

# レセプトデータセットを用いたがん登録の ケースファインディングに有効な統計手法の開発

○小原 仁<sup>1)</sup>, 水谷 駿介<sup>2)</sup>, 小森園 康二<sup>3)</sup>

1) 久慈大学バイオ統計センター, 2) 株式会社メハーゲン医療経営支援課, 3) 東北病院計画内科

## 【目的】

調査実施施設が有するレセプトデータからデータセットを新たに生成し、我々が開発したがん登録のケースファインディングに有効な統計的手法を適用させた判別能を評価する。



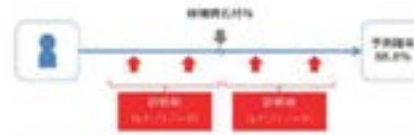
病名と診療行為の両方をもとに  
がん登録症例を識別できないか?



がん登録症例を多くの入院患者や外来患者のなかから  
効率的に検索することは容易ではない。

## がん登録症例の識別モデル

がん登録症例を識別する統計モデルのアイデアは  
病名と診療行為の組み合わせから予測確率を算出



患者基本情報	性別、年齢、紹介機関、既往歴情報
診療行為情報	内服薬情報、注射薬情報、CT/MRI、検査、エコー
診療行為情報	手術情報、腫瘍マーカー
診療実施機関	手術、がん科、放射線科、腫瘍ケア、診療科別ほか

## 【方法】

レセプトデータセットの対象は2015年1月から2015年12月までの間に、国立がん研究センターから提供されているがん登録の対象となるICD10コード、またはがん登録の対象となる可能性のあるICD10コードに該当した15,357件とした。レセプトデータセットをもとに予測モデルを追加抽出シグナチャーを構築した。構築した統計的手法の判別能は、未知となる評価用データセットに本手法を適用して求めた判別結果の感度と特異度をもとに評価した。



レセプトデータセットは、がん登録  
例の検索対象に該当する病名に該当し  
た症例を対象に、当該病名が付与され  
た前後の診療行為情報をレセプトデー  
タから生成した。

がん登録症例の検索は、がん病名の付与されていない症例やがん病名が付与  
されていない症例対象とからない症例をひとつひとつ確認している。

## 解析のフローチャート



データセットの生成

予測モデルの構築

予測モデルの評価

## 【結果】

レセプトデータセットから11項目の説明変数をもつ予測モデルが構築された。説明変数にはICD10のCコードの該当を示す変数のほか、病理診断や薬物療  
療特異物質必要管理料などの算定実績を示す変数が選択された。予測モデルのROC  
曲線から求めたAUCは0.978であった。評価用データセットから求めた予測モ  
デルによる判別の感度は93.6%、特異度は94.5%であった。また予測モデルの  
結果に追加抽出シグナチャーを加えた最終的な判別能の評価となる感度と特異度はそ  
れぞれ、96.3% (95%CI: 94.0%-97.9%)、80.2% (95%CI: 79.3%-  
81.1%)となった。



がん登録症例の予測モデル

## 評価用データセットを用いたがん登録症例の判定結果



## 【考察】

レセプトデータセットから構築した予測モデルの判別能は、感度93.6%、特異  
度94.5%の判別成績を得た。また本予測モデルに追加抽出シグナチャーを加えた結果、  
最終的な感度は95%を上回る成績を得た。これは施設類型や医療に関係なく、レ  
セプトデータをもつ多くの施設で効率的かつ精度の高いがん登録のケースファ  
インディングを実現できる可能性を示唆した。



本手法は感度96.3%に達して、がん登録症例の検索対象の75%を除外した。

## 本研究成果はどのような意義をもたらすか?



がん登録業務の効率化に寄与する作業の一部やその多くを効率化することで、  
がん登録業務者が行っている仕分け作業を中心とした業務からより付加  
価値の高い業務への移行機会になり得る。

※ 本論文は、小原仁、水谷駿介、小森園康二「レセプトデータセットを用いたがん登録のケースファインディング」  
東北病院計画 2018;24: 1-12