



2025 年 12 月 25 日

報道関係者各位

慶應義塾大学

電子カルテ「文章」の AI 解析で、既承認薬のがん化学療法における副作用抑制作用を可視化
ー 自然言語処理による臨床テキスト活用が、支持療法開発を加速ー

慶應義塾大学薬学部（土屋雅美特任講師、薬学科 2024 年度卒業生の井上真理氏、堀里子教授）、東京大学大学院医学系研究科（河添悦昌教授）、奈良先端科学技術大学院大学（荒牧英治教授）を中心とする研究グループは、電子カルテの自由記載を自然言語処理（NLP）（注 1）で解析し、がん薬物療法に伴う副作用を大規模に把握するとともに、副作用抑制作用を示す既承認薬を探索できる新規手法を構築しました。本研究では、アンジオテンシン II 受容体拮抗薬（ARB）およびラメルテオンが特定の副作用を抑制しうる可能性が示され、NLP を活用したがん化学療法における支持療法開発の新たなアプローチとなることが期待されます。

本研究成果は、2025 年 11 月 29 日に国際学術誌 European Journal of Cancer に掲載されました。

1. 本研究のポイント

- 電子カルテの自由記載に対して、日本語の医療関連表現に対応した自然言語処理（NLP）を適用し、がん化学療法に伴う主観的な副作用（口内炎、末梢神経障害など）を、5 万 6 千人規模の患者集団で定量的に評価した。
- 既存の高血圧治療薬であるアンジオテンシン II 受容体拮抗薬（ARB）がフルオロピリミジン系薬による口内炎を抑制し（ハザード比 0.58）、睡眠薬ラメルテオンが白金製剤による末梢神経障害を抑制する可能性が示された。
- これらの候補薬はいずれも基礎研究で抗がん薬の副作用抑制作用が報告されており、本研究により、追加の介入や特別なデータ収集を行わずに、既存の臨床データのみを用いて副作用抑制作用を検証できることが初めて示された。
- 本研究で用いた NLP 基盤は、これまでに日本語の医療テキストから副作用を抽出する技術の開発や、電子カルテを用いた既存薬の予防効果の再現検証、抗がん薬に伴う副作用の特徴解析など、研究チームが蓄積してきた一連の成果を発展させたものである。



図 1. 本研究の概念図

本図は、電子カルテの自由記載から副作用情報を自動抽出し、前臨床研究で示されていた候補薬の効果を臨床データで検証するまでの流れを示しています。NLP モデル（MedNERN-CR-JA）が文章中の副作用表現を認識し、標準化された症状名に整理した後、統計解析により副作用の発症パターンや予防薬の効果を評価します。

2. 研究の背景

がん化学療法では、治療薬そのものによる抗腫瘍効果に加え、患者の生活の質（Quality of Life: QOL）を損なう多様な副作用が問題となります。特に、末梢神経障害や口内炎、手足症候群などの「主観的で慢性的な苦痛を伴う副作用」は、患者の生活や就労に大きな影響を与える一方で、診療記録に明確な構造化データ（臨床検査値や診断コードなど）として残りにくいという課題があります。これらの症状を電子カルテなどの診療記録から抽出するためには、従来は患者のカルテを閲覧し、手作業でこれらの症状に関連する記録を見つけてくる必要がありました。

一方、既に臨床で広く用いられている医薬品の中には、抗がん薬に伴う副作用を軽減しうる「予防薬」の候補が多数存在し、細胞実験や動物実験など前臨床段階でその可能性が報告されているものもあります。しかし、これらの候補薬が「実際の患者」でどの程度副作用を抑制しうるのかを検証するには、大規模な観察研究や多数の患者を対象とした前向き介入試験が必要であり、時間的・費用的な制約が大きく、基礎から臨床への橋渡しが進まない現状がありました（Bench-to-Bedside Gap）。

研究グループはこれまでに、東京大学医学部附属病院の電子カルテ自由記載から自動的に副作用を抽出する NLP 基盤を開発してきました（Kawazoe Y et al. NPJ Digit Med. 2024）。併せて、本手法が実際の臨床現象を適切に捉えられるかを確認するため、すでに臨床研究で効果が示されている既承認薬を対

象に、パイプラインの妥当性を検証する研究を進めてきました。JCO Clinical Cancer Informatics 誌 (Tsuchiya M et al. 2025) では、セレコキシブがカペシタビン誘発手足症候群を予防するという既知の臨床知見を、本パイプラインを用いて再現できることを明らかにしました。また Scientific Reports 誌 (Kawazoe Y et al. 2025) では、アンスラサイクリン誘発性心毒性に対して心保護効果が報告されている ACE (アンジオテンシン変換酵素) 阻害薬/ARB、 β 遮断薬、スタチンなどについて、電子カルテデータに基づく解析により、これらの既知の臨床効果を再現可能であることを示しました。加えて、本研究で用いた解析枠組みの一部 (Tsuchiya M et al. 2025) は、医療情報学分野の国際会議 MEDINFO 2025 において Best Paper Award を受賞しており、国際的にも高く評価されています。

このような背景のもと、本研究では、日本語電子カルテの自由記載情報を NLP で解析することにより、がん化学療法に伴う副作用の発症パターンと、既承認薬の副作用抑制作用を同時に評価できるかどうかを検証しました。

関連論文 (*corresponding author)

- 1) Kawazoe Y*, Shimamoto K, Seki T, Tsuchiya M, Shinohara E, Yada S, Wakamiya S, Imai S, Hori S, Aramaki E. Post-marketing surveillance of anticancer drugs using natural language processing of electronic medical records. NPJ Digit Med. 2024 Nov 9;7(1):315. doi: 10.1038/s41746-024-01323-1.
- 2) Tsuchiya M, Kawazoe Y, Shimamoto K, Seki T, Imai S, Kizaki H, Shinohara E, Yada S, Wakamiya S, Aramaki E, Hori S*. Elucidating Celecoxib's Preventive Effect in Capecitabine-Induced Hand-Foot Syndrome Using Medical Natural Language Processing. JCO Clin Cancer Inform. 2025 Aug;9:e2500096. doi: 10.1200/CCI-25-00096. Epub 2025 Aug 12.
- 3) Kawazoe Y*, Tsuchiya M, Shimamoto K, Seki T, Shinohara E, Yada S, Wakamiya S, Imai S, Aramaki E, Hori S. Natural language processing of electronic medical records identifies cardioprotective agents for anthracycline induced cardiotoxicity. Sci Rep. 2025 Feb 24;15(1):6678. doi: 10.1038/s41598-025-91187-6.
- 4) Tsuchiya M, Shimamoto K, Kawazoe Y, Shinohara E, Yada S, Wakamiya S, Imai S, Kizaki H, Hori S, Aramaki E*. Natural Language Processing-Based Approach to Detect Common Adverse Events of Anticancer Agents from Unstructured Clinical Notes: A Time-to-Event Analysis. Stud Health Technol Inform. 2025 Aug 7;329:703-707. doi: 10.3233/SHTI250931.

3. 研究の内容・結果

本研究では、東京大学医学部附属病院の受診歴のある 56,326 名のがん患者を対象に、電子カルテの自由記載に含まれる診療録や看護記録を解析しました。日本語医療テキストに特化した Transformer ベースの NLP モデルを用いて、フルオロピリミジン系薬に関連する口内炎や手足症候群、白金製剤に関連する末梢神経障害などの副作用表現を自動抽出し、薬の投与日からの時間とあわせて時系列化しました。なお、データの定義を統一するため、副作用の発症日は電子カルテに記録された日付を採用しました。

次に、電子カルテの処方情報と連結することで、ARB やラメルテオンなど、これまで前臨床研究で副作用予防効果が示唆されてきた薬剤の投与状況を把握しました。患者背景や併用薬の違いによるバイアスを考慮するため、傾向スコアマッチングにより候補薬の使用群と非使用群を統計的に均衡化した上で、副作用の発症リスクを Cox 比例ハザードモデルにより評価しました。

その結果、フルオロピリミジン系薬による口内炎について、ARB 使用群では非使用群と比較して発症

リスクが有意に低下しており、ハザード比 0.58 と推定されました (図 2A)。また、白金製剤による末梢神経障害については、ラメルテオン使用群でハザード比 0.60 と抑制的な関連が示されました (図 2B)。一方、ネガティブコントロール (注 2) として設定した緩下剤では、口内炎および末梢神経障害のいずれにおいても副作用抑制作用は示されず (図 2C、2D)、本解析手法の妥当性を支持する結果となりました。

これらの所見は、既報の基礎研究で報告されている抗炎症作用や神経保護作用と矛盾しないものであり、NLP を用いた臨床テキスト解析が、がん化学療法における支持療法開発を加速しうることを示しています。

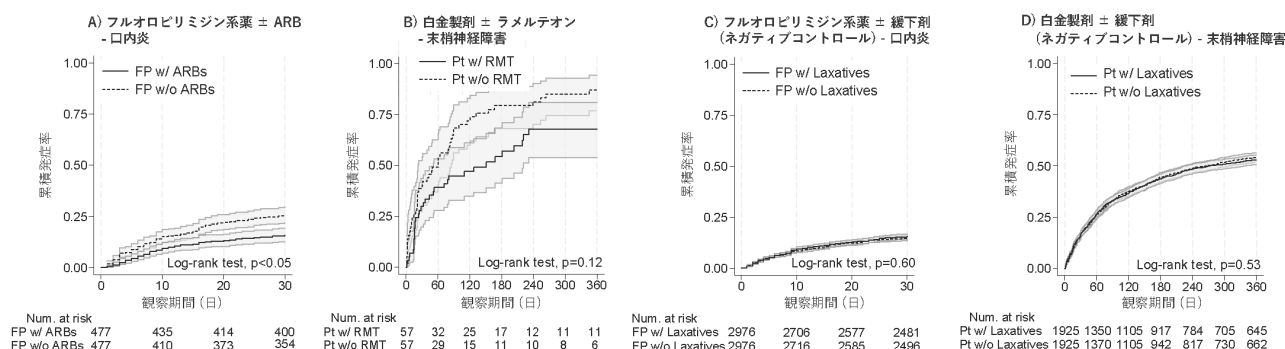


図 2. 候補薬の使用有無による副作用（口内炎・末梢神経障害）の累積発症率の比較

本図は、候補薬（ARB、ラメルテオン（RMT）、下剤）の使用群と非使用群における副作用（口内炎、末梢神経障害）の累積発症率を、傾向スコアマッチング後のデータに基づいて比較したものです。各曲線は観察期間内の累積発症率を示し、Log-rank 検定により群間差を評価しています。

(A) フルオロピリミジン系薬（FP）投与患者における ARB 使用有無での口内炎発症率の比較

(B) 白金製剤（Pt）投与患者における RMT 使用有無での末梢神経障害発症率の比較

(C) FP 投与患者における緩下剤（Laxatives）使用有無での口内炎発症率の比較（ネガティブコントロール）

(D) Pt 投与患者における緩下剤（Laxatives）使用有無での末梢神経障害発症率の比較（ネガティブコントロール）

NLP で抽出された副作用情報の信頼性を確認し、解析結果の妥当性を担保するため、本研究ではマッチング後の全症例を対象に人手による評価を行いました。がん化学療法に精通したがん専門薬剤師と臨床研修後の薬学生、さらにがん治療に携わる医師が電子カルテ記述を精査し、NLP によって抽出された副作用イベントが文脈に沿って適切に認識されているかを確認しました。その結果、フルオロピリミジン系薬による口内炎および白金製剤による末梢神経障害の抽出は手作業での判定と高い一致度を示し、抽出性能を示す F1 スコアは 0.87～0.89 と良好でした。また、誤りが生じたケースについても、副作用の見落としや、「他の治療や基礎疾患による症状を誤って抽出する」「否定的文脈を肯定的に解釈する」などの典型的なパターンに分類され、これらの誤りがハザード比などの解析結果に与える影響は限定的であることが確認されました。

以上より、本研究で用いた NLP 基盤は、臨床テキストを用いた大規模解析に十分適用可能であることが示されました。本研究の特徴は、追加の検査や介入を行うことなく、既に日常診療の中で蓄積された電子カルテ自由記載を活用することで、大規模な患者集団における副作用と医薬品による副作用抑制作用の関係を評価できる点にあります。これにより、患者に新たな負担をかけることなく、支持療法候補薬の探索と臨床検証を並行して進めることが可能になります。

4. 結論

本研究により、NLP を用いた臨床テキスト解析を活用することで、電子カルテの自由記載からがん化学療法に伴う副作用を大規模に定量化し、既承認薬の副作用抑制作用を臨床データに基づいて検証できることが示されました。本成果は、基礎研究と臨床研究の間に存在するギャップを埋める新たな手法として、がん支持療法研究の加速に貢献すると期待されます。

また、本研究は、医療機関に蓄積された臨床テキストの利活用価値を明確に示すものであり、医療 AI の社会実装や、患者の QOL 向上をめざした個別化支持療法の開発につながる可能性があります。今後は、多施設共同研究による検証や、前向き介入試験へと展開することで、実臨床における副作用予防戦略の確立が期待されます。

5. 特記事項

本研究は主に国立研究開発法人科学技術振興機構（JST）戦略的創造研究推進事業 CREST：研究領域「データ駆動・AI 駆動を中心としたデジタルトランスフォーメーションによる生命科学研究の革新」研究課題名「リアルワールドテキスト処理の深化によるデータ駆動型探索」（JPMJCR22N1）（研究代表者：荒牧英治、主たる共同研究者：河添悦昌、堀里子）、戦略的イノベーション創造プログラム（SIP）第3期「統合型ヘルスケアシステムの構築」：テーマ D-2「統合型の医学概念・知識連結データベースの構築及び医療文書の自動分析基盤の整備」（研究責任者：荒牧英治）の支援によって行われました。

6. 論文情報

〈タイトル〉 A scalable natural language processing framework for drug repurposing in chemotherapy-induced adverse events from clinical narrative records

〈著者名〉 Masami Tsuchiya[†], Mari Inoue[†], Yoshimasa Kawazoe, Kiminori Shimamoto, Tomohisa Seki, Shungo Imai, Hayato Kizaki, Emiko Shinohara, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki, Satoko Hori*（[†]co-first author, *corresponding author）

〈掲載誌〉 European Journal of Cancer

〈DOI〉 DOI: 10.1016/j.ejca.2025.116157

<用語説明>

（注1） 自然言語処理（Natural Language Processing：NLP）

人間が日常的に用いる文章や会話などの「自然言語」をコンピュータに処理させる技術の総称。医療分野では、電子カルテの自由記載から診断名や症状、検査値などの情報を抽出する用途で利用が進んでいる。

（注2） ネガティブコントロール

本来その効果が「ない」と考えられる比較対象を設定し、解析手法が誤った関連を示さないかを確認するための基準のこと。本研究では、口内炎や末梢神経障害の予防効果を持たないと考えられる緩下剤をネガティブコントロールとして用いることで、解析が特定薬剤の効果を過大評価していないかを検証している。

※ご取材の際には、事前に下記までご一報くださいますようお願い申し上げます。

※本リリースは文部科学記者会、科学記者会、各社科学部等に送信させていただいております。

<お問い合わせ先>

研究内容についてのお問い合わせ先：

慶應義塾大学薬学部 医薬品情報学講座

教授 堀 里子

TEL：03-5400-2650

E-mail：satokoh@keio.jp <https://keio-di.jp>

東京大学大学院医学系研究科 社会医学専攻 医療情報学分野

／東京大学医学部附属病院 企画情報運営部

教授 河添 悦昌

TEL：03-3815-5411（内線：35245）

E-mail：kawazoe-office@adm.h.u-tokyo.ac.jp

奈良先端科学技術大学院大学 先端科学技術研究科 情報科学領域

ソーシャル・コンピューティング研究室

教授 荒牧 英治

TEL：0743-72-5250

E-mail：aramaki@is.naist.jp

本発表資料についてのお問い合わせ先：

慶應義塾広報室（宮崎）

TEL：03-5427-1541 FAX：03-5441-7640

E-mail：m-pr@adst.keio.ac.jp