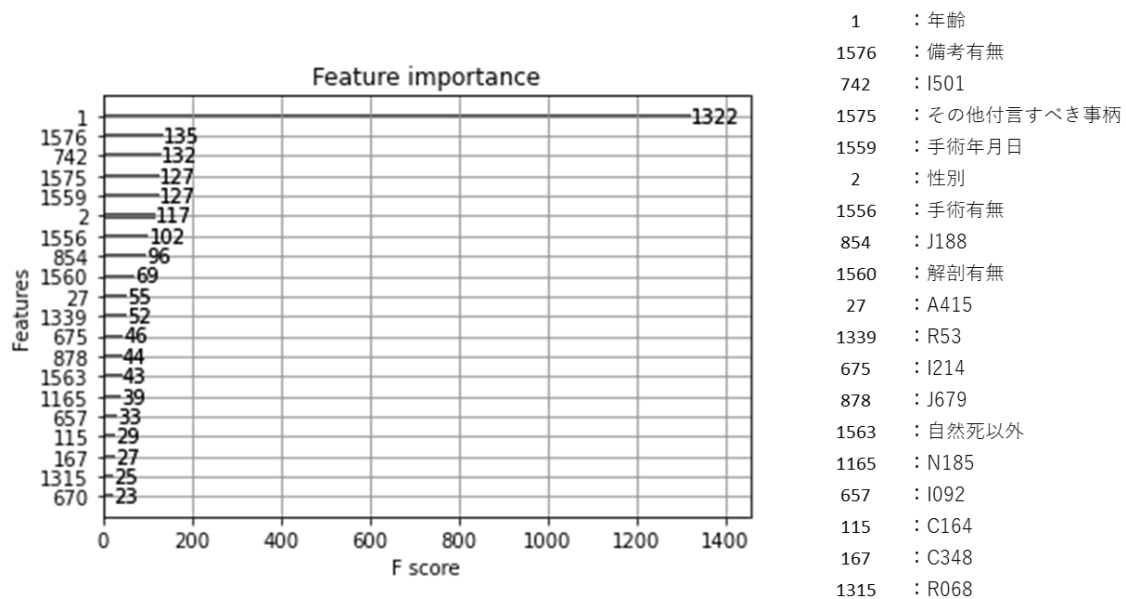


図 8 学習に影響を与えた項目 (テストデータの大きさ : 0.2)

考察

仮原死因の変更有無に関する機械学習を行った結果、約 9 割の精度で分類が可能であることがわかった。また、図 1~6 に示すテストデータの大きさ別の結果に関して、大きさを 0.1~0.3 まで変化させて実験を行ったが、精度に大きな違いはないと考えられた。また、図 7 に示す 5 分割交差検証とグリッドサーチを行った結果も精度に大きな違いはなかった。そのため、今回の実験ではテストデータの大きさやハイパーパラメータの選択の違いによる学習への影響は少ないと考えられる。

図 8 学習に影響を与えた年齢・性別や ICD-10 コード、付帯情報有無の項目に関して、最も影響を与える項目は年齢であることがわかった。また、そのほかにも備考欄の有無に関してやその他付言すべき事柄に記入があるかなど付帯情報が影響を与えていた。

5 分割交差検証とグリッドサーチにより最適とされたモデルによる予測と正解が異なるものについての考察を行う。

前提として、手順 1.でも述べたように、正解データは Iris (オートコーディングシステムと同じ挙動であると仮定) の算出した仮原死因とオートコーディングシステムによる確定原死因の違いをもとに算出しているため、Iris とオートコーディングシステムの仕様の違いが影響を与えている可能性がある。

また、機械学習による予測で「変更なし」であったが正解は「変更あり」であった場合のうち 108 件 (129 件中) に関しては、Iris が仮原死因を振っていなかった。これは、ICD-10 コードを付与したのちに Iris で仮原死因付与を行う際に、I II 欄に記載された傷病名に対して付与した ICD-10 コードが日本独自コードである場合や、入力された ICD-10 コードに不備 (前期破水は男性には使えないなど、性別や年齢と ICD-10 コードに不一致) がある場合、Iris は仮原死因を付与することができないためである。この問題は、ICD-10 コード付与時に改めて日本独自コードを修正することや Iris が出力するエラー (rejected) のメッセージを参照し仮原死因の付与を行い対策する必要がある。

機械学習による予測で「変更なし」であったが正解は「変更あり」であった場合に関して、各付帯情報の記載がある (1) 件数を算出すると、手術欄の詳細記述欄に記載がある場合が最も多い (686/1161 件) ことがわかった。より精度を上げるためには、詳細記述欄の記述内容に着目する必要がある (特定の単語や文言が仮原死因変更に影響を与える可能性があるため)、次年度以降の課題とする。

一方、機械学習による予測で「変更あり」であったが正解は「変更なし」であった場合に関しても手術欄の詳細記述欄そして手術の有りの番号にマークがされている場合が最も多かった (321/423 件)。機械学習による予測で「変更なし」であったが正解は「変更あり」であった場合と同様に付帯情報の詳細記述欄などを参照する必要がある。ただし、実用する

際は必ず人の目視確認が入るため、数は減らすべきであるが、他の対応を優先するべきである。

まとめ

320,008 件のデータを対象に XGBoost を用いた機械学習を行った。機械学習では、記載されている ICD-10 コードと付帯情報の有無に関する項目から、仮原死因の変更が行われるか行われないかを予測しており、精度は約 9 割であった。また、その中でも年齢が最も影響を与えていることがわかり、その他の付帯情報の記載有無も有効であることが示唆された。

今回は、データを訓練用とテスト用にのみ分割していたが、今後は、訓練・検証・テストデータに分け本実験を行う。また、現在は学習で用いる項目が単純な情報の有無にとどまっているが、特に付帯情報の詳細記述に関しては文章を tf-idf などの手法を用いてベクトル化し学習させるなど、より精度を上げるような機械学習用データの生成に取り組む。

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」
分担研究報告書(令和2年度)

死亡に関わる調査票情報提供に基づいた ICD10 コード自動付与ツールの作成

研究分担者 香川璃奈 (筑波大学医学医療系・講師)

研究要旨

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。診療報酬請求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。

本研究では、死因確定作業において目視確認に回る理由の中でも、死因を ICD-10 コードに変換できないという点に焦点をあて、標準病名と完全一致しない死因の記載を標準病名または ICD-10 コードに変換するルールを作成した。その結果をプログラムに実装し、平成 27 年～令和 2 年の死亡票とオンライン申請された死亡個票を突合した情報に ICD10 コードを自動付与したところ、作成したルールを利用しないとときと比較して、死因の記載のすべてに対して ICD10 コードを付与できた件数の割合が 57.3%から 80.06%に増加した。

来年度はこの結果に基づき機械学習手法の開発を行う。

A. 研究目的

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。診療報酬請求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。

我々は令和元年度に平成 27 年～平成 30 年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結合を行なった。結合した情報のことを、以下、突合死亡票 DB(データ数: 5, 169, 031 件)と呼ぶ。これを利用して、標準病名マスターを用いて、全ての I 欄・II 欄病名に対しほぼ原記載のまま、また助詞、接続詞の除去/展開と言い換えなどの比較的簡便な文字列処理を施

すことで、約 65%の I 欄・II 欄病名の自動 ICD10 コーディングが可能であるという感触を得た。さらに、I 欄・II 欄病名を ICD10 コードに変換できたものは約 9 割であった。しかし上記は、突合死亡票 DB の一部を目視で確認した結果である。

そこで本年度は以下の 3 点を行なった。

(1) 上記の結果をより詳細に解析した。これにより、標準病名と完全一致しない死因の記載を標準病名または ICD-10 コードに変換するために有効なルールを作成した。

(2) 解析結果に基づき、比較的簡便な文字列処理に基づき I 欄・II 欄病名の自動 ICD10 コーディングを行うツールを開発した。

(3) 開発ツールを利用して実際に突合死亡票

DB の全ての I 欄・II 欄病名を ICD10 コードに変換した。

この際に、(1)で作成したルールを利用した際と、利用しなかった際とで、ICD10 にコーディングできた件数がどれだけ変化したか確認した。

B. 研究方法

(1) I 欄・II 欄病名が標準病名と完全一致しない場合の対応ルールの作成

(1-1) I 欄・II 欄病名と標準病名の対応の抽出

I 欄・II 欄病名のうち、標準病名に完全一致しない病名を自然言語記載病名と呼ぶ。

(1-1-1) 自然言語記載病名の書き換えルール作成

漢字の誤記、カタカナの表記ゆれ、表記を漢字、ひらがな、カタカナ、英語の間で任意にすれば標準病名になるパターンに対しては、書き換えルールを作成し適用し ICD10 コードを自動付与できるようにすることで、IRIS による自動確定死因付与をより多くの症例に対して行うことが可能になる。これにより、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)による原死因変更の有無の把握およびそれに基づく予測をより高精度に行うことができる。

しかし、突合死亡票 DB の全データの確認にはコストがかかる。そこで、突合死亡票 DB の冒頭 30,000 件と最後 30,000 件の合計 60,000 件の I(ア)欄に記載された自然言語記載病名を目視で確認し、漢字の誤変換、カタカナの表記ゆれ、表記を漢字、ひらがな、カタカナ、英語の間で任意にすれば標準病名に変換されるルールを作成した。

自然言語記載病名のうち、複数の病名が並記されている記載は確認対象から除外した。

(1-1-2) 出現頻度の高い自然言語記載病名に対応する ICD10 コードの探索

(1-1)で作成した変換ルールを適用しても標準病名に変換できない自然言語記載病名のう

ち、特に出現頻度の高い自然言語記載病名に対して ICD10 コードを付与できれば、IRIS による自動確定死因付与をより多くの症例に対して行うことが可能になる。これにより、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)による原死因変更の有無の把握およびそれに基づく予測をより高精度に行うことができる。

そこで以下の作業を行なった。

まず、突合死亡票 DB の冒頭から 50 万行の中で、標準病名と完全一致しなかった自然言語記載病名の上位 10 種類を特定した。該当する 10 種類の自然言語記載病名の各々に対して、突合死亡票 DB において該当する自然言語記載のみが I(ア)欄に記載されており、その他の自然言語記載病名が存在しない死亡表の確定病名として付与されている ICD10 コードを、該当する自然言語記載病名に付与される ICD10 コードとみなした。

(1-2) 死亡表の特徴に基づく表記の前処理ルールの作成

(i) 表記の統一

カタカナ、数字とアルファベットを全て半角に統一した。

(ii) 死亡表の特徴に基づく表記の前処理

死亡表の病名欄には、ICD10 コードを付与する上では不要な詳細な情報(発症の経緯や、病期分類など)が記載される。そこで、以下の前処理を行った。

「(」から「)」までの文字列は削除した。

「(」からはじまり最後まで「)」が出てこない場合にも「(」のあとの文字列は削除した。自然言語病名が「)」からはじまる場合は、「)」のみ削除した。

(iii) 表記揺れの回収

日本医学会：医学用語管理/付表 1 日本語表記のゆれ[1]に基づいた。[1]における「その他の表記法」の記載を全て「本辞典で採用した表記法」に変換した。

例外処理：

(a) 「その他の表記法」に複数の単語が並列して記載されていた場合には、すべての単語を、対応する「本辞典で採用した表記法」に変換する。
(b) 6. 異なった用語のあるものにおいて、用語

の意味が「【旧】」のように「【】」を利用して記載されている場合には、「【】から「】」までは除外する。

(c)「本辞典で採用した表記法」に複数の用語が記載されている場合は、今回はまずは便宜的に、死亡者数が多いと考えられる方に限定した。

たとえば、「知的障害【小児】、精神遅滞【神経】」は(b)の処理も踏まえて「精神遅滞」のみにした。

(d)「その他の表記法」および「本辞典で採用した表記法」に複数の文字列を意味する「・・・」がある場合、それを除外して変換する。

たとえば、「・・・パシー」を「・・・パチー」に変換する。これによりミオパシーがミオパチーに変換される。「・・・」を除外して変換すると、仮にテレパシーという記載があったとするテレパチーに変換される。しかし、この変換は標準病名とのマッチに影響しないため問題ないと考えた。

(iv) 複数の病名が併記されている時の病名分割ルール

個別の病名欄に複数の病名が併記されている場合がある。この場合は、全角空白、および、読点により病名が区切られていると解釈した。

(2) 自動 ICD10 コーディングツールの作成

(1)の処理を行った上で、自然言語記載病名を ICD10 コードに変換するプログラムを作成した。

自然言語病名と ICD10 対応標準病名マスター ver5.00[2]の標準病名が完全一致した場合に、自然言語病名を ICD10 コードに変換した。

(3) 突合死亡票 DB の I 欄・II 欄病名の ICD10 コードへの変換

(1)のルールを適用せずに ICD10 コードへの対応を行う処理、および、(1)のルールを適用したあとに ICD10 コードへの対応を行う処理のそれぞれを実施した。

C. 研究結果

(1) I 欄・II 欄病名が標準病名と完全一致しない場合の対応ルールの作成

(1-1) I 欄・II 欄病名と標準病名の対応の抽出

(1-1-1) 自然言語記載病名の書き換えルール

作成

ユニーク数で 4,481 個(延べ 14,747 個)の I(ア)欄の自然言語記載病名が標準病名と完全一致しなかった。全て目視で確認を行い、140 個の変換ルールを追加した。

変換ルールの一部は以下の通りである。全ての変換ルール、および(1-2)(iii)で作成した変換ルールは別添資料 4「記載変換ルール」に示す通りである。

表 1: 変換ルールの例。before 列の文字列を after 列に置換することを意味する

before	after
消化管	消化管
縊首	首吊り自殺
胆管Ca	胆管癌
Ab血症	アルブミン血症

(1-1-2) 出現頻度の高い自然言語記載病名に対応する ICD10 コードの探索

得られた自然言語記載病名の上位 10 種類と、対応する ICD10 コードは以下の通りである。

なお、「不詳の内因死」という自然言語記載病名が特定されたが、これは「不詳」(表 2)に内包した。

表 2: 出現頻度の高い自然言語記載病名に対応する ICD10 コード。(YD+)は 1 個以上の任意の数の数字以外の文字が存在するという意味である。

自然言語記載病名	ICD10 コード
不詳	R99
不明	R99
縊死	X70
急性心臓死	I461
慢性心不全急性増悪	I509
低酸素脳症	G931
誤嚥	T179
致死性不整脈	I498
虚血性心不全	I255

なお、この結果を確認したあと、(2)では、変換ルール(別添資料 5「未コード化傷病名変換ルール」)を利用した。

(1-2) 死亡表の特徴に基づく表記の前処理ルールの作成

B. 研究方法に記載の通りのルールを作成した。(iii)の結果は先述の通り別添4に示す通りである。その他のルールの具体例は本統括・分担報告書全体の末尾にある別添資料「Iris 入力データ作成プログラム」を確認されたい。

(2)自動 ICD10 コーディングツールの作成

(1)の結果に基づき、自動 ICD10 コーディングツールを作成した。また (1)のルールを適用せずに ICD10 コードへの対応を行う処理と、(1)のルールを適用したあとに ICD10 コードへの対応を行う処理について結果を比較した。

(3)突合死亡票 DB の I 欄・II 欄病名の ICD10 コードへの変換

(1)のルールを適用せずに ICD10 コードへの対応を行う処理を行ったところ、I 欄・II 欄病名のすべての自然言語記載病名に ICD10 コードに変換できたのは 2,960,455 件(57.27%)であった。

(1)のルールを適用したあとに ICD10 コードへの対応を行う処理を行ったところ、I 欄・II 欄病名のすべての自然言語記載病名に ICD10 コードに変換できたのは 4,138,310 件(80.06%)であった。

(4)倫理面への配慮

本研究では統計法 33 条に基づき申請したデータを利用した。申請の通り、インターネットに繋がらない端末上でのみデータの閲覧作業を行うことで個人情報に配慮した。

D. 考察

今回の処理を通じて、医師が手書きで記載した死亡診断書に基づき、電子的な死亡票データに変換する作業、すなわちオートコーディングの前段階として行う前処理に関して以下の課題が明らかとなった。

作成した変換ルールは、そもそもの死亡診断書の入力記載が標準化されれば確認と作成の手間を除外できるものである。具体的には、「急性」「慢性」などの修飾語句が付与されていた自然言語記載病名、修飾語句に「(」 「)」などの記号を独自に付与している自然言語記載

病名、そもそも記載欄や記載内容を間違えているように解釈できた自然言語記載病名(例:「脾臓」)、自然言語記載病名が手書きだからこその誤記や省略表現(例:「下日支」「脳卒中」)などである。医師の死亡診断書記入時における標準病名の利用、あるいはその電子データへの転記時での標準病名への変換が実現すれば、このような入力、また入力に基づく解釈に人手を割く必要がなくなり、働き方改革にも貢献できると考える。

また、将来的に日本全国で死亡表の自然言語病名が ICD10 コードへの変換がオフライン上で安全に行われるようになる将来を見据えて、(2)自動 ICD10 コーディングツールの作成で作成したツールの一部の機能を docker コンテナとしても実装できることを確認した。

E. 結論

本年度研究では、実際の死亡表に記載された病名を ICD10 コードに変換するツールを作成した。独自の対応ルールを利用することで、I 欄・II 欄病名のすべての自然言語記載病名に ICD10 コードに変換できた件数が全体の 8 割を超えた。来年度はこの結果に基づき機械学習手法の開発を行う。

<参考文献>

[1] http://jams.med.or.jp/dic/kanji_variance2.html (2019年12月26日閲覧)

[2]

<https://www2.medis.or.jp/stdcd/byomei/download2019.html> (2019年12月26日閲覧)

F. 健康危険情報

なし

G. 研究発表

なし

H. 知的財産権の出願・登録状況

なし

別添資料4 記載変換ルール

before	after
アテトーゼ	アテトース
アミロイド症	アミロイドーシス
アンギナ	アンギーナ
アンドロジェン	アンドロゲン
いおう	イオウ
イボ	いぼ
インシュリン	インスリン
えい児	嬰兒
エストロジェン	エストロゲン
えら	鰓
えん下	嚥下
え死	壊死
おとがい	頤
おり	檻
カ	蚊
かいせん	疥癬
かい癬	疥癬
かかと	踵
がこう瘡	鷺口瘡
がん	癌
ガン	癌
くし	櫛
グルタミン酸オキサロ酢酸トランスフェラーゼ (GOT)	アスパラギン酸アミノトランスフェラーゼ (AST)
グルタミン酸ピルビン酸トランスフェラーゼ (GPT)	アラニンアミノトランスフェラーゼ (ALT)
くん煙	燻煙
コンジローム	コンジローマ
さや	莢
ざんごう	塹壕
ざん壕	塹壕
ジスキネジー	ジスキネジア
ジストニー	ジストニア
しゃ血	瀉血
しゅう酸	シュウ酸
じゆく状	粥状
じょくそう	褥瘡
じんま疹	蕁麻疹
し緩	弛緩
し骨	篩骨
し眠	嗜眠
すう勢	趨勢
ぜい弱	脆弱
せき	咳
せん	癬
せんもう	せん妄
そう	叢

そう病	躁病
そう痒	搔痒
そしゃく	咀嚼
そ径部	鼠径部
たばこ	タバコ
たん	痰
たんぱく質	タンパク質
つる	蔓
でん部	殿部
でん粉	デンプン
てん絡	纏絡
とう骨	橈骨
どもる	吃る
のう	囊
ばい煙	煤煙
はさみ	鋏
パシー	パチー
ハンチントン舞蹈病	ハンチントン病
ひとみ	瞳
ひょうせつ	剽窃
ひ骨	腓骨
ひ素	ヒ素
び漫性	びまん性
ふるい	篩
ふ化	孵化
べんち	胼胝
ポリプ	ポリープ
まつげ	睫毛
ミオトニー	ミオトニア
よう素	ヨード
らい病	ハンセン病
ラセン	らせん
り患	罹患
るい瘦	るいそう
レセプター	受容体
れんさ	連鎖
レンサ	連鎖
阿片	アヘン
暗きよ	暗渠
煙草	タバコ
横痃	よこね
牡蠣	カキ
可そ性	可塑性
鈎	鈎
覚せい	覚醒
間歇	間欠
顔ぼう	顔貌
砧骨	きぬた骨

交叉	交差 (交叉)
骨粗しょう	骨粗鬆
骨髄	骨格
産じょく	産褥
子かん	子癩
屍体	死体
漆	ウルシ
遮へい	遮蔽
充てん	充填
上皮小体	副甲状腺
色盲	色覚異常
食慾	食欲
新生物	腫瘍
震顫	振戦
靱帯	靱帯
洗滌	洗浄
繊維	線維
絨毛	線毛
鼠蹊部	鼠径部
蛋白質	タンパク質
痴呆	認知症
蜘蛛膜	くも膜
蔦	ツタ
殿粉	デンプン
澱粉	デンプン
尿路性器系	腎尿路生殖器系
函数	関数
非定型抗酸菌	非結核性抗酸菌
副睪丸	精巣上体
米ぬか	米糠
偏頭痛	片頭痛
弁逆流	弁閉鎖不全
縫帯	包帯
蜂窩織炎	蜂巣炎
沃素	ヨウ素
沃素	ヨード
螺旋	らせん
硫黄	イオウ
湾曲	彎曲
尙儂病	くる病
尙儂病	くる病
喘ぎ	あえぎ
嵌入	陥入
彎	弯
攪	攪
鬱血	うっ血
鬱病	うつ病
瀰漫性	びまん性

畸形	奇形
疝痛	仙痛
疣	いぼ
癩菌	らい菌
癩癧	てんかん
辜丸	精巣
砒素	ヒ素
糜爛	びらん
羸瘦	るいそう
羸?	るいそう
腔	腔
臀部	殿部
莓	莓
蓆酸	シュウ酸
蔗糖	ショ糖
蛔虫	回虫
訥る	吃る
頸	頸
鼠	鼠
鼠径部	鼠径部
齲齒	う齒
齲食	う食
精神薄弱	知的障害
精神薄弱	精神遅滞
精神病院	精神科病院
精神分裂病	統合失調症
_囊	そ囊
_氣	あい氣
_	啣
_	囊
_	囊
_	搔
_瘡	ぞ瘡
_痒	搔痒
_	瘦
_疽	ひょう疽
_	せつ
_	軀
_	齟
癰	よう
譫妄	せん妄
日支	肢
臍ぞう	臍臓
鬱血	うっ血
うつ血	うっ血
間塞	閉塞
皮フ	皮膚
懷死	壊死

委縮	萎縮
褥創	褥瘡
心源性	心原性
心厚性	心原性
脳硬塞	脳梗塞
消化管	消化管
消火管	消化管
壊疽	壊疽
姓	性
のがん	癌
悪性新生物	癌
イ頸	首吊り自殺
いっ頸	首吊り自殺
縊頸	首吊り自殺
縊首	首吊り自殺
縊死	首吊り自殺
?型	2型
?型	1型
StanfordA	大動脈解離StanfordA
大動脈乖離	大動脈解離
A型大動脈解離	大動脈解離StanfordA
ナルコース	ナルコース
ラルコース	ナルコース
憎悪	増悪
アルツハイマー認知症	アルツハイマー型認知症
くも膜出血	くも膜下出血
くま膜下出血	くも膜下出血
クモ膜下出血	くも膜下出血
みぎ	
ひだり	
右	
左	
低酸素脳症	低酸素性脳症
うっ血性心不全	うっ血性心不全
うっ血心不全	うっ血性心不全
うっけつ性心不全	うっ血性心不全
ㇿ血性心不全	うっ血性心不全
えん下性肺炎	嚥下性肺炎
がん悪液質	癌性悪液質
がん性	癌性
クワイフェルトヤコブ病	クワイフェルトヤコブ病
ごえんせい肺炎	誤嚥性肺炎
肺炎	肺炎
(誤嚥性)肺炎	誤嚥性肺炎
びまん性大細胞型B細胞リンパ腫	びまん性大細胞型B細胞性リンパ腫
びまん性大細胞B細胞性リンパ腫	びまん性大細胞型B細胞性リンパ腫
びまん性大細胞性B細胞性リンパ腫	びまん性大細胞型B細胞性リンパ腫
びまん性大細胞性B細胞リンパ腫	びまん性大細胞型B細胞性リンパ腫

びまん性大細胞柄B細胞リンパ腫	びまん性大細胞型B細胞性リンパ腫
一酸化炭素中毒死	一酸化炭素中毒
急性一酸化炭素中毒死	一酸化炭素中毒
急性一酸化炭素中毒	一酸化炭素中毒
COPD急性増悪	慢性呼吸不全急性増悪
胃(\D+)癌	胃癌
悪期	悪性
下顎歯肉(\D+)癌	下顎歯肉癌
両側肺炎	肺炎
両肺炎	肺炎
線癌	腺癌
(\D+)非浸潤性乳管癌	非浸潤性乳管癌
([^非]+)浸潤性乳管癌	浸潤性乳管癌
患性	悪性
第動脈	大動脈
解離性大動脈流	解離性大動脈瘤
拡張性心筋症	拡張型心筋症
幹細胞癌	肝細胞癌
感染症心内膜炎	感染性心内膜炎
胆肝癌	胆管癌
胆管	胆管
管内胆管癌	肝内胆管癌
肝間部胆管癌	肝門部胆管癌
肝質性肺炎	間質性肺炎
肝門部(\s+)胆管癌	肝門部胆管癌
肝門部(\D+)胆管癌	肝門部胆管癌
間接性肺炎	間質性肺炎
悪質液	悪液質
癌悪液質	癌性悪液質
癌悪液質	癌性悪液質
腹膜園	腹膜炎
急性くも膜下出血	くも膜下出血
急性一酸化炭素中毒	一酸化炭素中毒
急性呼吸(\D+)不全	急性呼吸不全
白血症	白血病
氏不全	心不全
急性室不全	急性心不全
新不全	心不全
筋萎縮性	筋萎縮性
筋萎縮性側管硬化症	筋萎縮性側索硬化症
筋萎縮性側策硬化症	筋萎縮性側索硬化症
筋萎縮側索硬化症	筋萎縮性側索硬化症
加仏	K
高K血圧	高K血症
骨髓異形性症候群	骨髓異形成症候群
骨髓異型性症候群	骨髓異形成症候群
骨髓異型成症候群	骨髓異形成症候群
骨髓異形症候群	骨髓異形成症候群

骨髓異形性症候群	骨髓異形成症候群
骨髓異常形成症候群	骨髓異形成症候群
骨髓性異形成症候群	骨髓異形成症候群
腺維	線維
腎盂	腎盂
上向結腸癌	上行結腸癌
腸管膜	腸間膜
心原性脳梗塞症	心原性脳梗塞
心原性脳梗塞症	心原性脳梗塞
心原性脳梗塞症	心原性脳梗塞
多系統萎縮\$	多系統萎縮症\$
胆管Ca	胆管癌
Na血症	ナトリウム血症
Ab血症	アルブミン血症
低体温症	低体温
特発性肺腺症	特発性肺線維症
脳卒中	脳卒中
播腫	播種
播種性血管内凝固症	播種性血管内凝固
播種性血管内凝固異常	播種性血管内凝固
播種性血管内凝固異常	播種性血管内凝固
播種性血管内凝固症候群	播種性血管内凝固
廃用性症候群	廃用症候群
癯用症候群	廃用症候群
敗血栓	敗血症
敗血病	敗血症
肺血症	敗血症
非閉塞性腸間膜虚血	非閉塞性腸間膜虚血
非閉塞性腸管虚血	非閉塞性腸管虚血
非閉塞性腸管膜虚血	非閉塞性腸管膜虚血
腹膜Ca	腹膜癌
慢性腎不全	慢性腎不全
老すい	老衰
老スイ	老衰
嚥下傷害	嚥下障害
攝食障害	摂食障害
濾胞性悪性リンパ腫	濾胞性リンパ腫
膠芽種	膠芽腫

別添資料5 未コード化傷病名変換ルール

不詳	R99
不明	R99
縊死	T71
急性心臓死	I461
慢性心不全急性増悪	I509
低酸素脳症	G931
不詳の内因死	R99
誤嚥	T179
致死性不整脈	I498
虚血性心不全	I255
左乳癌	C509
右乳癌	C509
なし	R99
縊頸	T71
クモ膜下出血	I609
慢性心不全の急性増悪	I509
右肺癌	C349
慢性心不全増悪	I509
大動脈解離	I710
虚血性心疾患（推定）	I259
脳幹出血	I613
虚血性心疾患の疑い	I259
心臓死の疑い	I461
不詳の病死	R99
縊頸	T71
左肺癌	C349
蘇生後脳症	G931
重症肺炎	J189
心疾患の疑い	I519
急性心筋梗塞の疑い	I219

III. 研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年	ページ
なし							

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
明神 大也, 大井川 仁美, 香川 璃奈, 今村 知明, 今井 健	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	医療情報学	40(Suppl.)	674-676	2020
大井川 仁美, 明神 大也, 香川 璃奈, 今村 知明, 今井 健	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	医療情報学	40(Suppl.)	677-682	2020
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 の 指標	2020;67(3)	17-24	2020

別添資料

以下、本統括・分担研究報告書全体に係る別添資料を示す。
順に、以下の構成となっている。

- Iris 用入力データ生成プログラムソース（主要な部分抜粋）
 - メインプログラム (parse.pl)
 - 処理用ライブラリソース
 - ◇ GKAI::Coding.pm
 - ◇ GKAI::Kikan.pm
 - ◇ GKAI::Util.pm
 - Iris 入力データ作成用 日本独自 ICD コード変換ルール
 - Iris 入力データ作成用 期間表現正規化ルール

- 機械学習による原死因変更有無予測結果（一部抜粋）


```

1 #! /usr/bin/env perl
2
3 #=====
4 #==                                     ==#
5 #==          [ parse.pl ]              ==#
6 #==                                     ==#
7 #==                                     ==#
8 #=====
9
10 use warnings;
11 use v5.28;
12
13 use utf8;
14 use Encode;
15 use FindBin;
16 use lib "$FindBin::Bin/lib";
17 use GKAI::Util;
18 use GKAI::Coding;
19 use GKAI::Kikan;
20
21 binmode STDIN, ":utf8";
22 binmode STDOUT, ":utf8";
23
24 if ( $ARGV{HELP} ) {
25     &printHelp();
26 }
27
28 #== 初期設定 =====
29 my $switch = ""; # スイッチ
30 my $kikanrule = decode_utf8($FindBin::Bin) . '/rules/期間なおし_ver2.csv';
31 my $tikanrule = decode_utf8($FindBin::Bin) . '/rules/Iris用ICD変換.csv';
32 my $sdm = decode_utf8($FindBin::Bin) . '/rules/504-han-withICD.txt';
33 my $unicode = decode_utf8($FindBin::Bin) . '/rules/未code化傷病名用リスト.txt';
34 my $normalizekanji = decode_utf8($FindBin::Bin) . '/rules/記載変換ルール.txt';
35 my $dup = {};
36 my $gu = new GKAI::Util();
37 my $gc = new GKAI::Coding('SDM' => $sdm, 'ICDtikan' => $tikanrule, 'UNICODE' =>
38     $unicode, 'ICDtikan2' => $normalizekanji);
39 my $gk = new GKAI::Kikan('kikanRule' => $kikanrule);
40
41 #== 処理本体 =====
42
43 #== ALL
44 open IDENTALL, ">:utf8", "ALL/Ident.txt";
45 open MEDCODALL, ">:utf8", "ALL/MedCod.txt";
46
47 #== 5分割
48 open my $OUTI1, ">:utf8", "5分割/Ident1.txt";
49 open my $OUTI2, ">:utf8", "5分割/Ident2.txt";
50 open my $OUTI3, ">:utf8", "5分割/Ident3.txt";
51 open my $OUTI4, ">:utf8", "5分割/Ident4.txt";
52 open my $OUTI5, ">:utf8", "5分割/Ident5.txt";
53
54 open my $OUTM1, ">:utf8", "5分割/MedCod1.txt";
55 open my $OUTM2, ">:utf8", "5分割/MedCod2.txt";
56 open my $OUTM3, ">:utf8", "5分割/MedCod3.txt";
57 open my $OUTM4, ">:utf8", "5分割/MedCod4.txt";
58 open my $OUTM5, ">:utf8", "5分割/MedCod5.txt";
59
60 #== テスト用 (10件ごと抽出)
61 open TESTI, ">:utf8", "Test/Ident.txt";
62 open TESTM, ">:utf8", "Test/MedCod.txt";
63
64 #== テスト用 (500件ごと抽出)
65 open TESTIS, ">:utf8", "Test2/Ident.txt";
66 open TESTMS, ">:utf8", "Test2/MedCod.txt";
67
68 #== IRIS にかけたリスト
69 open IRIS, ">:utf8", "ALL/ForIris.txt";
70 open IRIS5, ">:utf8", "5分割/ForIris.txt";
71 open IRIST, ">:utf8", "Test/ForIris.txt"; # テスト用
72 open IRISTS, ">:utf8", "Test2/ForIris.txt"; # テスト用
73
74 #== 期間のパターン
75 open KIKAN, ">:utf8", "期間表現/Kikan_match_all.txt";
76
77 open IN, "<:utf8", "../01_NewShiboJoin/new_shibo_join.tsv";
78 my $c = 0;
79 my $c2 = 0;
80 while (my $str = <IN>) {
81     $c ++;
82     $c2 ++;
83 }

```

```

83 my $o1 = $OUT11;
84 my $o2 = $OUTM1;
85 if ($c <= 1000000){ $o1 = $OUTI1; $o2 = $OUTM1 }
86 elsif ($c <= 2000000){ $o1 = $OUTI2; $o2 = $OUTM2}
87 elsif ($c <= 3000000){ $o1 = $OUTI3; $o2 = $OUTM3}
88 elsif ($c <= 4000000){ $o1 = $OUTI4; $o2 = $OUTM4}
89 elsif ($c <= 5500000){ $o1 = $OUTI5; $o2 = $OUTM5}
90
91 if ($c != 0 ) {
92     my @a = split(/\t/, $str);
93
94     # 病名は 10,12,14,16,18
95     my $d = $gc->ICDCoding($a[10], $a[12], $a[14], $a[16], $a[18]);
96
97     # 期間は 11,13,15,17,19
98     my @kikan = ();
99     my $k1 = $gk->getKikan($a[11]);
100    my $k2 = $gk->getKikan($a[13]);
101    my $k3 = $gk->getKikan($a[15]);
102    my $k4 = $gk->getKikan($a[17]);
103    my $k5 = $gk->getKikan($a[19]);
104    @kikan = ($k1, $k2, $k3, $k4, $k5);
105
106    #-- 期間のDEBUG用 -----
107    my $dk1 = $gk->debug($a[11]);
108    my $dk2 = $gk->debug($a[13]);
109    my $dk3 = $gk->debug($a[15]);
110    my $dk4 = $gk->debug($a[17]);
111    my $dk5 = $gk->debug($a[19]);
112    if ($dk1 ne '') {say KIKAN $dk1}
113    if ($dk2 ne '') {say KIKAN $dk2}
114    if ($dk3 ne '') {say KIKAN $dk3}
115    if ($dk4 ne '') {say KIKAN $dk4}
116    if ($dk5 ne '') {say KIKAN $dk5}
117
118    # 全てにICDコードが当たった
119    if ($d->{HIT}){
120
121        my $key = $gu->keta($a[0]);
122
123        # Identテーブル
124        my @identArr = ();
125        for (my $i = 0; $i < 44; $i++){ $identArr[$i] = '' };
126        $identArr[0] = $key;
127        $identArr[2] = $gu->ynd($a[47]); # DateBirth (birth_ynd)
128        $identArr[3] = $gu->ynd($a[50]); # DateDeath (death_ynd)
129        $identArr[5] = $a[45]; # Sex (sex)
130        $identArr[6] = '0'; # MannerOfDeath ('0')
131        $identArr[9] = 'Initial'; # Status ('Initial')
132        $identArr[10] = 'No'; # Reject ('No')
133        $identArr[11] = 'Automatic'; # Coding ('Automatic')
134        my $identRes = join("\t", @identArr);
135
136        # ALL
137        print IDENTALL "$identRes\r\n";
138        # 5分割
139        print $o1 "$identRes\r\n";
140        # テスト
141        if ($c % 10 == 0){ print TESTI "$identRes\r\n" }
142        if ($c % 500 == 0){ print TESTIS "$identRes\r\n" }
143
144        # MedCodテーブル
145        for (my $i = 0; $i <= 4; $i++){
146            my @medcodArr = ();
147            for (my $j = 0; $j < 7; $j++){ $medcodArr[$j] = '' };
148            if($d->{DIS}[$i] ne ''){
149                my $icd = $d->{DIS}[$i];
150                # 期間がある場合付け加える。
151                if ($kikan[$i] ne '' and $kikan[$i] ne 'N/A'){ $icd .= '('
152                . $kikan[$i] . ' )' }
153                #----
154                my $medcodnum = 0;
155                if ($i <= 3) { $medcodnum = $i }
156                else { $medcodnum = 5 }
157                #----
158                $medcodArr[0] = $key;
159                $medcodArr[1] = $medcodnum;
160                $medcodArr[2] = '';
161                $medcodArr[3] = $icd;
162                $medcodArr[4] = '';
163                $medcodArr[5] = 1;
164                $medcodArr[6] = 1;
165                my $medcodRes = join("\t" , @medcodArr);

```

```

166         # ALL
167         print MEDCODALL "$medcodRes\r\n";
168         # 5分割
169         print $o2 "$medcodRes\r\n";
170         # テスト
171         if ($c % 10 == 0){ print TESTM "$medcodRes\r\n" }
172         if ($c % 500 == 0){ print TESTMS "$medcodRes\r\n" }
173     }
174 }
175
176 #== ForIris.txt にID記入
177 # ALL
178 say IRIS $a[0];
179 # 5分割
180 say IRIS5 $a[0];
181 # テスト
182 if ($c % 10 == 0){ say IRIST $a[0] }
183 if ($c % 500 == 0){ say IRISTS $a[0] }
184 }
185 # INFO
186 if ($c % 10000 == 0 ) { say STDERR "[INFO]: $c death certificates
processed." }
187 }
188 }
189 close(IN);
190
191
192 #== ALL
193 close(IDENTALL);
194 close(MEDCODALL);
195
196 #== 5分割
197 close($OUTI1);
198 close($OUTI2);
199 close($OUTI3);
200 close($OUTI4);
201 close($OUTI5);
202
203 close($OUTM1);
204 close($OUTM2);
205 close($OUTM3);
206 close($OUTM4);
207 close($OUTM5);
208
209 #== テスト用 (10件ごと)
210 close(TESTI);
211 close(TESTM);
212
213 #== テスト用 (5000件ごと)
214 close(TESTIS);
215 close(TESTMS);
216
217 #== IRIS にかけたリスト
218 close(IRIS);
219 close(IRIS5);
220 close(IRIST);
221 close(IRISTS);
222
223 #== 期間表現
224 close(KIKAN);
225
226

```

```

1 package GKAI::Coding;
2
3 #=====
4 #==                                     ==#
5 #==          [ GKAI::Coding ]         ==#
6 #==                                     ==#
7 #==                                     ==#
8 #=====
9
10 use warnings;
11 use v5.28;
12 use utf8;
13 use GKAI::Util;
14 use Encode;
15 use Carp qw(croak);
16 use Data::Dumper;
17
18 binmode STDIN, ":utf8";
19 binmode STDOUT, ":utf8";
20 binmode STDERR, ":utf8";
21
22 #== 初期設定 =====
23 our $VERSION = 1.00;
24 our %DEFAULT = ();
25
26 #== コンストラクタ =====
27 sub new {
28     my $class = shift;
29     my $this = {};
30     bless $this, $class;
31
32     #== 変数の設定 =====
33     my (%option) = @_;
34     my $u = new GKAI::Util;
35     $this->{UTIL} = $u;
36     $this->{SDM} = {};
37     $this->{TIKANRULE} = {};
38     $this->{UNCODE} = {};
39     $this->{TIKANRULE2} = {};
40
41     #== 初期設定 =====
42     $this->loadSDM($option{"SDM"}, $option{"UNCODE"});
43     $this->loadICDtikanRule($option{"ICDtikan"});
44     $this->loadICDtikanRule2($option{"ICDtikan2"});
45     return $this;
46 }
47
48 #== デストラクタ =====
49 sub DESTROY {
50     my $this = shift;
51 }
52
53 #== Methods =====
54
55 #-----
56 # [loadSDM]
57 #-----
58 sub loadSDM{
59     my $this = shift;
60     my ($in, $in2) = @_;
61     open SDM, "<:utf8", $in or croak("[ERROR]: cannot load SDM");
62     say STDERR '[INFO]: 標準病名マスター (' . $in . ") loaded";
63     while (<SDM>){
64         chomp;
65         my ($d, $i) = split(/\t/);
66         $this->{SDM}{$d} = $i;
67     }
68     close(SDM);
69
70     open UNCODE, "<:utf8", $in2 or croak("[ERROR]: cannot load UNCODE");
71     say STDERR '[INFO]: 未コード化傷病名用マスター (' . $in2 . ") loaded";
72     while (<UNCODE>){
73         chomp;

```

```

74     my ($d, $i) = split(/\t/);
75     $this->{SDM}{$d} = $i;
76 }
77 close(UNCODE);
78 }
79
80 #-----
81 # [loadICDtikanRule]
82 #-----
83 sub loadICDtikanRule{
84     my $this = shift;
85     my ($in) = @_ ;
86     my $info = {};
87     $info->{NUM} = 0;
88     $info->{PATH} = $in;
89     say STDERR "[INFO]: ICD10コード強制置換ルール (" . $info->{PATH} . ")
loading...";
90     my $csv = Text::CSV::Encoded->new({
91         encoding_in => 'UTF-8',
92         encoding_out => 'UTF-8'
93     });
94     open my $IN, '<:utf8', $in or die "cannot open $in !";
95     while (my $fields = $csv->getline($IN)){
96         # 現在有効なルールのみ
97         if ($fields->[4] =~ /^(1|2)$/){
98             $info->{NUM} ++;
99             my $p = $fields->[0];
100             $this->{TIKANRULE}{$p}{FREQ} = $fields->[2];
101             $this->{TIKANRULE}{$p}{NEW} = $fields->[1];
102         }
103     }
104     close($IN);
105
106     say STDERR "[INFO]: 有効ICD10コード強制置換ルール数 " . $info->{NUM};
107 }
108
109
110 sub loadICDtikanRule2{##異字体の正規化
111     my $this = shift;
112     my ($in) = @_ ;
113     my $num = 0;
114     open my $fh, '<:encoding(utf8)', $in; # auto decoding on read
115     while (my $line = <$fh> ) {
116         chomp ($line);
117         Encode::_utf8_off($line);
118         $line= decode('UTF-8', $line);
119         my @d=split(/\t/, $line, -1);
120         next if (length($d[0])<1 );
121         $num++;
122         $this->{TIKANRULE2}{$num}{before}=$d[0];
123         $this->{TIKANRULE2}{$num}{after}=$d[1];
124     }
125     $this->{TIKANRULE2}{NUM}=$num;
126     close($fh);
127 }
128
129
130 sub tikanICD2{
131     my $this = shift;
132     my ($in) = @_ ;
133     for my $i(1..$this->{TIKANRULE2}{NUM})
134     {
135         my $before=$this->{TIKANRULE2}{$i}{before};
136         my $after=$this->{TIKANRULE2}{$i}{after};
137         $before=quotemeta $before;
138         $after=quotemeta $after;
139         if ($in =~ /$before/) {
140             $in =~ s/$before/$after/g;
141         }
142     }
143     return $in;
144 }
145

```

```

146
147 #-----
148 # [tikanICD]
149 #-----
150 sub tikanICD{
151     my $this = shift;
152     my ($in) = @_ ;
153     my $ret = '';
154     if (exists($this->{TIKANRULE}{$in})) {
155         $ret = $this->{TIKANRULE}{$in}{NEW};
156     }
157     else {
158         $ret = $in;
159     }
160     return $ret;
161 }
162
163
164 #-----
165 # [byomei2icd] 病名から何らかのICDコードを返す。なければ空文字
166 #-----
167 sub byomei2icd{
168     my $this = shift;
169     my ($in) = @_ ;
170     my $code = '';
171     my $ret = '';
172
173     my $sdm = $this->{SDM};
174
175     #--- 仮のICDコードを引く
176     # 標準病名 + 未コード化傷病名用セットを引く
177     if (exists($sdm->{$in})) {
178         $code = $sdm->{$in};
179     }
180
181     # 部分一致の場合
182     else {
183         $code = ''; #*
184     }
185
186     #--- IRIS入力用にICDコードを無理矢理変更
187     $ret = $this->tikanICD($code);
188     return $ret;
189 }
190
191 #-----
192 # [ICDCoding]
193 #-----
194 #-----
195 sub ICDCoding{
196     my $this = shift;
197     my @a = @_ ;
198     my $ret = {};
199     my $u = $this->{UTIL};
200     my $sdm = $this->{SDM};
201     $ret->{HIT} = 1;
202     $ret->{DIS} = [];
203     $ret->{UNKNOWN} = {};
204     for (my $i = 0; $i < @a; $i++){
205         my $byomei = $a[$i];
206         $ret->{DIS}[$i] = '';
207
208         # 前後の全角/半角空白の連続、「、」の前後の全角空白のみを除去
209         my $tmp1 = $u->rmSpaceSimple($byomei);
210
211         # カッコで囲まれた部分を除外する
212         # 括弧を取る
213         $tmp1 =~ s/<.*?>//g if( $tmp1 );
214         $tmp1 =~ s/\(.*?\)//g if( $tmp1 );
215         $tmp1 =~ s/(.*?) //g if( $tmp1 );
216         $tmp1 =~ s/[.*?] //g if( $tmp1 );
217         $tmp1 =~ s/(.*?) //g if( $tmp1 );
218

```

```

219 # スタート文字の場合削除
220 my @mark1=((">","\)"," " ",") " ",") " ");
221
222 foreach my $m (@mark1){
223     $m=quotemeta $m;
224     $tmp1=~s/^\$m//g if( $tmp1 );
225 }
226
227 # 閉じ括弧がなければ全部削除
228 my @mark2=("<",'('," ("," ["," (");
229
230 foreach my $m (@mark2){
231     $m=quotemeta $m;
232     $tmp1 =~ s/$m.*$//g if( $tmp1 );
233 }
234
235 $tmp1 = $this->tikanICD2( $tmp1 );
236
237 # 空文字列でなければ
238 if ( $tmp1 ne '' ) {
239
240     my $res = ''; # ICD Coding 結果
241
242     #--- パターン解析用
243     #my $pattern = $tmp1;
244     #$pattern =~ s/[^\ \. ]/x/g;
245     #$pattern =~ s/x+/x/g;
246     #say $pattern;
247
248     #--- 複雑なパターン
249     # 全角空白あり 読点なし
250     if ( $tmp1 =~ / / and $tmp1 !~ /、 / ) {
251
252         my @b = split(/ +/, $tmp1);
253         for (my $i = 0; $i < @b; $i++) {
254             my $tmp2 = $u->zen2han($b[$i]);
255             my $tmp3 = $this->byomei2icd($tmp2);
256             if ( $tmp3 ne '' ){
257                 $res .= ','. $tmp3;
258             }
259             else {
260                 $res .= ',*';
261             }
262         }
263     }
264
265     # 全角空白なし 読点あり
266     elsif ( $tmp1 !~ / / and $tmp1 =~ /、 / ) {
267         #先頭と最後の読点は削除
268         $tmp1 =~ s/^\ //g;
269         $tmp1 =~ s/、 $//g;
270         my @b = split(/、 +/, $tmp1);
271         for (my $i = 0; $i < @b; $i++) {
272             my $tmp2 = $u->zen2han($b[$i]);
273             my $tmp3 = $this->byomei2icd($tmp2);
274             if ( $tmp3 ne '' ){
275                 $res .= ','. $tmp3;
276             }
277             else {
278                 $res .= ',*';
279             }
280         }
281     }
282
283     #--- 残りは"1病名"とみなす
284     else {
285         my $tmp2 = $u->zen2han($tmp1);
286         my $tmp3 = $this->byomei2icd($tmp2);
287         if ( $tmp3 ne '' ){
288             $res .= ','. $tmp3;
289         }
290         else {
291             $res .= ',*';
292         }
293     }

```

```
292     }
293   }
294   my $tmpres = substr($res, 1);
295
296   # *を含んだら HIT を 0にする
297   if ($tmpres =~ /*/){
298     $ret->{HIT} = 0;
299   }
300   $ret->{DIS}[$i] = $tmpres;
301 }
302 }
303 return $ret;
304 }
305
306 1;
307
308
```



```

1 package GKAI::Kikan;
2
3 #=====
4 #==                                     ==#
5 #==          [ GKAI::Kikan ]          ==#
6 #==                                     ==#
7 #==                                     ==#
8 #=====
9
10 use warnings;
11 use v5.28;
12 use utf8;
13 use Encode;
14 use Carp qw(croak);
15 use Text::CSV::Encoded;
16 use GKAI::Util;
17
18 binmode STDIN, ":utf8";
19 binmode STDOUT, ":utf8";
20 binmode STDERR, ":utf8";
21
22 #== 初期設定 =====
23 our $VERSION = 1.00;
24 our %DEFAULT = ();
25
26 #== コンストラクタ =====
27 sub new {
28     my $class = shift;
29     my $this = {};
30     bless $this, $class;
31
32     #== 変数の設定 =====
33     my (%option) = @_ ;
34     $this->{KIKANRULE} = {};
35     $this->{UTIL} = new GKAI::Util();
36
37     #== 初期設定 =====
38     $this->loadKikanRule($option{"kikanRule"}); #
39
40     #=====
41     return $this;
42 }
43
44 #== デストラクタ =====
45 sub DESTROY {
46     my $this = shift;
47 }
48
49 #== Methods =====
50
51 #-----
52 # [getKikan]
53 #-----
54 sub getKikan{
55     my $this = shift;
56     my ($in) = @_ ;
57     my $u = $this->{UTIL};
58     my $in2 = $u->rmSpaceSimple($in);
59     my $str = $u->zen2han($in2);
60     my $p = $this->getKikanPattern($in);
61     my $ret = '';
62     if ($str ne '') {
63         if (exists($this->{KIKANRULE}{$p})) {
64             my $f = $this->{KIKANRULE}{$p}{FREQ};
65             my $v = $this->{KIKANRULE}{$p}{VAR};
66             my $v2 = $v;
67             my $u = $this->{KIKANRULE}{$p}{UNIT};
68             my $r = '';
69             my @a = (); # 元の数値表現
70             while ($str =~ /\d{1,}/g){ push(@a, $&) }
71             my @b = (); # 正規化数値表現 ('09分' 等対策)
72             for (my $i = 0; $i < @a; $i++){
73                 my $tmp = $a[$i] + 0;

```

```

74         push(@b, $tmp);
75     }
76     if ($v2 =~ /\$1/){ $v2 =~ s/\$1/$b[0]/ }
77     if ($v2 =~ /\$2/){ $v2 =~ s/\$2/$b[1]/ }
78     if ($v2 =~ /\$3/){ $v2 =~ s/\$3/$b[2]/ }
79     if ($v2 ne ''){ $r = eval($v2)};
80     my $r2 = $r . $u;
81     if ($r2 ne '' or $r2 ne 'N/A') {$ret = $r2}
82 }
83 }
84 return $ret;
85 }
86
87 #-----
88 # [debug]
89 #-----
90 sub debug{
91     my $this = shift;
92     my ($in) = @_ ;
93     my $u = $this->{UTIL};
94     my $in2 = $u->rmSpaceSimple($in);
95     my $str = $u->zen2han($in2);
96     my $p = $this->getKikanPattern($in);
97     my $ret = '';
98
99     if ($str ne '') {
100         if (exists($this->{KIKANRULE}{$p})) {
101             my $f = $this->{KIKANRULE}{$p}{FREQ};
102             my $v = $this->{KIKANRULE}{$p}{VAR};
103             my $v2 = $v;
104             my $u = $this->{KIKANRULE}{$p}{UNIT};
105             my $r = '';
106             my @a = (); # 元の数値表現
107             while ($str =~ /\d{1,}/g){ push(@a, $&) }
108             my @b = (); # 正規化数値表現 ('09分' 等対策)
109             for (my $i = 0; $i < @a; $i++){
110                 my $tmp = $a[$i] + 0;
111                 push(@b, $tmp);
112             }
113             if ($v2 =~ /\$1/){ $v2 =~ s/\$1/$b[0]/ }
114             if ($v2 =~ /\$2/){ $v2 =~ s/\$2/$b[1]/ }
115             if ($v2 =~ /\$3/){ $v2 =~ s/\$3/$b[2]/ }
116             if ($v2 ne ''){ $r = eval($v2)};
117             my $r2 = $r . $u;
118             $ret = $str . "\t" . $p . "|" . $f . "|" . $v . "|" . $u . "\t" .
119 $v2 . "|" . $r . "\t" . $r2;
120         }
121     }
122     return $ret;
123 }
124 #-----
125 # [loadKikanRule]
126 #-----
127 sub loadKikanRule{
128     my $this = shift;
129     my ($in) = @_ ;
130     my $info = {};
131     $info->{NUM} = 0;
132     $info->{PATH} = $in;
133     say STDERR "[INFO]: 期間正規化ルール (" . $info->{PATH} . ") loading...";
134     my $csv = Text::CSV::Encoded->new({
135         encoding_in => 'UTF-8',
136         encoding_out => 'UTF-8'
137     });
138     open my $IN, '<:utf8', $in or die "cannot open $in $!";
139     while (my $fields = $csv->getline($IN)){
140         # 現在有効なルールのみ
141         if ($fields->[5] eq '1'){
142             $info->{NUM} ++;
143             my $p = $fields->[1];
144             $this->{KIKANRULE}{$p}{FREQ} = $fields->[0];
145             $this->{KIKANRULE}{$p}{VAR} = $fields->[3];

```

```

146     $this->{KIKANRULE}{$p}{UNIT} = $fields->[4];
147     }
148 }
149 close($IN);
150
151 say STDERR "[INFO]: 有効期間正規化ルール数 " . $info->{NUM};
152 }
153
154 #-----
155 # [getKikanPattern]
156 #-----
157 sub getKikanPattern{
158     my $this = shift;
159     my ($in) = @_ ;
160     my $u = $this->{UTIL};
161     my $in2 = $u->rmSpaceSimple($in);
162     my $in3 = $u->zen2han($in2);
163     $in3 =~ tr/0-9/xxxxxxxxxxx/;
164     return $in3;
165 }
166
167
168 1;
169
170

```

```

1 package GKAI::Util;
2
3 #=====
4 #==                                     ==#
5 #==          [ GKAI::Util ]          ==#
6 #==                                     ==#
7 #==                                     ==#
8 #=====
9
10 use warnings;
11 use v5.28;
12 use utf8;
13 use Encode;
14 use Carp qw(croak);
15 use Text::CSV::Encoded;
16
17 binmode STDIN, ":utf8";
18 binmode STDOUT, ":utf8";
19 binmode STDERR, ":utf8";
20
21 #== 初期設定 =====
22 our $VERSION = 1.00;
23 our %DEFAULT = ();
24
25 #== コンストラクタ =====
26 sub new {
27     my $class = shift;
28     my $this = {};
29     bless $this, $class;
30
31     #== 変数の設定 =====
32     my (%option) = @_ ;
33     #== 初期設定 =====
34
35     return $this;
36 }
37
38 #== デストラクタ =====
39 sub DESTROY {
40     my $this = shift;
41     #untie( %{ $this->{DATA} } );
42 }
43
44 #== Methods =====
45
46 #-----
47 # [ymd] 年月日のフォーマット合わせ
48 #     3020421 4280502
49 #-----
50 sub ymd{
51     my $this = shift;
52     my $in = shift or croak("[ERROR]: No Input");
53     my $y = '';
54     my $t1 = substr($in, 0, 1);
55     my $t2 = substr($in, 1, 2);
56     my $m = substr($in, 3, 2);
57     my $d = substr($in, 5, 2);
58     my $ret = '';
59
60     # 明治
61     if ($t1 eq '1'){
62         $y = 1867 + $t2;
63     }
64     # 大正
65     elsif ($t1 eq '2'){
66         $y = 1911 + $t2;
67     }
68     # 昭和
69     elsif ($t1 eq '3'){
70         $y = 1925 + $t2;
71     }
72     # 平成
73     elsif ($t1 eq '4'){

```

```

74 $y = 1988 + $t2;
75 }
76 $ret = $y . '/' . $m . '/' . $d;
77 if ($in =~ /VV/){
78 $ret = '1940/01/01';
79 }
80 return $ret;
81 }
82
83 #-----
84 # [keta] certificate key 用桁合わせ
85 #-----
86 sub keta{
87     my $this = shift;
88     my ($in) = @_ ;
89     my $ret = '';
90     $ret = "0" x (7 - length($in)) . $in;
91     return $ret;
92 }
93
94 #-----
95 # [zen2han] (英数記号全て) 全角 => 半角変換
96 #-----
97 sub zen2han {
98     my $this = shift;
99     my ($str) = @_ ;
100     $str =~ tr/0 1 2 3 4 5 6 7 8 9/0-9/;
101     $str =~ tr/ABCDEFGHIJKLMNOPQRSTUVWXYZ/A-Z/;
102     $str =~ tr/abcdefghijklmnopqrstuvwxyz/a-z/;
103     $str
104     =~ tr/\. \, \' \" \~ \ \- \- \- \ / \ ! \# \% \& \ ( \) \* \+ \: \; \< \= \> \? \
[\ ] \ { } \ ^ \_ \ ` \ | \ \! \# \% \& \ ( \) \* \+ \: \; \< \= \> \? \ [ \ ] \ { \ } /;
105     return $str;
106 }
107
108 #-----
109 # [rmSpaceSimple] <安全> (1) 先頭と文末 (2) 読点まわり だけ空白を削除
110 #-----
111 sub rmSpaceSimple {
112     my $this = shift;
113     my ($str) = @_ ;
114
115     #== 先頭と文末のいらぬ空白を削除 ==
116     $str =~ s/^\x{3000}+//;
117     $str =~ s/\x{3000}+$//;
118     $str =~ s/^\s+//;
119     $str =~ s/\s+$//;
120
121     #== 読点まわり ==
122     while ( $str =~ /(、)(?: +)/g ) {
123         $str = "$`" . $1 . "$'";
124     }
125     while ( $str =~ /(?: +)(、)/g ) {
126         $str = "$`" . $1 . "$'";
127     }
128     return $str;
129 }
130
131 #-----
132 # [medcodHead] MedCod テーブルの項目リストを引く
133 #-----
134 sub medcodHead{
135     my $this = shift;
136     my @medcod = ('CertificateKey',
137                 'LineNb',
138                 'TextLine',
139                 'CodeLine',
140                 'IntervalLine',
141                 'CodeOnly',
142                 'LineCoded');
143     return \@medcod;
144 }
145

```

```

146 #-----
147 # [identHead] Ident テーブルの項目リストを引く
148 #-----
149 sub identHead{
150     my $this = shift;
151     my @idents = ('CertificateKey',
152                 'LastChange',
153                 'DateBirth',
154                 'DateDeath',
155                 'Age',
156                 'Sex',
157                 'MannerOfDeath',
158                 'UCCode',
159                 'MainInjury',
160                 'Status',
161                 'Reject',
162                 'Coding',
163                 'CodingVersion',
164                 'CodingFlags',
165                 'SelectedCodes',
166                 'SubstitutedCodes',
167                 'ErnCodes',
168                 'AcmeCodes',
169                 'MultipleCodes',
170                 'Comments',
171                 'FreeText',
172                 'ToDoList',
173                 'CoderReject',
174                 'DiagnosisModified',
175                 'Residence',
176                 'Name',
177                 'Address',
178                 'AutopsyRequested',
179                 'AutopsyUsed',
180                 'RecentSurgery',
181                 'DateOfSurgery',
182                 'ReasonSurgery',
183                 'DateOfInjury',
184                 'PlaceOfOccurrence',
185                 'ActivityCode',
186                 'ExternalFreeText',
187                 'Pregnancy',
188                 'PregnancyContributeDeath',
189                 'Stillbirth',
190                 'MultiplePregnancy',
191                 'CompletedWeeks',
192                 'BirthWeight',
193                 'AgeOfMother',
194                 'ConditionsMother',
195                 'CertImage');
196     return \@idents;
197 }
198
199 1;
200
201

```

Iris入力データ作成用 日本独自 ICDコード変換ルール

対象	変換後	数	対象症病名例	Valid Flag (2は確信度が低いもの)	対応方法・備考
J9699	J969	5338	呼吸不全	1	桁落とし
J9609	J960	4009	急性呼吸不全	1	桁落とし
E14	E149	1767	糖尿病	1	9合併症を伴わない
J9619	J961	1141	慢性呼吸不全	1	桁落とし
G903	G239	237	多系統萎縮症	2	シャイ・ドレーガー症候群 => G23.8, 多系統萎縮症 => G23.2 or G23.3
E11	E119	223	2型糖尿病	2	9合併症を伴わない
B59	J173	105	ニューモシスチス肺炎	1	併記されてる別コード
I7020	I702	31	末梢動脈硬化症	1	桁落とし
J9691	J969	30	2型呼吸不全	1	桁落とし
E10	E109	17	1型糖尿病	2	9合併症を伴わない
M7265	M726	10	外陰部壊死性筋膜炎	1	桁落とし
R17	R198	10	黄疸	1	消化器系および腹部に関するその他の明示された症状および兆候
J9611	J961	8	慢性2型呼吸不全	1	桁落とし
E13	E139	6	B型インスリン受容体異常症	2	9合併症を伴わない
J9601	J961	1	急性2型呼吸不全	1	桁落とし

Iris入力データ作成用期間表現正規化ルール

頻度	期間表現	Iris 表記	変数	単位	Valid Flag	memo
1253817	不詳			N/A	1	
533962	x日	xDays	\$1	Days	1	
450503	約xヶ月	xMonths	\$1	Months	1	
390626	約x年	xYears	\$1	Years	1	
368978	xx日	xxDays	\$1	Days	1	
364311	xヶ月	xMonths	\$1	Months	1	
282086	x年	xYears	\$1	Years	1	
271392	短時間		2	Hours	1	解釈
209669	不明			N/A	1	
207877	x日間	xDays	\$1	Days	1	
206337	約x週間	xWeeks	\$1	Weeks	1	
199044	約x日	xDays	\$1	Days	1	
188073	x年xヶ月	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	
162854	約x時間	xHours	\$1	Hours	1	
161570	約xx年	xxYears	\$1	Years	1	
143965	xx年	xxYears	\$1	Years	1	
137907	xx日間	xxDays	\$1	Days	1	
133864	x時間	xHours	\$1	Hours	1	
115740	約x日間	xDays	\$1	Days	1	
115673	約xx日	xxDays	\$1	Days	1	
101846	xか月	xMonths	\$1	Months	1	
100864	約x年xヶ月	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	
99877	約x年間	xYears	\$1	Years	1	
98383	x週間	xWeeks	\$1	Weeks	1	
84620	約xか月	xMonths	\$1	Months	1	
82609	約xヶ月	xMonths	\$1	Months	1	
75554	約xx日間	xxDays	\$1	Days	1	
70673	約xカ月	xMonths	\$1	Months	1	
67646	約xヵ月	xMonths	\$1	Months	1	
64885	xヶ月	xMonths	\$1	Months	1	
61412	xカ月	xMonths	\$1	Months	1	
60175	xヵ月	xMonths	\$1	Months	1	
55943	x年xか月	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	
53675	約xヶ月間	xMonths	\$1	Months	1	
50109	x年間	xYears	\$1	Years	1	
45212	約xx時間	xxHours	\$1	Hours	1	
44825	xx時間	xxHours	\$1	Hours	1	
38774	約xx年間	xxYears	\$1	Years	1	
35001	数時間		2	Hours	1	解釈
31651	xx分	xxMinutes	\$1	Minutes	1	
30922	x時間xx分	$x*60+xxMinutes$	$\$1 * 60 + \2	Minutes	1	
30801	x年xカ月	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	
30338	x年xヶ月	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	
30091	x年xヵ月	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	
28310	xxヶ月	xxMonths	\$1	Months	1	
28113	約xx分	xxMinutes	\$1	Minutes	1	
27672	約xxヶ月	xxMonths	\$1	Months	1	
25236	xx年間	xxYears	\$1	Years	1	
23092	x年xxヶ月	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	

22188	約x年半	x*12+6Months	\$1 * 12 + 6	Months	1	
20461	約x年xか月	x*12+xMonths	\$1 * 12 + \$2	Months	1	
17932	xヶ月間	xMonths	\$1	Months	1	
16965	約x年xヶ月	x*12+xMonths	\$1 * 12 + \$2	Months	1	
16438	数分		2	Minutes	1	解釈
16376	同上	前の期間と同じ		N/A	q	他セル(10,12,14,16:l*_p)情報利用
16360	数日		2	Days	1	解釈
16010	xヶ月xx日	31*x+xxDays	\$1 * 31 + \$2	Days	1	
15201	約x年xカ月	x*12+xMonths	\$1 * 12 + \$2	Months	1	
14242	数年		2	Years	1	解釈
14167	約x年xヵ月	x*12+xMonths	\$1 * 12 + \$2	Months	1	
12905	数日間		2	Days	1	解釈
12632	約x時間xx分	x*60+xxMinutes	\$1 * 60 + \$2	Minutes	1	
12573	約xヶ月半	31*x+15Days	\$1 * 31 + 15	Days	1	
11409	約半年	6Months	6	Months	1	
10246	約x年xxヶ月	x*12+xxMonths	\$1 * 12 + \$2	Months	1	
9957	約xヶ月間	xMonths	\$1	Months	1	
9909	数年間		2	Years	1	解釈
9737	短時間*推定+		2	Hours	1	解釈
9300	約x週	xWeeks	\$1	Weeks	1	
8528	約xヵ月間	xMonths	\$1	Months	1	
8436	x月		\$1	Months	1	解釈
8363	急死		1	Days	1	解釈
8343	xx年xヶ月	xx*12+xMonths	\$1 * 12 + \$2	Months	1	
8214	平成xx年x月x	逆算する			q	他セル (49:death_ymd) 利用
8214	xxか月	xxMonths	\$1	Months	1	
7598	約xヵ月間	xMonths	\$1	Months	1	
7543	数ヶ月		3	Months	1	解釈
7406	x年半	x*12+6Months	\$1 * 12 + 6	Months	1	
7375	約xヵ月間	xMonths	\$1	Months	1	
7305	x年xxか月	x*12+xxMonths	\$1 * 12 + \$2	Months	1	
7106	約x年xヶ月間	x*12+xMonths	\$1 * 12 + \$2	Months	1	
6335	xヶ月半	x*31+15Days	\$1 * 31 + 15	Days	1	
6256	数分間		2	Minutes	1	解釈
6213	即死		2	Minutes	1	解釈
6011	xxx日	xxxDays	\$1	Days	1	
5999	約x時間半	x*60+30Minutes	\$1 * 60 + 30	Minutes	1	
5740	xヶ月x日	31*x+xDays	\$1 * 31 + \$2	Days	1	
5721	長期間				q	不明と同義→短期間と比較
5563	約xxか月	xxMonths	\$1	Months	1	
5373	約x月	xMonths	\$1	Months	1	
5010	x週	xWeeks	\$1	Weeks	1	
4992	約xxヶ月	xxMonths	\$1	Months	1	
4830	xxヶ月	xxMonths	\$1	Months	1	
4683	xxカ月	xxMonths	\$1	Months	1	
4533	xか月xx日	x*31+xxDays	\$1 * 31 + \$2	Days	1	
4457	約半日	12Hours	12	Hours	1	
4425	約xxヵ月	xxMonths	\$1	Months	1	
4160	x年x月		\$1 * 12 + \$2	Months	1	解釈
4150	平成xx年xx月	逆算する			q	他セル (49:death_ymd) 利用
4069	xxヵ月	xxMonths	\$1	Months	1	
4041	xx時間xx分	xx*60+xxMinutes	\$1 * 60 + \$2	Minutes	1	

4037	x年xxカ月	x*12+xxMonths	\$1 * 12 + \$2	Months	1	
3971	数週間		3	Weeks	1	解釈
3875	約xxカ月	xxMonths	\$1	Months	1	
3837	記入なし			N/A	1	
3697	約x.xヶ月	x.x*31Days	\$1	Months	1	解釈
3677	x年xxヶ月	x*12+xxMonths	\$1 * 12 + \$2	Months	1	
3676	x年xxカ月	x*12+xxMonths	\$1 * 12 + \$2	Months	1	
3671	約xx分間	xxMinutes	\$1	Minutes	1	
3501	x年以上	xYears	\$1	Years	1	
3475	x分	xMinutes	\$1	Minutes	1	
3334	x年xヶ月間	x*12+xMonths	\$1 * 12 + \$2	Months	1	
3236	約xカ月半	x*31+15Days	\$1 * 31 + 15	Days	1	
3181	不詳-不詳			N/A	1	
3055	xx年以上	xxYears	\$1	Years	1	
3028	xカ月xx日	x*12+xxDays	\$1 * 31 + \$2	Days	1	
2987	xカ月xx日	x*12+xxDays	\$1 * 31 + \$2	Days	1	
2972	xヶ月間	xMonths	\$1	Months	1	
2912	約xx年xヶ月	xx*12+xMonths	\$1 * 12 + \$2	Months	1	
2878	直後			N/A	1	
2865	詳細不明			N/A	1	
2826	xか月間	xMonths	\$1	Months	1	
2815	x.xヶ月	x.x*31Days	\$1	Months	1	解釈
2772	xカ月間	xMonths	\$1	Months	1	
2741	約xxヶ月間	xxMonths	\$1	Months	1	
2719	xx年xカ月	xx*12+xMonths	\$1 * 12 + \$2	Months	1	
2659	平成xx年x月	逆算する			q	他セル (49:death_ymd) 利用
2615	空欄			N/A	1	
2590	xヶ月xx日	x*31+xxDays	\$1 * 31 + \$2	Days	1	
2576	約半年間	6Months	6	Months	1	
2503	約xヶ月半	x*31+15Days	\$1 * 31 + 15	Days	1	
2434	xx分間	xxMinutes	\$1	Minutes	1	
2433	約x分	xMinutes	\$1	Minutes	1	
2430	xカ月間	xMonths	\$1	Months	1	
2409	まもなく			N/A	1	
2407	x時間半	x*60+30Minutes	\$1 * 60 + 30	Minutes	1	
2407	約x年前	xYears	\$1	Years	1	
2361	短期間		1	Months	1	解釈
2356	約xか月半	x*31+15Days	\$1 * 31 + 15	Days	1	
2327	約xカ月半	x*31+15Days	\$1 * 31 + 15	Days	1	
2325	数十分		30	Minutes	1	解釈
2199	約x.x年	x.x*12Months	\$1	Years	1	解釈
2196	半年	6Months	6	Months	1	
2133	xカ月半	x*31+15Days	\$1 * 31 + 15	Days	1	
2118	約x年xxカ月	x*12+xxMonths	\$1 * 12 + \$2	Months	1	
2081	約xx年前	xxYears	\$1	Years	1	
1975	数時間*推定+		3	Hours	1	解釈
1929	約数時間		3	Hours	1	解釈
1903	x年前	xYears	\$1	Years	1	
1832	約x年-約x年	xYears	\$2	Years	1	解釈
1823	不祥			N/A	1	
1810	約xヶ月xx日	x*31+xxDays	\$1 * 31 + \$2	Days	1	
1772	約x年xxヶ月	x*12+xxMonths	\$1 * 12 + \$2	Months	1	

1760	約x年xxカ月	x*12+xxMonths	\$1 * 12 + \$2	Months	1	
1679	約x.x時間	x.x*60Minutes	\$1	Hours	1	解釈
1636	x時間x分	x*60+xMinutes	\$1 * 60 + \$2	Minutes	1	
1616	約x年x月	x*12+xMonths	\$1 * 12 + \$2	Months	1	
1584	xxxx年x月x	逆算する			q	他セル (49:death_ymd) 利用
1581	xか月x日		\$1 * 31 + \$2	Days	1	
1563	数十年		20	Years	1	解釈
1562	まもなく*推定+					
1561	xか月半	x*31+15Days	\$1 * 31 + 15	Days	1	
1559	x年xヶ月xx日	x*365+x*31	\$1 * 12 + \$2	Months	1	解釈
1535	xx年前	xxYears	\$1	Years	1	
1522	即時					
1464	xxx日間	xxxDays	\$1	Days	1	
1454	約x日半	x*24+12Hours	\$1 * 24 + 12	Hours	1	
1433	xx年xカ月	xx*12+xMonths	\$1 * 12 + \$2	Months	1	
1425	約x年xxヵ月	xx*12+xMonths	\$1 * 12 + \$2	Months	1	
1395	xカ月半	x*31+15Days	\$1 * 31 + 15	Days	1	
1393	半日	12Hours	12	Hours	1	
1392	xx年xヵ月	xx*12+xMonths	\$1 * 12 + \$2	Months	1	
1388	約x年xヶ月間	x*12+xMonths	\$1 * 12 + \$2	Months	1	
1358	xx年xヶ月	xx*12+xMonths	\$1 * 12 + \$2	Months	1	
1341	やや短時間					
1327	数か月	6Months	3	Months	1	解釈
1309	xヶ月半	x*31+15Days	\$1 * 31 + 15	Days	1	
1308	約xxx日	xxxDays	\$1	Days	1	
1304	Hxx.x.xx	逆算する			q	他セル (49:death_ymd) 利用
1272	平成xx年	逆算する			q	他セル (49:death_ymd) 利用
1213	xヶ月xx日間	x*31+xxDays	\$1 * 31 + \$2	Days	1	
1208	数カ月		3	Months	1	解釈
1208	約数分		3	Minutes	1	解釈
1187	記載なし				q	不明のため
1187	短時間推定					
1167	約xx年-約x年	xYears	\$1	Years	1	解釈
1164	瞬時					
1161	約x時間*推定+	xHours	\$1	Hours	1	解釈
1161	x日前	xDays	\$1	Days	1	
1156	xヵ月x日	x*31+xDays	\$1 * 31 + \$2	Days	1	
1146	xx年xxヶ月	xx*12+xxMonths	\$1 * 12 + \$2	Months	1	
1144	xxヶ月間	xxMonths	\$1	Months	1	
1139	数ヶ月間	6Months	3	Months	1	
1131	ごく短時間					
1129	平成xx年x月xx日	逆算する			q	他セル (49:death_ymd) 利用
1122	x日xx時間	x*24+xxHours	\$1 * 24 + \$2	Hours	1	
1112	x.x年	x.x*12Months	\$1	Years	1	解釈
1070	xカ月x日	x*31+xDays	\$1 * 31 + \$2	Days	1	
1065	x時間程度	xHours	3	Hours	1	
1062	x年-x年	xYears	\$2	Years	1	解釈
1007	x月xx日					
1001	約一週間	1Weeks	1	Weeks	1	
1001	x年xx日		\$1 * 365 + \$2	Days	1	
989	約xx年-約xx年	xxYears	\$2	Years	1	解釈
974	x.x時間	x.x*60Minutes	\$1	Hours	1	解釈

970	約x年xヵ月間	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	
965	xヶ月x日	$x*31+xDays$	$\$1 * 31 + \2	Days	1	
953	xx月	xxMonths	\$1	Months	1	
942	数ヶ月		3	Months	1	
941	長期					
939	約x年xヵ月間	$x*12+xMonths$	$\$1 * 12 + \2	Months	1	
937	約xx月		\$1	Months	1	
922	x～x時間		\$1	Hours	1	
921	約x年xヵ月間		$\$1 * 12 + \2	Months	1	

機械学習による原死因コード変更有無予測結果（一部抜粋）

仮ID	予測結果 1:変更あり 0:変更なし	正解 1:変更あり 0:変更なし	仮原死因 (Iris) コード	Iris Reject	確定原死因 コード	外因符号	正誤
90	0	0	E46	No	E46		TRUE
560	0	0	I639	No	I639		TRUE
750	1	0	C349	No	C349		FALSE
1530	0	0	C679	No	C679		TRUE
2230	0	0	R54	No	R54		TRUE
2530	0	0	I619	No	I619		TRUE
2660	0	0	C169	No	C169		TRUE
3800	0	0	C23	No	C23		TRUE
4350	0	0	T58	No	T58	X474	TRUE
4760	0	0	I219	No	I219		TRUE
4830	0	0	D376	No	D376		TRUE
5370	0	0	R579	No	R579		TRUE
5730	0	0	J189	No	J189		TRUE
5810	0	0	R54	No	R54		TRUE
6070	0	0	R579	No	R579		TRUE
6110	0	0	C509	No	C509		TRUE
6180	0	0	C349	No	C349		TRUE
6190	0	0	R688	No	R688		TRUE
6870	0	0	C80	No	C80		TRUE
7620	0	0	C080	No	C080		TRUE
7840	1	1	I743	No	I709		TRUE
8280	0	0	F03	No	F03		TRUE
8630	0	0	C250	No	C250		TRUE
9410	0	0	C187	No	C187		TRUE
9890	0	0	C187	No	C187		TRUE
10300	0	1	I619	No	I611		FALSE
10380	1	0	S269	MayBe	S269	X788	FALSE
10490	0	0	I711	No	I711		TRUE
10650	1	1	C349	No	C343		TRUE
10700	0	0	F03	No	F03		TRUE
10730	0	0	I693	No	I693		TRUE
11030	0	0	C250	No	C250		TRUE
11320	0	0	T58	No	T58	X678	TRUE
11470	0	0	J189	No	J189		TRUE
12080	0	0	R54	No	R54		TRUE

12270	0	0	C169	No	C169		TRUE
12280	0	0	C55	No	C55		TRUE
12860	1	1	R090	No	T179	W790	TRUE
13320	0	0	C220	No	C220		TRUE
13520	0	0	I693	No	I693		TRUE
13550	0	0	N179	No	N179		TRUE
14830	0	0	I710	No	I710		TRUE
15010	0	0	R54	No	R54		TRUE
15510	0	0	J841	No	J841		TRUE
15570	0	0	R54	No	R54		TRUE
15590	0	0	C211	No	C211		TRUE
15730	0	0	C20	No	C20		TRUE
16030	0	0	C20	No	C20		TRUE
16200	1	0	C189	No	C189		FALSE
16900	0	0	C184	No	C184		TRUE
17490	0	0	C679	No	C679		TRUE
17540	0	0	T751	No	T751	W650	TRUE
17680	1	1	M069	No	I509		TRUE
18000	0	0	T751	No	T751	W650	TRUE
18160	1	1	I639	No	I693		TRUE
18970	0	0	R54	No	R54		TRUE
19340	0	0	C187	No	C187		TRUE
20390	0	0	C900	No	C900		TRUE
20550	0	0	C221	No	C221		TRUE
20710	0	0	C56	No	C56		TRUE
22340	0	0	J439	No	J439		TRUE
22700	0	0	M471	No	M471		TRUE
22740	0	0	I249	No	I249		TRUE
23290	0	0	I509	No	I509		TRUE
24520	0	0	C250	No	C250		TRUE
24670	0	0	C20	No	C20		TRUE
25100	0	0	C182	No	C182		TRUE
25150	0	0	I219	No	I219		TRUE
25340	0	1	C169	No	C162		FALSE
25580	0	0	I609	No	I609		TRUE
25630	0	0	I609	No	I609		TRUE
26710	0	1	C349	No	C343		FALSE
26750	0	0	C509	No	C509		TRUE
26770	0	0	J189	No	J189		TRUE
27100	0	0	I639	No	I639		TRUE
27180	0	0	Q042	No	Q042		TRUE

27280	0	0	I609	No	I609		TRUE
27520	0	0	I219	No	I219		TRUE
27590	0	0	N179	No	N179		TRUE
27760	0	0	C250	No	C250		TRUE
28510	0	0	C711	No	C711		TRUE
28570	0	0	I610	No	I610		TRUE
28660	0	0	C549	No	C549		TRUE
28780	0	1	I219	No	I251		FALSE
28910	0	0	I693	No	I693		TRUE
29470	0	0	C509	No	C509		TRUE
29540	0	1	T794	No	S409	W20	FALSE
29600	0	0	I609	No	I609		TRUE
29750	0	0	I619	No	I619		TRUE
29830	0	0	F03	No	F03		TRUE
29880	0	0	N189	No	N189		TRUE
30160	0	0	S065	No	S065	W010	TRUE
30670	0	0	C349	No	C349		TRUE
30710	0	0	C445	No	C445		TRUE
30780	0	0	I422	No	I422		TRUE
30810	0	0	C20	No	C20		TRUE
31480	0	0	K922	No	K922		TRUE
31710	0	0	J690	No	J690		TRUE
31800	0	0	R54	No	R54		TRUE
32520	1	0	C349	No	C349		FALSE
33560	0	0	S328	No	S328	W13	TRUE
33860	0	0	C159	No	C159		TRUE
34240	0	0	I509	No	I509		TRUE
34250	0	0	C169	No	C169		TRUE
34620	0	0	R54	No	R54		TRUE
35020	0	0	C539	No	C539		TRUE
36170	0	0	C184	No	C184		TRUE
36490	0	0	C159	No	C159		TRUE
36560	0	0	C159	No	C159		TRUE
36660	0	0	C349	No	C349		TRUE
36820	0	0	C162	No	C162		TRUE
36900	0	0	I211	No	I211		TRUE
37730	0	0	C221	No	C221		TRUE
37810	0	0	K746	No	K746		TRUE
38310	0	0	J690	No	J690		TRUE
38630	0	0	C20	No	C20		TRUE
38910	0	0	N19	No	N19		TRUE

38950	0	0	C220	No	C220		TRUE
39030	0	0	R54	No	R54		TRUE
39440	0	0	I219	No	I219		TRUE
39490	1	1	C248	No	C240		TRUE
40140	1	1	M625	No	R54		TRUE
41370	1	1	C248	No	C240		TRUE
41450	0	0	C184	No	C184		TRUE
41570	0	0	J189	No	J189		TRUE
41780	0	0	K802	No	K802		TRUE
41890	0	0	J690	No	J690		TRUE
41980	0	0	C240	No	C240		TRUE
42160	0	0	C20	No	C20		TRUE
42360	0	0	J690	No	J690		TRUE
42490	0	0	I693	No	I693		TRUE
42630	0	0	R54	No	R54		TRUE
43320	0	0	N390	No	N390		TRUE
43660	0	0	G301	No	G301		TRUE
44090	1	1	R090	No	T598	X670	TRUE
44120	0	0	J189	No	J189		TRUE
44380	0	0	I716	No	I716		TRUE
44450	0	1	R688	No	S828	X599	FALSE
44500	0	0	D381	No	D381		TRUE
44530	0	0	C509	No	C509		TRUE
45750	0	0	I619	No	I619		TRUE
46030	1	0	D65	No	D65		FALSE
46510	0	0	C169	No	C169		TRUE
46590	0	0	C187	No	C187		TRUE
46840	0	0	C65	No	C65		TRUE
46980	0	0	C220	No	C220		TRUE
47100	0	0	C20	No	C20		TRUE
47720	0	0	C509	No	C509		TRUE
47850	0	0	I219	No	I219		TRUE
48430	0	0	I48	No	I48		TRUE
48470	0	0	C169	No	C169		TRUE
48830	0	0	J449	No	J449		TRUE
48880	0	0	C349	No	C349		TRUE
48960	1	1	R571	No	T794	V031	TRUE
49030	1	1	J189	No	J690		TRUE
49050	0	0	C20	No	C20		TRUE
49080	0	0	I619	No	I619		TRUE
49310	0	0	E146	No	E146		TRUE

50020	0	0	I249	No	I249		TRUE
50620	1	0	E149	No	E149		FALSE
51040	0	0	R54	No	R54		TRUE
51210	0	0	I249	No	I249		TRUE
51260	0	0	J189	No	J189		TRUE
51460	0	0	J690	No	J690		TRUE
51670	1	1	C189	No	C180		TRUE
51870	0	1	J690	MayBe	I209		FALSE
52180	0	0	C20	No	C20		TRUE
52700	0	0	I249	No	I249		TRUE
52890	0	0	I509	No	I509		TRUE
52990	0	0	S062	No	S062	V031	TRUE
53380	0	0	C250	No	C250		TRUE
53460	0	0	T751	No	T751	V922	TRUE
53480	0	0	C169	No	C169		TRUE
53520	0	0	C220	No	C220		TRUE
53550	0	0	I509	No	I509		TRUE
53830	0	0	I259	No	I259		TRUE
53930	0	0	C187	No	C187		TRUE
55400	0	0	I714	No	I714		TRUE
55460	1	1	C349	No	C343		TRUE
55600	0	0	C169	No	C169		TRUE
55860	0	0	I609	No	I609		TRUE
55960	0	0	J189	No	J189		TRUE
56100	0	0	N19	No	N19		TRUE
57110	0	1	C259	No	C250		FALSE
57580	0	0	C349	No	C349		TRUE
57860	0	0	I712	No	I712		TRUE
58190	0	0	I219	No	I219		TRUE
58260	0	0	I219	No	I219		TRUE
58310	0	0	C169	No	C169		TRUE
58990	0	0	K562	No	K562		TRUE
59020	0	0	C187	No	C187		TRUE
59070	0	0	T751	No	T751	W65	TRUE
59260	1	0	C189	No	C189		FALSE
59530	0	0	D469	No	D469		TRUE
59810	0	0	I219	No	I219		TRUE
60090	0	0	C252	No	C252		TRUE
60110	0	0	R54	No	R54		TRUE
60190	0	0	I110	No	I110		TRUE
60250	0	1	C259	No	C220		FALSE

61570	0	0	J849	No	J849		TRUE
61640	0	0	C169	No	C169		TRUE
61990	0	0	J439	No	J439		TRUE
62250	0	0	C184	No	C184		TRUE
62290	0	0	K703	No	K703		TRUE
62520	1	1	C349	No	C348		TRUE
62640	0	0	C56	No	C56		TRUE
62890	0	0	K650	No	K650		TRUE
63140	0	0	C187	No	C187		TRUE
63700	0	0	C184	No	C184		TRUE
63850	0	0	C349	No	C349		TRUE
64300	0	0	J439	No	J439		TRUE
64930	0	0	C20	No	C20		TRUE
65080	1	1	C349	No	C343		TRUE
65300	0	0	C679	No	C679		TRUE
65400	0	0	K261	No	K261		TRUE
65690	0	0	I619	No	I619		TRUE
66060	0	0	C180	No	C180		TRUE
66230	1	0	C349	No	C349		FALSE
66670	0	0	C844	No	C844		TRUE
67190	0	0	C915	No	C915		TRUE
67220	0	0	S069	No	S069	W100	TRUE
67230	1	0	C349	No	C349		FALSE
67850	0	0	C169	No	C169		TRUE
67860	0	0	C341	No	C341		TRUE
68090	0	0	R54	No	R54		TRUE
68220	0	0	C509	No	C509		TRUE
68610	0	0	C220	No	C220		TRUE
68900	0	0	I619	No	I619		TRUE
69840	0	0	J459	No	J459		TRUE
70200	0	0	C509	No	C509		TRUE
70280	0	0	C349	No	C349		TRUE
71010	0	0	T07	No	T07	V234	TRUE
71080	0	0	C220	No	C220		TRUE
71110	0	1	M100	No	M103		FALSE
71790	1	1	C542	No	C55		TRUE
71880	0	0	C259	No	C259		TRUE
73990	1	1	J849	No	C341		TRUE
74040	0	0	S062	No	S062	W130	TRUE
74210	1	1	I620	No	S065	X599	TRUE
74490	0	0	R54	No	R54		TRUE

74630	0	0	I258	No	I258		TRUE
74830	0	0	I619	No	I619		TRUE
75630	0	0	C73	No	C73		TRUE
75670	0	0	J849	No	J849		TRUE
75790	0	0	C549	No	C549		TRUE
75810	0	0	C20	No	C20		TRUE
76080	1	1	I48	MayBe	N189		TRUE
76600	0	0	C060	No	C060		TRUE
77040	0	0	J849	No	J849		TRUE
77050	0	0	I609	No	I609		TRUE
77170	0	0	C859	No	C859		TRUE
77600	0	1		No	R99		FALSE
77660	0	0	C349	No	C349		TRUE
77760	0	0	F03	No	F03		TRUE
78020	0	0	C679	No	C679		TRUE
78230	0	0	J42	No	J42		TRUE
78380	0	0	R54	No	R54		TRUE
78620	0	0	M314	No	M314		TRUE
78850	0	0	I219	No	I219		TRUE
79140	0	0	C679	No	C679		TRUE
79160	0	0	C182	No	C182		TRUE
79240	1	1	I692	No	T179	W780	TRUE
79380	0	0	C20	No	C20		TRUE
79510	0	0	J189	No	J189		TRUE
79960	0	0	C169	No	C169		TRUE
80070	0	0	J189	No	J189		TRUE
80290	0	0	I713	No	I713		TRUE
81070	0	0	C169	No	C169		TRUE
81200	0	0	S062	No	S062	V031	TRUE
81570	0	1		No	R99		FALSE
82010	0	0	I249	No	I249		TRUE
82520	0	0	I710	No	I710		TRUE
82690	0	0	C20	No	C20		TRUE
83000	0	0	I340	No	I340		TRUE
83200	0	0	C159	No	C159		TRUE
83340	0	0	R54	No	R54		TRUE
83460	0	0	C169	No	C169		TRUE
83490	0	0	T751	No	T751	X71	TRUE
83620	0	1	S069	No	S064	X808	FALSE
84320	0	0	C240	No	C240		TRUE
84640	0	0	C159	No	C159		TRUE

85080	1	0	C349	No	C349		FALSE
85160	0	0	C64	No	C64		TRUE
85400	0	1	C169	No	C168		FALSE
85880	0	0	C182	No	C182		TRUE
86240	0	0	C187	No	C187		TRUE
86490	0	0	C169	No	C169		TRUE
86510	1	1	I714	No	I509		TRUE
86940	0	0	T751	No	T751	W650	TRUE
87440	0	0	I639	No	I639		TRUE
87470	0	0	I619	No	I619		TRUE
87600	0	0	R54	No	R54		TRUE
88010	0	0	C250	No	C250		TRUE
88090	1	1	C349	No	C341		TRUE
88110	0	0	E142	No	E142		TRUE
88180	0	0	C180	No	C180		TRUE
88190	0	0	I509	No	I509		TRUE
88310	0	0	I619	No	I619		TRUE
88480	1	1	C189	No	C187		TRUE
89010	1	0	C189	No	C189		FALSE
89190	0	0	C20	No	C20		TRUE
89510	0	0	T751	No	T751	W69	TRUE
90020	0	0	C509	No	C509		TRUE
90090	0	0	T751	No	T751	W650	TRUE
90350	0	0	I509	No	I509		TRUE
90670	0	0	I219	No	I219		TRUE
91180	0	0	C20	No	C20		TRUE
91220	1	1	S141	No	S197	X800	TRUE
91950	0	0	C679	No	C679		TRUE
92080	0	0	E669	No	E669		TRUE
92880	0	0	C180	No	C180		TRUE
93070	0	0	D619	No	D619		TRUE
93270	0	0	C509	No	C509		TRUE
93500	0	0	C159	No	C159		TRUE
93600	0	1	C20	No	C349		FALSE
93610	0	0	J42	No	J42		TRUE
93710	0	0	I509	No	I509		TRUE
94030	1	0	C189	No	C189		FALSE
94080	0	1	J841	No	J849		FALSE
94090	0	0	C719	No	C719		TRUE
94910	0	0	T751	No	T751	W650	TRUE
95890	0	0	T751	No	T751	W650	TRUE

95930	0	0	I219	No	I219		TRUE
96600	0	0	J189	No	J189		TRUE
97280	0	0	C169	No	C169		TRUE
97330	1	0	C189	No	C189		FALSE
97720	0	0	M318	No	M318		TRUE
97890	0	0	I219	No	I219		TRUE
98480	0	0	I500	No	I500		TRUE
98640	0	0	C169	No	C169		TRUE
98650	0	1	E149	No	E144		FALSE
98700	0	0	I219	No	I219		TRUE
98950	0	0	B181	No	B181		TRUE
99160	0	0	C169	No	C169		TRUE
99400	0	0	F03	No	F03		TRUE
99850	0	0	C182	No	C182		TRUE
100850	1	0	C349	No	C349		FALSE
101250	0	0	C169	No	C169		TRUE
101320	0	0	I490	No	I490		TRUE
101490	0	0	C20	No	C20		TRUE
101890	0	0	C679	No	C679		TRUE
102440	0	0	I509	No	I509		TRUE
102660	0	0	C56	No	C56		TRUE
102980	0	0	C240	No	C240		TRUE
102990	0	1	D150	No	C37		FALSE
103280	0	0	R54	No	R54		TRUE
103290	0	0	R54	No	R54		TRUE
104050	0	0	R54	No	R54		TRUE
104180	0	0	I219	No	I219		TRUE
104870	0	0	C20	No	C20		TRUE
105230	0	0	C162	No	C162		TRUE
105310	0	0	C240	No	C240		TRUE
105590	0	0	S299	No	S299	X80	TRUE
105790	0	1	C169	No	C160		FALSE
106570	0	0	I691	No	I691		TRUE
106830	0	0	I350	No	I350		TRUE
108180	1	1	K859	No	K851		TRUE
108640	0	0	I509	No	I509		TRUE
108970	0	0	D469	No	D469		TRUE
109160	0	0	I259	No	I259		TRUE
109240	1	1	C189	No	C187		TRUE
109790	0	0	S141	No	S141	V051	TRUE
110530	0	0	I619	No	I619		TRUE

110600	0	0	I609	No	I609		TRUE
110850	0	1	C169	No	C162		FALSE
110860	0	1	I48	No	I630		FALSE
110960	1	1	G309	No	J189		TRUE
111300	0	0	S099	No	S099	X599	TRUE
111820	0	0	I619	No	I619		TRUE
111850	0	0	R54	No	R54		TRUE
111960	0	0	I200	No	I200		TRUE
112090	0	0	I693	No	I693		TRUE
112170	0	0	J189	No	J189		TRUE
112290	0	0	C480	No	C480		TRUE
112460	0	0	R54	No	R54		TRUE
112580	0	0	T07	No	T07	W110	TRUE
112590	0	0	I639	No	I639		TRUE
112820	0	0	N390	No	N390		TRUE
113080	0	0	C221	No	C221		TRUE
114210	0	0	G20	No	G20		TRUE
114330	1	1	C189	No	C184		TRUE
114710	0	0	C220	No	C220		TRUE
114940	1	0	C19	No	C19		FALSE
115060	0	0	I48	No	I48		TRUE
115200	0	0	C240	No	C240		TRUE
115830	0	0	J189	No	J189		TRUE
116100	0	0	C679	No	C679		TRUE
116600	0	0	C169	No	C169		TRUE
117110	0	0	I509	No	I509		TRUE
117500	0	0	C65	No	C65		TRUE
118400	0	0	C220	No	C220		TRUE
118500	0	0	K746	No	K746		TRUE
118690	0	0	A162	No	A162		TRUE
118740	0	0	C23	No	C23		TRUE
118800	0	0	C549	No	C549		TRUE
119610	0	0	D329	No	D329		TRUE
119640	1	1	I259	No	I249		TRUE
119710	0	1	I619	No	I509		FALSE
119990	0	0	S099	No	S099	X808	TRUE
120020	0	0	G301	No	G301		TRUE
120230	0	0	K255	No	K255		TRUE
120280	0	0	J189	No	J189		TRUE
121010	0	0	C20	No	C20		TRUE
121450	0	0	T751	No	T751	W65	TRUE

令和3年5月20日

厚生労働大臣 殿

機関名 国立大学法人東京大学

所属研究機関長 職名 総長

氏名 藤井 輝夫



次の職員の令和2年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

- 1. 研究事業名 政策科学総合研究事業（統計情報総合研究事業）
- 2. 研究課題名 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
- 3. 研究者名 (所属部局・職名) 大学院医学系研究科・准教授
(氏名・フリガナ) 今井 健 ・イマイタケシ

4. 倫理審査の状況

	該当性の有無		左記で該当がある場合のみ記入（※1）		
	有	無	審査済み	審査した機関	未審査（※2）
ヒトゲノム・遺伝子解析研究に関する倫理指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
遺伝子治療等臨床研究に関する指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
人を対象とする医学系研究に関する倫理指針（※3）	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
その他、該当する倫理指針があれば記入すること (指針の名称：)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他（特記事項）

(※2) 未審査の場合は、その理由を記載すること。

(※3) 廃止前の「疫学研究に関する倫理指針」や「臨床研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 <input checked="" type="checkbox"/> 未受講 <input type="checkbox"/>
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合はその理由：)
当研究機関におけるCOI委員会設置の有無	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合は委託先機関：)
当研究に係るCOIについての報告・審査の有無	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合はその理由：)
当研究に係るCOIについての指導・管理の有無	有 <input type="checkbox"/> 無 <input checked="" type="checkbox"/> (有の場合はその内容：)

(留意事項) ・該当する□にチェックを入れること。
・分担研究者の所属する機関の長も作成すること。

令和3年2月9日

厚生労働大臣 殿

機関名 国立大学法人筑波大学

所属研究機関長 職名 学長

氏名 永田 恭介



次の職員の令和2年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

- 研究事業名 令和2年度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））
- 研究課題名 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
- 研究者名（所属部局・職名） 医学医療系 講師
(氏名・フリガナ) 香川 璃奈 ・ カガワ リナ

4. 倫理審査の状況

	該当性の有無		左記で該当がある場合のみ記入（※1）		
	有	無	審査済み	審査した機関	未審査（※2）
ヒトゲノム・遺伝子解析研究に関する倫理指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
遺伝子治療等臨床研究に関する指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
人を対象とする医学系研究に関する倫理指針（※3）	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
その他、該当する倫理指針があれば記入すること (指針の名称：)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>

（※1）当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他（特記事項）

（※2）未審査の場合は、その理由を記載すること。

（※3）廃止前の「疫学研究に関する倫理指針」や「臨床研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 <input checked="" type="checkbox"/> 未受講 <input type="checkbox"/>
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合はその理由:)
当研究機関におけるCOI委員会設置の有無	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合は委託先機関:)
当研究に係るCOIについての報告・審査の有無	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合はその理由:)
当研究に係るCOIについての指導・管理の有無	有 <input type="checkbox"/> 無 <input checked="" type="checkbox"/> (有の場合はその内容:)

（留意事項） ・該当する□にチェックを入れること。

・分担研究者の所属する機関の長も作成すること。