

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)

総括 研究報告書

ビッグデータからの機械学習による前立腺癌小線源療法の予後予測法の
開発と均てん化への応用

研究代表者 中村 和正 浜松医科大学 医学部 教授

研究要旨：

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) のコホート1の5年PSA非再発生存率は、Phoenix基準にて89.1%、JPOPS基準にて91.6%と良好でPhoenix基準では年齢のみが有意な因子となった。また、治療症例数が多いほど、放射線治療の質の指標である治療パラメータが良好となる傾向にあることが判明した。ロジスティック回帰、サポートベクターマシン(SVM)、ランダムフォレスト(RF)、ディープラーニング(DNN)を用いて、PSA再発、有害事象発生率の予測を改善できるかを検討したが、学習データではRFやDNNで予測精度が上がる傾向にあったが、テスト症例の予測精度は現時点ではいずれも低かった。PSA再発とされた180例のうち、ホルモン療法等の治療の介入無しに再びPSAが低下傾向にある74例 (PSA bounce) を除き、他病死53例を打ち切りとし、T因子、PSA値、Gleasonスコア、年齢、前立腺体積、Total BED (biological equivalent dose)、外照射有無の計7つの因子について、ロジスティック回帰分析を行い、各々の症例でのPSA非再発確率を求め、この確率によりリスク分類を行うと従来のリスク分類より、明確にPSA再発を予測できた。今後、SVM、RF、DNNにて解析を行う予定で、最終年度は、JPOPSコホート1, 2を合わせた約7000例にて本手法のバリデーションを行う予定である。

研究分担者

斉藤史郎 東京医療センター・医長
萬 篤憲 東京医療センター・医長
馬込大貴 駒沢大学・講師
小島伸介 臨床研究情報センター・TRI専門職
菊池 隆 臨床研究情報センター・上席研究員

A . 研究目的

前立腺癌の放射線治療の予後因子としては、T因子、PSA値、Gleason分類などがある。これらによって低・中・高リスクに分類され、大まかな予後予測が可能で、

それに沿ったノモグラムが作成されている。しかし、それ以外の患者背景や検査所見、照射線量、治療パラメータなどの多くの因子を網羅的に分析し、治療結果を予測する方法については国内外を含めてほとんど研究が進められていない。

我々は、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) を実施してきた。本研究に

は全国74施設(小線源療法を施行する施設の約70%)が参加し、2005年から2010年までに小線源療法で治療された約7000例(同時期に本邦で小線源療法により治療された症例の約40%)が前向き登録された。調査項目は、患者・家族背景、生活歴、T因子、PSA値、Gleason分類などの腫瘍因子、小線源治療パラメータ、外照射併用の有無等の治療因子、PSA再発の有無、生死、有害事象の有無などで、少なくとも5年以上の経過観察が行われている世界的に類を見ないビッグデータである。2005年から2007年までのコホート1の2339例のデータクリーニングが終了し、2010年までのコホート2の約4600例についても平成28年11月で5年の経過観察が終了し、これらのビッグデータが使用可能となる予定である。

本研究の目的は、JPOPS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し、放射線治療の質の均てん化に資することである。

B . 研究方法

JPOPS 研究のコホート1(2005-2007年)で収集された約2000例のデータのクリーニングはすでに終了している。平成29年度にはコホート1の患者背景、臨床情報、治療パラメータと、PSA非再発率、全生存率、有害事象発生率等の治療結果との関係について機械学習による評価を開始した。平成30年度は本研究を継続し、以下のように研究を進めた。

1) 放射線治療施設の規模と治療の質

本邦の放射線治療の均てん化に資するために、先進国における放射線治療施設数、外部照射機器数との関係を調べた。一人当たりGNI25,000ドル以上の国で、癌罹患数が見える先進23カ国で放射線治療施設と放射線治療装置数を比較した。放射線治療施設および外照射台数については、DIRAC (Directory of Radiotherapy Centres) database 2017を用いた。また、各国の癌罹患数については、2012年のOECD Health Dataを用いた。

次に、治療症例数と治療の質の関係を調べるために、JPOPSコホート1に120例以上登録している大規模5施設と、120例未満を登録した20施設について、D90(Gy), V100(%), V150(%), 尿道D5(Gy), U200(ml), 直腸R100(ml), R150(ml)の治療パラメータと登録数との関係を調べた。Learning curveの影響を除くため、治療開始から25例は除き、シード治療単独例のみを解析した。

2) JPOPSコホート1での予後因子解析

予後予測因子のエンドポイントであるPSA再発については、Phoenix基準(PSAが治療後最低値Nadirとなり、その後PSAが上昇し、Nadir+2ng/mLとなった時点を再発とする)が国際的に用いられている。しかし、前立腺癌の放射線治療、特に小線源治療では、PSAが一時的に上昇し、その後再び低下する、いわゆるPSA bounceと呼ばれる現象が発生することが知られている。そこで、JPOPSでは、PSA bounceを正しく評価するために、Phoenix基準に加えて、JPOPS基準(PSAが1を超えて3回以上上昇した場合を再発とする)を定義している。

まず、JPOPS コホート1の 2316 例全例を用いて、PSA 非再発生存率および予測される予後因子（年齢、T 因子、治療前 PSA 値、Gleason score (GS)、照射線量等）が PSA 非再発生存率に影響するかどうかを調べた。

3) JPOPS コホート1での機械学習による予後予測精度向上の検証

予後予測因子のエンドポイントである PSA 非再発生存率については、国際的に用いられている Phoenix 基準を用いた。POPS コホート1のデータを用いて、ロジスティック回帰、サポートベクターマシン(SVM)、ランダムフォレスト(RF)、ディープラーニング(DNN)の機械学習手法を用いて、PSA 再発、有害事象発生率の予測を改善できるかを検討した。入力特徴量は、予後と関連がありそうな特徴量のみ厳選した Limited Database (12 項目) となるべく多くの特徴量を採用した Large Database(44 項目)の2つのデータセットを用いた。評価法は 10-fold cross-validation 法を用いて Accuracy (全症例にて正しく予測できた症例の割合)、Precision (再発と予測した症例にて真に再発した症例の割合)、Recall (再発症例にて再発と予測した症例の割合)を評価指標とした。

4) 小線源治療後の PSA kinetics の解析

POPS コホート1のデータを用いて、小線源治療後の PSA の経時的な動き、すなわち PSA kinetics を調べた。統計解析ソフトウェア JMP13 を用い、小線源挿入からの PSA 値の変化の平均値を実線にて、存在確率を4段階の密度分布で表示させた。

5) 新しい基準での PSA 非再発率と予後因子との関係

PSA 非再発率と予後因子との関係を評価するために、T 因子、PSA 値、GS に加え、単変量解析や過去の論文から PSA 非再発率に影響を与えると考えられる、年齢、前立腺体積、Total BED (biological equivalent dose)、外照射有無の計7つの因子について、ロジスティック回帰分析を行った。回帰係数より、各々の症例での PSA 非再発確率 $p(x)$ を下記式より求めた。

$$\log \frac{p(x)}{1-p(x)} = 0.032 \times age - 0.749 \times GS - 0.181 \times T - 0.069 \times PSA + 0.008 \times volume - 0.549 \times ERT + 0.004 \times BED + 5.996$$

$$z = \log \frac{p(x)}{1-p(x)} \text{ とすると、} p(x) = \frac{1}{1+\exp(-z)}$$

*ただし、T: T1a-c=1, T2a=2, T2b=3, T2c=4, T3a-b=5 外照射有無:無し=1

PSA 非再発確率 $p(x)$ により、が 0.97 以上を Risk A、0.97-0.92 を Risk B、0.92 未満を Risk C として、新たなリスク分類とし、PSA 非再発率を求めた。

(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。また、JPOPS研究のコホート1およびコホート2のデータセット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。

本年度、JPOPS研究のコホート1のデータを用いて解析を開始したが、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。また、解析用のパソコンにお

いてもパスワード等で厳重に管理されている。

研究の透明性確保のため、研究代表者の所属機関（浜松医科大学）において倫理審査を受け、平成29年7月26日に倫理審査委員会の承認を得ている（研究番号17-095）。また、解析を担当する駒沢大学においても同様に平成29年7月21日に倫理審査委員会の承認を得ている（通知番号17-18）。

匿名化された既存データのみを用いる観察研究ではあるが、今後も、ハード、ソフトウェアおよびデータの移送のすべてのレベルでセキュリティを強固にする。

C . 研究結果

1) 放射線治療施設の規模と治療の質

癌患者数当たりの外部照射装置台数は、先進国間ではほぼ同じであったが、放射線治療施設数当たりの外部照射装置数は、癌患者当たりの施設数と反比例の関係にあることが判明した。本邦では放射線治療施設数当たりの外部照射装置数は 1.24 と先進国中で最も小さかった。また施設当たり 3 台以上の外部照射数を持つ大規模施設の割合も最も少なかった。すなわち、外部照射装置台数が 1 - 2 台の小規模な施設が多数分散して存在していることが明らかとなった（文献 1）。

次に、JPOPS コホート 1 に 120 例以上登録している大規模 5 施設と、120 例未満を登録した 20 施設にて、小線源単独症例（1126 例）の治療パラメータを比較した。大規模施設では、ほとんどの症例で D90 は 144 Gy 以上が投与されていたが、小規模施設では 144 Gy 未満の症例も有る程度

の比率で認められた。V100 は大規模施設で有意に良好であった。一方、尿道 U200(ml)は小規模施設でより大きくばらついていた（文献 2）。

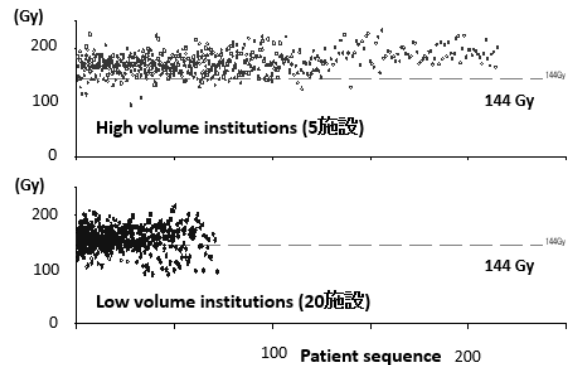


図 1. 施設規模と治療パラメータ

大規模施設とそれ以外の20施設での、前立腺への投与線量（D90）を登録症例順に表示した。

144 Gy以上の投与線量が推奨されており、大規模施設ではほとんどの症例で144 Gy以上が投与されており、症例を登録するほど、十分な線量が投与される傾向にあることがわかる。

2) JPOPS コホート 1 での予後因子解析

JPOPS コホート 1 の 2316 例全例の PSA 非再発生存率は、観察期間中央値 60 カ月にて、5 年 PSA 非再発生存率は、Phoenix 基準にて 89.1%、JPOPS 基準にて 91.6%と極めて良好であった。SA 非再発生存率に影響する因子について調べたところ、JPOPS 基準では、PSA 値、GS、T 因子が有意な因子であったが、Phoenix 基準では年齢のみが有意な因子となった（文献 3）。

3) JPOPS コホート 1 での機械学習による予後予測精度向上の検証

Phoenix 基準を用いて、ロジスティック回帰、SVM、RF、DNN の機械学習手法を用いて、PSA 再発、有害事象発生率の予測を改善できるかを検討した。いずれも学習データでは予測精度は 0.7-0.9 以上と高い値を示し、は RF や DNN で予測精度

表1. 機械学習によるPSA再発予測

		Accuracy		Precision		Recall	
		training	test	training	test	training	test
Large DB PHOENIX定義 PSA再発予測	Logist	0.765	0.680	0.736	0.129	0.828	0.472
	SVM	0.717	0.581	0.674	0.111	0.841	0.551
	RF	0.992	0.900	0.994	0.163	0.991	0.027
	DNN	0.866	0.834	0.807	0.112	0.826	0.183
Limited DB PHOENIX定義 PSA再発予測	Logist	0.573	0.590	0.578	0.091	0.542	0.460
	SVM	0.568	0.660	0.592	0.095	0.452	0.376
	RF	0.994	0.852	0.994	0.091	0.993	0.106
	DNN	0.727	0.559	0.681	0.097	0.875	0.510

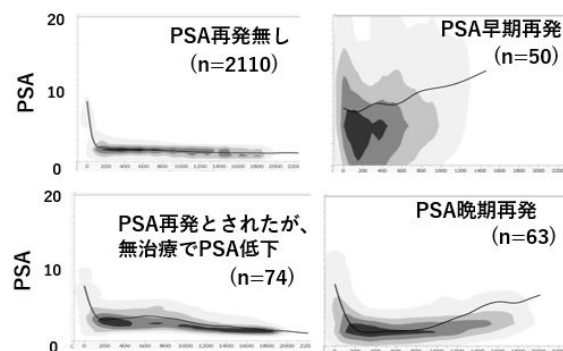
が上がる傾向にあったが、テスト症例の予測精度は現時点ではいずれも低かった (Magome T, et al: RSNA 104th scientific assembly and annual meeting, Chicago, 2018)。

4) 小線源治療後の PSA kinetics の解析

上記のように、国際的に用いられている Phoenix 基準では、PSA 非再発生存率と一般的な前立腺癌の予後因子との関係は機械学習を用いてもほとんど見いだせなかった。

その原因を探るため、小線源治療後の PSA の経時的な動き、すなわち PSA kinetics を調べた。

Phoenix 基準にて PSA 再発とされた 180 例のうち、74 例が PSA 上昇後、ホルモン療法等の治療の介入無しに再び PSA が低下傾向にあることが判明した。また、PSA 再発にも早期に PSA 上昇が認められる早期パターンとかなり時間がたって PSA が上昇する 2 つのパターンがあることが判明した (図 2)。



小線源挿入からの日数
PSAの平均値を実線で、存在確率を4段階の密度分布で表示

図 2. 小線源治療後の PSA kinetics

5) 新しい基準での PSA 非再発率と予後因子との関係

上記 74 例を PSA 再発とはせずに解析を行うこととした。また、前立腺癌小線源治療後は、通常は再発しない限り治療の介入は行わないため、他病死は小線源治療に関連していないと解釈し、他病死 53 例を打ち切りとして、再度予後因子解析を行った。まず、PSA 再発に寄与する因子を明確にするために、PSA 非再発例で、経過観察期間が 3 年未満の症例 160 例を除いて、2135 例で解析した。NCCN ガイドラインでは、前立腺癌では、T 因子、PSA 値および GS にて、低リスク (T1-T2a, PSA<10、かつ GS 6 以下)、高リスク (T3 以上、PSA>20、または GS 8-10)、中リスク (上記以外) に分類される。5 年 PSA 非再発率は、低、中、高リスク各々 96.3%、93.9%、91.0% で、有意差を認めしたが、その差はわずかであった。

そこで、機械学習による新たなリスク評価を T 因子、PSA 値、GS に加え、単変量解析や過去の論文から PSA 非再発率に影響を与えると考えられる、年齢、前立腺体

積、Total BED (biological equivalent dose) 外照射有無の計 7 つの因子について、ロジスティック回帰分析を行い、回帰係数より、各々の症例での PSA 非再発確率 $p(x)$ を求め、PSA 非再発確率から、Risk A、Risk B、Risk C に分類した。

各々の 5 年 PSA 非再発率は、Risk A 98.8%, Risk B 94.7%, Risk C 87.8% ($P < 0.0001$) となり、従来のリスク分類より、明確に PSA 再発を予測できた。

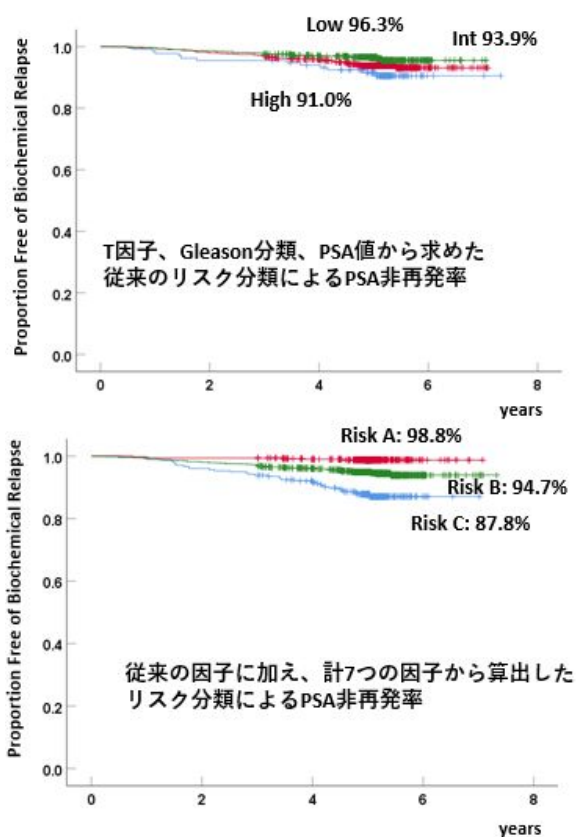


図 3. 従来のリスク分類と PSA 非再発確率 $p(x)$ から求めたリスク分類

検証のため、2135 例を Training 1706 例、Test 429 例に分け、Training 群から回帰計数を求め、Test 群の各症例の PSA 非再発確率 $p(x)$ を計算し、リスク群に分類したところ、Risk C の PSA 非再発確率は

やや不良となり、ほぼ同様な傾向が認められた。

D. 考察

今回の研究で、本邦では、他の先進国と比較し、外部照射装置台数が 1 - 2 台の小規模な施設が多数分散して存在していることが明らかとなった。また、JPOPS 登録施設において、登録症例数が多い施設では、治療パラメータが良好な傾向にあることが明らかとなった。そのような中、治療の質の均てん化のためには、治療パラメータも含めた、新しい予後予測システムを確立することは極めて重要である。

我々は、まず、JPOPS コホート 1 のデータを用いて、ロジスティック回帰、SVM、RF、DNN の機械学習手法にて、PSA 再発、有害事象発生率の予測を改善できるかを検討した。いずれも学習データでは予測精度は 0.7-0.9 以上と高い値を示したものの、テスト症例の予測精度は現時点ではいずれも低く、機械学習で正確に再発するかどうかを予測することは困難であることが明らかとなった。

その原因として、前立腺癌小線源療法では、PSA 再発の頻度が極めて少ないため、予測が難しいこと、および、今回の JPOPS で収集したデータ以外、例えば腫瘍の遺伝子情報、患者の生活習慣等が再発に係わっている可能性があること、の 2 つを考えている。

そのため、個々の症例が再発するかどうかを予測するのではなく、予後に関与している因子を用いて、機械学習によって、従来のリスク分類よりも、より正確に、予後の良いまたは悪い群の抽出ができないかを調べることにした。

まず、PSA非再発確率 p を初歩的な「機械学習」のひとつであるロジスティック回帰分析にて求め、リスク分類することにより、従来のリスク分類より、明確にPSA再発を予測できる可能性があることが判明した。

今後、SVM、RF、DNNにて解析を行う予定で、さらに予測精度が向上する可能性がある。最終年度は、JPOPSコホート1, 2を合わせた約7000例について、約6000例で予後予測アルゴリズムの構築を行い、残り1000例でそのバリデーションを行う予定である。

E . 結論

今回の研究で、機械学習によりPSA非再発確率を求めてリスク分類することにより、PSA非再発率をより正確に予測できる可能性が示唆された。

さらに平成31年度にてコホート2と合わせて7000例を解析し、予後予測アルゴリズムを確立できれば、以下のような活用が期待できる。

1) 小線源療法にて治療された患者で、治療時点の背景因子、治療パラメータ等を入力することで、従来のリスク分類より、さらに正確にPSA非再発率を予測できる。

2) 治療結果が正確に予測されれば、再発リスクの低い症例では経過観察期間を延ばすことが可能となる。また、再発リスクが高いと判断されれば、密な経過観察を行い、再発の時点で早期に治療介入することが可能となる。このように個別化した介入が可能となり、医療費削減効果が期待できる。

3) 特にPSAが一時的に上昇するPSA

Bounceでは、再発か非常に判断が難しく、しばしばホルモン療法が施行されることも多いが、予後予測アルゴリズムで再発の確率が低いと判断されれば、無駄にホルモン療法が施行することがなくなる可能性がある。

4) 本研究で機械学習による予後予測法が確立されれば、その研究プロセスを応用して、小線源療法のみならず、他の前立腺癌放射線治療モダリティの予後予測アルゴリズムの確立が期待される。

F . 健康危険情報

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。

G . 研究発表

1. 論文発表

- 1) Nakamura K, Konishi K, Komatsu T, Sasaki T, Shikama N. Patterns of radiotherapy infrastructure in Japan and in other countries with well-developed radiotherapy infrastructures. *Jpn J Clin Oncol* 2018, 48:476–479.
- 2) Nakamura K, Ohga S, Yorozu A, Saito S, Kikuchi T, Dokiya T, Fukushima M, Yamanaka H. Institutional patient accrual volume and the treatment quality of I-125 prostate seed implantation in a Japanese nationwide prospective cohort study. *Strahlenther Onkol.* 2019, 195: 412–419.
- 3) Ito K, Saito S, Yorozu A, Kojima S,

- Kikuchi T, Higashide S, Aoki M, Koga H, Satoh T, Ohashi T, Nakamura K, Katayama N, Tanaka N, Nakano M, Shigematsu N, Dokiya T, Fukushima M; J-POPS Investigators. Nationwide Japanese Prostate Cancer Outcome Study of Permanent Iodine-125 Seed Implantation (J-POPS): first analysis on survival. *Int J Clin Oncol*. 2018, 23:1148-1159.
- 4) Katayama N, Yorozu A, Nakamura K, Fukushima M, Kikuch T, Saito S, Dokiya T. Biochemical outcomes and predictive factors by risk group after permanent iodine-125 seed implantation: Prospective cohort study in 2,316 patients. *Brachytherapy* 2019, in press.
- 5) Tanaka N, Yorozu A, Takashi Kikuchi, Higashide S, Kojima S, Ohashi T, Katayama N, Nakamura K, Saito S, Dokiya T, Fukushima M; the J-POPS Study Group. Genitourinary Toxicity after Permanent Iodine-125 Seed Implantation: The nationwide Japanese Prostate Cancer Outcome Study of Permanent Iodine-125 Seed Implantation (J-POPS). *Brachytherapy* 2019, in press.
- 2. 学会発表**
- 1) 中村和正. 「機械学習による前立腺癌小線源療法の子後予測法の開発の概要と今後」第14回 J-POPS中間報告会プログラム. 2019.01.19 東京
- 2) 中村和正. 外照射療法. パネルディスカッション「合併症・QOL からみた各種限局がん治療の比較」第34回 前立腺シンポジウム 2018.12.9-12.9 東京
- 3) Katsumasa Nakamaura. Fractionation and time factor in radiation therapy for prostate cancer. (Symposium) JASTRO-ESTRO Joint Symposium. The 31st Annual Meeting of the Japanese Society for Radiation Oncology (JASTRO2018) Kyoto, H30.10.11-13.
- 4) 中村和正、小西憲太、小松哲也、石場領、太田尚文、若林紘平. 機械学習による音声認識技術および汎用知的音声合成システムを利用した大学病院における放射線治療業務効率化への試み. 第二回山梨静岡放射線治療研究会. 静岡市 2018.5.3
- 5) Nakamura K, Yorozu A, Saito S. The pattern of infrastructure of radiotherapy in Japan and the treatment quality of brachytherapy for prostate cancer. The 6th Taiwan-Japan Radiation Oncology Symposium. Taichung, Taiwan, 2018.5.9
- H. 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