

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慮す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。

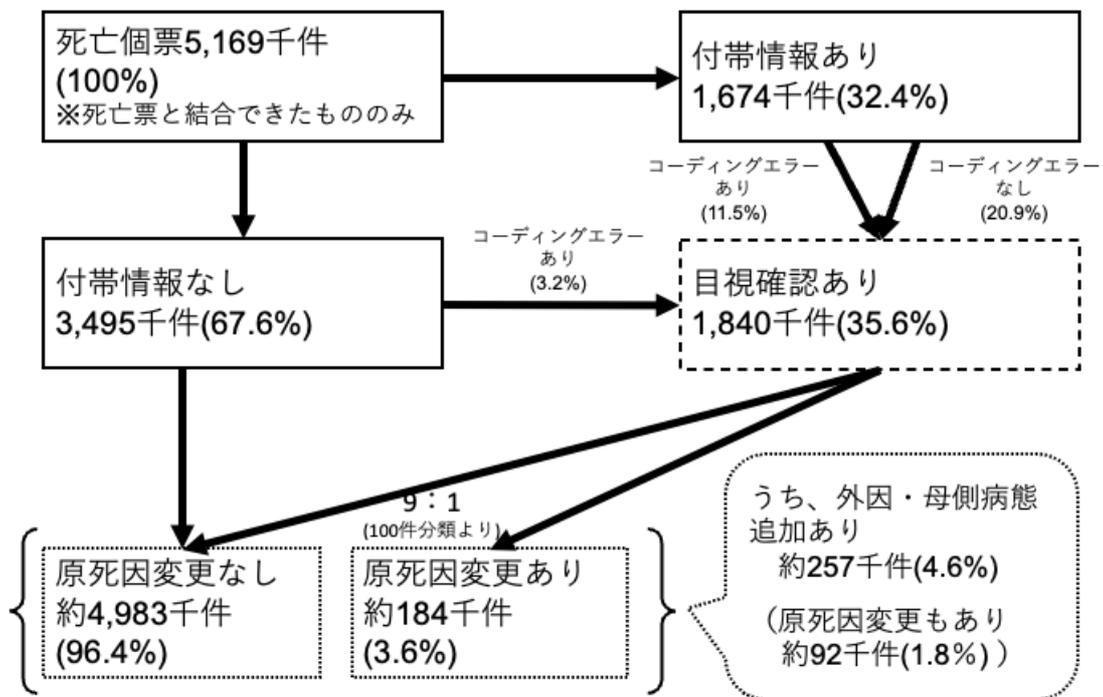


図1 原死因決定プロセス

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 届出	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		inl_ymd	記述あり				傷害が発生したとき	193719			障害発生年月日
29		inl_hm	記述あり				傷害が発生したとき	150902			障害発生時分
30		inl_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		inl_detail	記述あり				傷害が発生したところの他の記述	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		inl_tdfk	記述あり				傷害発生場所	187845			傷害発生都道府県
33		inl_city	記述あり				傷害発生場所	186349			傷害発生市郡
34		inl_town	記述あり				傷害発生場所	78935			傷害発生区町村
35		inl_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		inl_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

