

NDB に対する死亡決定ロジックの手法開発と検証

久保慎一郎^{*1}, 野田龍也^{*1}, 西岡祐一^{*1}, 明神大也^{*1}, 降旗志おり^{*2}, 東野恒之^{*2},
瀬楽丈夫^{*2}, 今村知明^{*1}

^{*1} 奈良県立医科大学 公衆衛生学講座,
^{*2} (株) 三菱総合研究所 ヘルスケア・ウェルネス事業本部

Development and validation of a Logic-based Method for Determining Mortality for the National Database

Kubo Shinichiro ^{*1}, Noda Tatsuya ^{*1}, Nishioka Yuichi ^{*1}, Myojin Tomoya ^{*1},
Furihata Shiori ^{*2}, Higashino Tsuneyuki ^{*2}, Seraku Takeo ^{*2}, Imamura Tomoaki ^{*1}

^{*1} Department of Public Health, Nara Medical University,

^{*2} Healthcare and Wellness Division, Mitsubishi Research Institute, Inc.

日本の保険診療の悉皆データである NDB には、死亡した患者のレセプトの「転帰区分」に死亡が付与されるが、医療機関の付与忘れや付与間違い等で正確とはいえない。診療行為や薬剤等から死亡を推定することで転帰死亡の有効性向上を目的とした。4 年分の奈良県 KDB レセプトと 3 年分の NDB レセプトを用いて、KDB の保険者マスターに記載されている死亡転帰(確実に死亡)を教師データとし、KDB の死亡転帰(不完全な死亡)で必要となる決定木の診療行為を基に決定木分析を行った。その仕組みを用いて NDB でも検証し、死亡数を人口動態統計と比較した。偽陽性を低減するために決定木の条件を調整し、NDB における偽陽性は 0.2%、偽陰性は 5.8%であった。人口動態統計の死亡者総数と死亡決定ロジックによる死亡数の比(2015 年度)は、適用前が 73.3%であったのに対し、適用後が 82.2%となった。死亡アウトカム指標の正確度を高めることで日本のコホート研究が大きく前進する。

キーワード NDB、ナショナルデータベース、レセプト、死亡追跡、人口動態統計

1. はじめに

レセプト情報・特定健診等情報データベース(以下、NDB)とは、電子レセプトデータと特定健診等のデータを格納・構築されているデータベースである。日本の保険診療の悉皆データであり、臨床研究や政策研究の応用が期待されている。

ただし、NDB には様々な問題点や障壁が存在する。そのひとつに、死亡や罹患のようにアウトカムにできる指標が乏しいことが挙げられる。

NDB には転帰区分に「死亡」が付与される。ただし、①入力 of 正確性、②医療機関外の死亡の補足、の課題から正確とは言えない状況である。特に、医療機関の付与忘れや付与間違い等は一定程度発生していると推察される。

本研究は診療行為や薬剤等から死亡推定し、NDB の死亡情報の正確度向上を目的とした。

2. 方法

本研究は医療計画策定に係る評価指標作成の一環として(5 年生存率など死亡関連指標のため)行った。4 年分(2013 年 4 月~)の奈良県の国保および後期高齢者レセプト(以下、KDB)と、3 年分(2013 年 4 月~)の NDB レセプトを用いた。KDB の被保険者マスターに記載されている資格

喪失日は死亡を判定する上で正確なため目的変数とし、診療行為や使用薬剤などのレセプト情報を説明変数として機械分析(rpart)を行った。これで、記載漏れの死亡を推定し正確度向上を目指した。なお、被保険者マスターで最終診療日と被保険者マスターの資格喪失日が2日以上乖離した場合は医療外の死亡とみなして除外した。

決定木分析の説明変数には、退院日または死亡日に算定された診療行為、医薬品、病名、病棟区分を使用した。作成した決定木で死亡推定ロジックを作成し、NDB に適用した。レセプトの死亡転帰に比べてどの程度付与できたかを検証し、人口統計の死亡数と比較し有効性を検証した。

3. 結果

1) 決定木の分類方法について

KDB 上で医療機関が報告した死亡転帰の正確度を検証した。特異度は 99.8%と高い正確度だが、感度が 91.4%と低かった。死亡転帰日以降にレセプトが発生しているケースは、約 2%存在し、間違って死亡とされていた。転帰区分に死亡が付与されている場合は、死亡転帰の誤りを修正するルールを作成することで、死亡の正確度を向上させる必要があることが分かった。

死亡転帰が生存となっている患者においては、機械分析を行った。レセプトにおける正例(生存)と負例(死亡)の比率の関係で、サンプリングの補正処理を行い正確度を高めた。外来の来院で不均衡が顕著であったため、入院、在宅、来院の3つのモデルに分けて機械分類を実施した。これらのモデルの総称を死亡決定ロジックと呼ぶ。

2) KDB での入院・在宅・来院検証結果

Table.1 に各モデルの検証結果を示す。入院モデルを除いてある一定の正確度が担保されたが、入院は偽陽性率が 5.2%と他に比べ高かった。

	転帰死亡ありの場合		転帰死亡なしの場合			⑤全適用 (①～④)
	転帰区分のまま	①転帰死亡の精度向上	②入院	③在宅	④来院	
感度	91.4%	91.3%	94.7%	93.2%	94.4%	94.2%
特異度	99.8%	99.8%	94.8%	99.6%	96.9%	98.8%
偽陽性率	0.2%	0.2%	5.2%	0.4%	3.1%	0.2%
偽陰性率	8.6%	8.7%	5.3%	6.8%	5.6%	5.8%

3) NDB への死亡推定ルールの適用

上記3つのモデルを基として、NDB に死亡を付与した。その際、偽陽性率が高いルールは第一種過誤を避けるため除外した。死亡推定ルールを Table.2 に示す。適用後の患者数と人口動態と比較した患者数の比を Table.3 に示す。

モデル	死亡推定ルール	偽陽性率
入院	最終診療日において、食事療養・生活療養なし、かつ大腸ファイバースコープ(D313)なし、かつ酸素吸入(J024)あり	0.17%
	最終診療日において、食事療養・生活療養なし、かつ大腸ファイバースコープ(D313)なし、かつ酸素吸入(J024)なし、かつ一般病棟以外、かつ人工呼吸(J045)あり	0.20%
	最終診療日において、食事療養・生活療養あり、かつ療養病棟以外、かつ酸素吸入(J024)なし、かつ人工呼吸(J045)あり	0.26%
在宅	最終診療日において、往診料(C000)なし、かつ在宅ターミナル加算なし、かつ看取り加算あり	0.00%
	最終診療日において、往診料(C000)なし、かつ在宅ターミナル加算あり	0.00%
	最終診療日において、往診料(C000)あり、かつ検体検査判断料(D026)なし、かつ処方箋料(F400)あり、かつ死亡診断加算あり	0.00%
	最終診療日において、往診料(C000)あり、かつ検体検査判断料(D026)なし、かつ処方箋料(F400)なし、かつ調剤料(F000)なし、かつ在宅患者訪問診療料(C001)なし、かつ死亡診断加算あり	0.00%
	最終診療日において、非開胸的心マッサージ(J046)あり	0.00%
来院	最終診療日において、非開胸的心マッサージ(J046)なし、かつ呼吸心拍監視(D220)なし、かつ心停止・詳細不明(I469)あり	0.08%

年度	人口動態統計	転帰死亡精度向上		決定木ルール		全ルール適用
		推定死亡数	(比)	推定死亡数	(比)	推定死亡数 (比)
2013年度	1,267,908	914,115	(72.1%)	92,528	(7.3%)	1,006,643 (79.4%)
2014年度	1,272,534	932,053	(73.2%)	95,542	(7.5%)	1,027,595 (80.8%)
2015年度	1,290,444	946,060	(73.3%)	113,595	(8.8%)	1,060,255 (82.2%)

なお死亡者の 99%は 40 歳以上であるため、NDB への適用は 40 歳以上に限定した。

4. 考察

1) 決定木の説明変数の選択の妥当性

説明変数の選択において、レセプトの最終日の診療行為発生に着目した理由は、最終日の処

置が死亡の傾向をつかむうえで重要と判断したためである。検証段階で、集計対象に含める診療行為を期間別(最終日、3 日以内、5 日以内)で検証したが、期間が長くなれば長くなるほど余計なノイズ(日々の治療)が含まれた。最終日に限定したところが正確度を高めた重要な点である。

2) 決定木が臨床において説明できるか

決定木の条件が臨床に合うものでなければ懷疑的と言わざるを得ない。これまでにトライアンドエラーを繰り返しながら決定木作成している。たとえば AUC や感度、特異度が高くても臨床的に説明できないモデルは不採用とした。入院であれば「最終日に食事をしていない人が内視鏡検査を受けたかどうか」という条件が存在する。これは絶食中でも内視鏡検査の場合は検査のみ受けて帰宅する(=予後もいい)ということの表れであるといえる。在宅の場合は、在宅訪問診療料や、死亡診断加算、看取り加算などが説明変数として抽出された。来院に関しては、心臓マッサージや、心停止、心不全、ショックといった病名が説明変数として抽出された。これらの診療行為や薬剤、病名は臨床として違和感なく、今回の決定木の説明変数としては可用性が高いといえる。

3) 死亡数が死亡統計と比べて正確か

人口動態統計と比較すると、転帰区分のみの死亡では 73.3%の正確度であったものを 82.2%まで高めることができた。人口動態統計には医療機関外で死亡したものも含まれているため、医療機関で補足可能な死亡の大部分を補足可能となったと考えられる。

5. 結語

KDB を用いて死亡患者の動向を機械学習し決定木分析を行うことで NDB の死亡フラグを策定した。死亡情報の正確度を高めることが可能となり NDB を用いたコホート研究が前進する。

参考文献

[1]久保 慎一郎, 野田 龍也, 明神 大也、他: NDB(ナショナルデータベース)の課題および留意点と今後の展望, 医療情報学連合大会論文集, 36(1): 272-275, 2016.