

一般口演 | 医療データ解析

一般口演25

医療データ解析

2019年11月24日(日) 13:40 ~ 15:10 F会場 (国際会議場 3階中会議室302)

[4-F-3-02] クリニカルパスと診療テキストを用いた糖尿病併存胃がん患者の Patient Reported Outcome 指標の検討

○山下 貴範¹、伊豆倉 理江子¹、野尻 千夏¹、野原 康伸¹、中島 直樹¹（1.九州大学病院）

キーワード：Patient Reported Outcome, Clinical Pathway, Dependency parsing, Electronic medical record

【背景】近年、糖尿病有病者のがん罹患率が高いことが明らかとなった。糖尿病の併存は、がんの治療成績を下げ、生存に関する予後が悪く、合併症のリスクが高く、化学療法の効果下がるとの報告がある。糖尿病併存のがん患者は非併存患者に比べ、病態や治療に伴う症状・機能・QOLの変化が大きいと推察される。この状態の把握には、患者自身の主観的指標：Patient Reported Outcome (PRO)が重要となるが、本邦では、質の高いPRO指標が実用化されておらず、糖尿病併存がん患者への信頼性・妥当性の高いPRO指標はない。

周術期の過程において、治療方針やアウトカム指標はクリニカルパス（以下、パス）で一元化されつつあるが、現行のパスのアウトカム評価は医療者視点のため患者発信の情報収集は困難であり、それらは自由文の診療テキストに記載されている。本研究では、胃がん患者を対象としてパスと診療テキストからQOLの向上を目指した経時的なPRO指標を検討する。

【方法】九州大学病院における2013年から2018年の胃がん症例のパスと診療テキストデータを対象とする。当院のパスは、アウトカム志向型パスを運用しており、患者状態や患者アウトカムが構造化データとして蓄積される特徴を持つ。診療テキストは、医師記載の経過記録と看護記録の Subjective、Objective項目を抽出した。各データを入院日や手術日を考慮した時系列変数として整理した。さらに糖尿病併存、非併存で分類し機械学習を用いて、特徴語を抽出した。

【まとめ】胃がん患者のPROとしての自覚症状や機能などに関連する語を対象にして、時系列推移と糖尿病併存患者の特徴語を評価した。その結果とQOL評価のEORTC QLQ-C30やSF-12などと比較を行い、PRO指標として考察した。

今後は、本結果を用いて電子版PRO（e-PRO）評価ツールを開発する予定である。

クリニカルパスと診療テキストを用いた糖尿病併存胃がん患者の Patient Reported Outcome 指標の検討

山下貴範^{*1}、伊豆倉理江子^{*1}、野尻千夏^{*1}、野原康伸^{*1}、中島直樹^{*1}

^{*1}九州大学病院メディカル・インフォメーションセンター

Investigation of Patient Reported Outcome Index for Gastric Cancer Patient with Diabetes Mellitus Using Clinical Pathway and Electronic Medical Record

Takanori Yamashita^{*1}, Rieko Izukura^{*1}, Chinatsu Nojiri^{*1}, Yasunobu Nohara^{*1}, Naoki Nakashima^{*1}

^{*1} Medical Information Center, Kyushu University Hospital

The comorbidity of diabetes mellitus has been reported to worsen cancer treatment, be poor prognosis, increased risk for complications and reduce the effectiveness of chemotherapy. The cancer patients with diabetes mellitus are presumed to have large changes in symptoms, functions and quality of life associated with clinical condition and treatment. Patient's own Patient Reported Outcome (PRO) is important for understanding those conditions. This study aimed to investigate PRO index to patient of gastric cancer using an outcome-oriented clinical pathway and electronic medical record. We applied dependency parsing to electronic medical record to extract the temporal words that represent PRO. Variance data of clinical pathway and the words were analyzed for diabetes mellitus and long-term cases by machine learning.

Keywords: Patient reported outcome, Clinical pathway, Dependency parsing, Electronic medical record.

1. はじめに

平成 28 年「国民調査・栄養調査」(厚生労働省)によると、糖尿病の有病者数と予備軍はいずれも約 1000 万人と推計しており、合併症の中でもがんにおいて糖尿病有病者が非有病者に比べ発症のリスクが 1.2 倍に上昇し、中でも消化器がんが有意に高い。がんの治療法の進歩により、5 年相対生存率も 6 割まで上昇していることから、今後は糖尿病を併存するがん患者数が急増していくと推定される。糖尿病の併存は、がんの治療成績を下げ、生存に関する予後が悪く、合併症のリスクが高く、化学療法の効果が下がるとの報告がある¹⁻⁴⁾。糖尿病併存のがん患者は非併存患者に比べ、病態や治療に伴う症状・機能・QOL の変化が大きいと推察される。この状態の把握には、患者自身の主観的指標: Patient Reported Outcome (PRO) が重要となるが、本邦では、質の高い PRO 指標が実用化されておらず、継続的なデータ収集基盤も構築されていない。さらに糖尿病併存がん患者への信頼性・妥当性の高い PRO 指標はない。

周術期の過程において、治療方針やアウトカム指標はクリニカルパス(以下、パス)で一元化されつつある。本来、病院情報システムにて、患者状態や診療行為の目的、診療のプロセス、患者のアウトカム評価については自由文で診療テキストデータとして記録されているが、パスはこれらの情報を構造化データとして蓄積することが可能である。日本クリニカルパス学会監修の BOM(ベーシックアウトカムマスタ)⁵⁾による標準化により、複数施設の統合データにより、精緻な診療プロセス解析の取り組みが始まっている(AMED ePathプロジェクト)⁶⁾。但し、現行のパスのアウトカム評価は医療者視点のため患者発信の情報収集は困難であり、それらは自由文の診療テキストに記載されていると考える。

本研究では、胃がん患者を対象としてパスデータと診療テキストデータから自然言語処理と機械学習を用いた解析を行い、QOL の向上を目指した経時的な PRO 指標を検討する。

2. データ・手法

九州大学病院における 2013 年から 2018 年の胃がん症例の 638 例を対象とする。その入院基本情報とパスデータと診療テキストデータを病院情報システムから抽出した。

当院のパスは、アウトカム志向型パスを運用しており、BOM を元にアウトカムを設定している。該当する観察項目により、アウトカムの達成 or 未達成(バリエーション)の評価が記録される。診療テキストデータは、医師記録と看護記録の Subjective、Objective 項目を対象とした。さらに、入院期間中に 2 型糖尿病の病名の有無、入院日、手術日、パス適用期間、退院日を紐づけた。本パスは術後 10 日目で標準パス終了であり、13 日以上を長期パス症例として扱った。

638 症例のうち、2 型糖尿病の有病は 131 例、長期パス症例は 62 例であった。

そして、以下の 2 つの手順で解析した(図 1)。

- 1) 該当症例の診療テキストデータ(医師記録、看護記録)は、CaboCha(ver. 0.69)⁷⁾による係り受け処理を行い、“係り元”と“係り先”の単語に分類した。次に、“係り元”を主語、“係り受”を述語の形式を抽出するため、“係り元”と“係り先”の単語に対して McCab(ver. 0.996)⁸⁾による形態素解析を行い、“係り元”は名詞または副詞、“係り受”は名詞、動詞、形容詞、形容動詞を抽出した。辞書は ComeJisyoSjis-1⁹⁾を利用した。その中から出現頻度の大きい単語から、人手によるチェックで PRO に関連する単語と“係り元・係り先”の組合せを抽出した。そして、時系列に出現頻度の傾向を可視化した。
- 2) “係り元・係り先”のバリエーション表現を選択し、パスのバリエーションデータと組合せた。これらに術後相対日を付与し、時系列データセットを作成した。そして、糖尿病併存と長期パス症例に対する Gradient Boosting Decision Tree (GBDT)¹⁰⁾と、野原ら(2019)¹¹⁾による Shapley Additive Explanation (SHAP)を用いた特徴量抽出を行った。

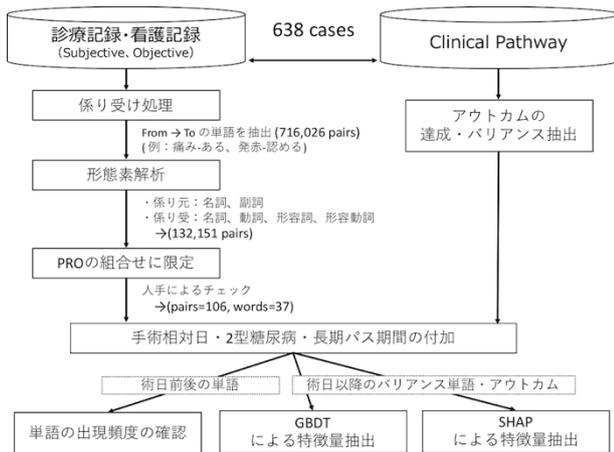


図1 実験手順

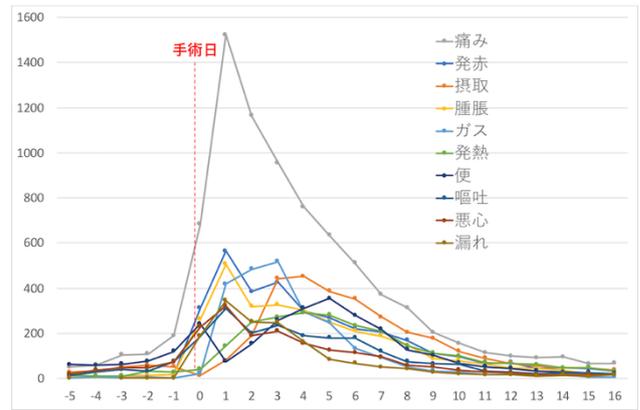


図2 手術日相対日でみた単語の出現頻度(上位10)

3. 結果

3.1 頻出単語の傾向

医師記録と看護記録の Subjective、Objective 項目に対して、係り受けを実施し、“係り元・係り先”の組合せは 710,026 件抽出された。次に、“係り元”は名詞または副詞、“係り先”は名詞、動詞、形容詞、形容動詞の組合せとして、132,151 件に絞った。そして、“係り元”に対して出現頻度の大きい単語について、人手により PRO に関連する組合せを 106 件選択し、名詞を中心に類義語を集約し、PRO に関連する単語を 37 件とした。

PRO に関連する単語について、術後の出現頻度順では、“痛み”、“発赤”、“摂取”、“腫脹”、“ガス”と続いた(表1)。各単語の手術日相対日での出現頻度の推移をみると、術後 1 日目に“痛み”、“発赤”、“腫脹”、“悪心”、“漏れ”の単語のピークが見られた。術後 2 日目以降では、“ガス”、“摂取”、“発熱”、“便”の単語のピークが見られた(図 2)。単語の出現頻度の傾向について、糖尿病の有無、長期パス・短期パスにおける大きな差は見られなかった。

表1 術後の出現頻度順の単語(上位20)

	単語	出現頻度		単語	出現頻度
1	痛み	7,810	11	出血	1,414
2	発赤	3,361	12	浸出液	1,182
3	摂取	3,039	13	歩行	1,124
4	腫脹	2,912	14	介助	934
5	ガス	3,377	15	汚染	725
6	発熱	2,400	16	尿	671
7	便	2,402	17	雑音	644
8	嘔吐	1,961	18	離床	585
9	悪心	1,711	19	覚醒	535
10	漏れ	1,572	20	自立	502

3.2 糖尿病有病症例と長期パス適用症例の術後バリエーションの特徴抽出

3.1 章で PRO に関連する単語が含まれる“係り元・係り先”で、術後のバリエーション表現の組合せを 31 にグルーピングした。同様に、術後の 25 個のパスアウトカムのバリエーションデータを統合し、説明変数の数は 56 となった。

糖尿病有病を目的変数とした GBDT (AUC=0.607) は、“痛み”、“摂取(食事)”、“漏れ”、“不眠”、“パスアウトカムは“循環動態が安定している”のバリエーションが上位に出現した(図 3)。SHAP の解析は、GBDT と比較してほとんど同じ傾向である。赤いプロットがバリエーション発生なので、それぞれのバリエーション発生が寄与している(図 4)。

長期パスを目的変数とした GBDT (AUC=0.719) は、“摂取(食事)”、“発赤”、“痛み”、パスアウトカムは“食事摂取ができていない”、“身体的準備ができていない”、“自立した入院生活を送れる”のバリエーションが上位に出現した(図 5)。SHAP の解析では GBDT の変数と若干の上下はあるが、ほとんど同じ傾向である。そして、バリエーション発生が寄与していることが解釈できる(図 6)。

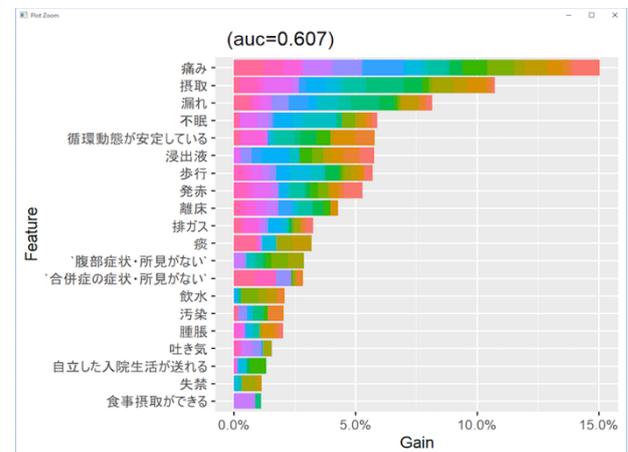


図3 GBDTによる変数重要度グラフ(糖尿病)

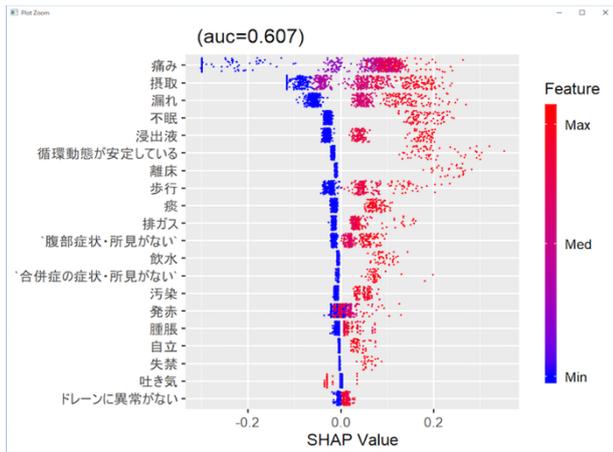


図 4 SHAP による変数重要度グラフ(糖尿病)

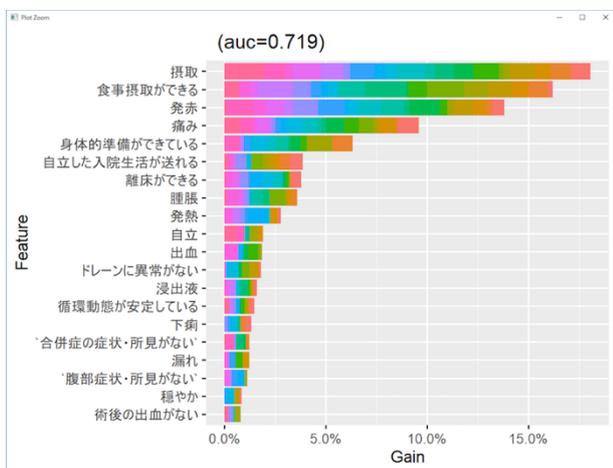


図 5 GBDT による変数重要度グラフ(長期パス)

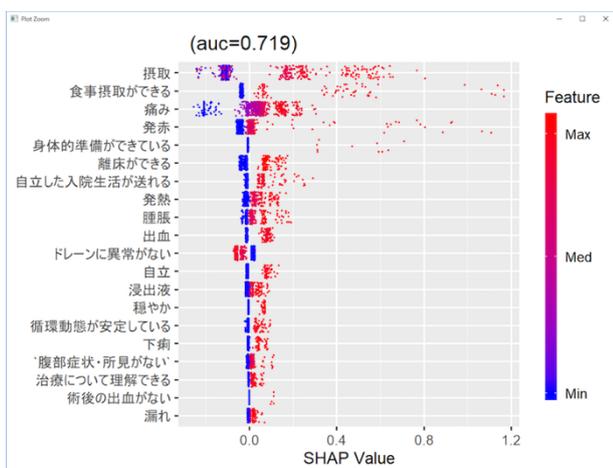


図 6 SHAP による変数重要度グラフ(長期パス)

4. 考察

単語の頻出傾向をみると、術後すぐにピークが出現する単語と、数日遅れてピークが出現する単語(“ガス”、“食事”、“発熱”、“悪心”)がある。これらを参考にしてパスのアウトカム設定することで、診療プロセスの改善につながる可能性もある。また、構造化データとして蓄積されるため、さらに精緻化されたデータ解析も可能となる。

GBDT と SHAP の解析結果として、糖尿病有病では単語の出現頻度と同様の変数が抽出された。また、AUC は 0.607 と低い精度であった。長期パスについては、AUC は 0.719 と比較的高い結果となった。食事に関連する単語とパスアウトカムのバリエーションが 1 位と 2 位であり、SHAP をみてもバリエーション発生の場合に寄与していることがわかった。パス長期では、食事バリエーションがクリティカルな事象であることが考えられる。

病院情報システムに蓄積されたデータは、リアルワールドデータと呼ばれ、利活用によって副作用検知や発症予測、診療プロセス改善などに期待されているが、解析を目的としていない。信頼のある解析結果のためには、正確なデータ収集が必要である。パスは患者状態を構造化しているため精度の高いデータと言える。今回は、それに加えて診療テキストデータの係り受け処理から品詞を絞り、人手による作業で PRO に関連する単語として整理し、解析データとした。これらの処理は必要であるが精度の高いデータとして生成できたと考える。

今回の解析結果からは子版 PRO (e-PRO) 評価ツールの検討材料として、「痛み・食事・ガス」について、特徴的な変数として抽出された。EORTC QLQ-C30¹²⁾や SF-12¹³⁾と比較すると、EORTC QLQ-C30 について、「食欲がないと感じましたか」という表現があるが、もう少し具体的な表現でも良いと考える。また、ガスについても検討対象である。EORTC QLQ-C30 や SF-12 はがん全体のスケールである。胃がん用の EORTC QLQ-STO22¹⁴⁾なども含めて症例特異的な項目も検討する必要がある。

5. 結語

病院情報システムのパスデータと診療テキストデータからバリエーション表現を対象に、胃がん症例の継時的な PRO 指標を検討するために、データ解析を行った。診療テキストデータに対しては、自然言語処理と人手による確認で解析可能なデータセットが生成できた。長期パスの解析では AUC が高く、食事関係のバリエーションが抽出された。今後は、本結果を用いて e-PRO 評価ツールを開発する予定である。

謝辞

本研究は科研費・基盤研究 C (18K10312) の助成を受けたものである。

参考文献

- 1) 加藤明彦, 石井正之, 大田貢由, 森田浩文, 山口茂樹, 大腸がん患者では、耐糖能異常が高率に認められる, 日本病態栄養学会誌 2005; 8 (3):209-212.
- 2) がん患者の周術期管理のすべて・合併症を持つがん患者の周術期管理・糖尿病, 濱田康弘, 上野公彦, 青山倫子, 三好真琴, 宇佐美眞, 外科治療 2011; 104: 192-197
- 3) 春日雅人, 植木浩二郎, 田嶋尚子ほか, 糖尿病と癌に関する委員会報告, 糖尿病 2013; 56 (6):374~390
- 4) 田島加奈子, 福嶋清香, 齋藤聡子, 米田千裕, 荻野淳, 橋本尚武, 高血糖者でのがん発症の臨床的背景と正常血糖者との比

較検討, 糖尿病 2017; 60(6): 434-441.

- 5) <http://www.jscp.gr.jp/bom.html>
- 6) <https://e-path.jp>
- 7) CaboCha/南瓜 : Yet Another Japanese Dependency Structure Analyzer, <http://taku910.github.io/cabocha/>
- 8) MeCab : <https://taku910.github.io/mecab/>
- 9) Comejisyo : <https://ja.osdn.net/projects/comedic/>
- 10) Jerome H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 2001; 29 (5): 1189-1232.
- 11) 野原康伸, 松本晃太郎, 副島秀久, 中島直樹. Shapley Additive Explanation を用いた機械学習モデルの解釈と医療実データへの応用. 2019; 第 23 回日本医療情報学会春季学術大会: 60-61.
- 12) Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993; 85:365-376.
- 13) <https://www.sf-36.jp/index.html>
- 14) Kobayashi K, Takeda F, Teramukai S, et al. A cross-validation of the European Organization for Research and Treatment of Cancer QLQ-C30 (EORTC QLQ-C30) for Japanese with lung cancer [see comments]. Eur J Cancer 1998; 34:810-815.