

厚生労働科学研究費補助金(臨床研究等ICT基盤構築・人工知能実装研究事業)

分担研究報告書

研究計画・予後予測法への統計学的助言

研究分担者 菊池 隆 臨床研究情報センター 上席研究員

研究要旨：2005/7/1 から2007/6/30までに登録されたJPOPSコホート1症例2354例から、重複登録例2例、登録後不適格例1例、小線源治療未実施12例を除き、安全性解析対象例2339例を選択した。このうち、JPOPS登録の同意取得前に保険医療として小線源療法を実施された23例を除いて2316例を有効性解析対象例とし、匿名化された本データを使用し、PSA非再発率、および有害事象発生率と種々の特徴量との関係についての、機械学習での解析を開始した。

A．研究目的

本研究の目的は、ヨウ素125シード線源を用いた小線源療法に関する前向きコホート研究(JPOPS, Japanese Prostate Cancer Outcome Study of Permanent I-125 seed Implantation)によって得られたビッグデータを用いて、詳細な臨床情報を機械学習させることにより、新しい前立腺癌の予後予測システムを開発することである。

B．研究方法

2017年7月28日に臨床研究情報センターにて、研究代表者中村和正、研究分担者馬込大貴と、JPOPS研究の登録データセットをどのように取り扱うか、特許等を取得できた場合の契約等について討議した。

2017年10月19日、2018年1月20日の班会議にて、研究分担者および研究協力者に、本研究の概要、今後の研究計画について、討議を行った。

さらに、2018年3月28日に臨床研究情報センターにて、研究代表者中村和正、研

究分担者馬込大貴と現在までの機械学習による予後予測結果の精度について、討議を行った。

(倫理面への配慮)

本研究はすでにJPOPSで登録され、匿名化された既存データのみを用いる観察研究であり、患者への侵襲は伴わない。JPOPS研究のコホート1およびコホート2のデータセット原本については臨床研究情報センターにおいて厳重に管理されている。また、駒澤大学へのデータの移送においては、フォルダにパスワードにて暗号化した。

C．研究結果

2005/7/1 から2007/6/30までに登録されたJPOPSコホート1症例2354例から、重複登録例2例、登録後不適格例1例、小線源治療未実施12例を除き、安全性解析対象例2339例を選択した。このうち、JPOPS登録の同意取得前に保険医療として小線源療法を実施された23例を除いて2316例を有効性解析対象例とした。

浜松医科大学、駒沢大学にて、本解析について倫理審査で承認を得た後に、匿名化されたJPOPSデータをセキュリティを確保した上で送付し、駒澤大学にて、PSA非再発率、および有害事象発生率と、種々の特徴量との関係についての、機械学習での解析を開始した。

また、「研究方法」で述べた主任研究者、研究分担者との研究討議を行った。

D . 考察

J-POPSのデータ解析の結果、治療開始後5年以内に原病死する確率は極めて低いことが判明した。この状況で原病死をアウトカムとした解析したモデルを作成しても感度・特異度共に高い結果は得られにくいと考えられた。そこで、今後の方向として、アウトカムを変更して、臨床再発を目的変数としてモデル開発をすることを推奨した。説明因子としてはPSA再発が考えられるが、PSA再発は前向きで評価した場合、PSAバウンスと区別できないことが最大の問題となる。また、PSA再発と判定されても、救済治療が行われるのはPSA濃度が4ng/mlを超える場合がほとんどであった。つまり、医師の治療介入判断はPSA再発の有無ではなく、「PSA濃度が4ng/ml以上」に結びついていた。従って、予後因子としてのPSA再発に臨床的意義がある可能性は小さく、慎重に判断する必要がある。

PSAバウンスの有無と臨床再発との関連はJ-POPSのデータ解析の結果によって認められており、今後、臨床現場では「PSA再発の有無」よりも「PSAバウンスの有無」に関心が移行してくると予想される。した

がって、構築すべきモデルとしては、臨床再発を目的変数、PSAバウンスを説明因子に加えたモデルである。ただし、PSAバウンスの定義は様々であるので、まず治療開始後600日以内のPSA濃度の局所最大値から局所最小値を引いた値(連続変数)をいれたモデルで探索を始めるのが良いと考える。また、PSAバウンスが認められず、典型的なexponential curveを描いて治療に反応した症例も多いことから、他の説明因子、特に年齢によって層別して検討する必要がある。これらの点を考慮すると、まずtree based model から初めて、アンサンブル学習(ランダムフォレスト)にグレードを上げて最適モデルを探している手順が研究としては効率的と思われる。最終的には、機械学習によって、臨床再発を予測する感度・特異度の高いPSAバウンスの定義(局所最初値及び最大値の時期(治療開始からの期間)、最小値及び高低差:これらを変数としてモデルに入れる)を発見することができれば、アルゴリズムとして特許申請を考慮すべきと考えられた。

E . 結論

JPOPSコホート1症例について、PSA非再発率、および有害事象発生率と、種々の特徴量との関係についての機械学習での解析を開始した。

F . 研究発表

1. 論文発表

なし

2. 学会発表

なし

G . 知的財産権の出願・登録状況

(予定を含む)

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし