

電子カルテ由来データベースの可能性 ～米国転移性乳がんデータベースを例に～

中外製薬株式会社 安全性RWDS部 織田 哲朗, 松田 真一

1. 背景

- 薬剤疫学や医薬品安全性監視におけるリアルワールドデータ(RWD)の活用が進んでいるが、電子レセプトデータやDPCデータでは、画像・病理診断の結果や遺伝子バイオマーカーの検査結果が不明である。そのため、既存のコード体系では特定の難しいがん種やサブタイプが存在する。
- 電子カルテ情報に基づけば、フリーテキストや診断レポート情報を構造化する、あるいは、これらを参照した上で特定のがん種をデータベース化しうる。
- Flatiron Health社の提供する転移性乳がんデータベースの利用経験を元に、薬剤疫学や医薬品安全性監視における電子カルテの非構造化データの利用可能性について考察する。

2. Flatiron Health社データの概要

- Flatiron Health社のデータベースに含まれるデータは、OncoEMR（Flatiron Health社のがん特化型電子カルテ）を利用している施設で治療を受けたがん症例（**21種類**）の情報。約280以上のオンコロジークリニックと7のアカデミック施設、2021年6月時点で分析対象症例として**約30万例**が含まれる。
- 電子カルテから**直接入力される構造化データテーブル**と、**フリーテキスト欄から入力される非構造化データや画像診断・病理診断結果も照合した上で構築されたデータテーブル**がある。非構造化データの抽出と統合は、**機械学習を用いた自動カルテ情報抽出技術**と**abstractorsと呼ばれる人材による目視作業**の組み合わせによって行われている。
- 常時使えるデータベース(**Enhanced Data Mart, EDM**)のほかに、利用ごとに都度申請することで追加のデータ収集を依頼できるサービス(**Spotlight Project**)、Foundation Medicine社の遺伝子情報データベースと突合されたClinico-Genomics Database(**CGDB**)がある。

3. データベース(EDM)の特徴

- **原疾患特定がより正確:** ICD9/10コード情報だけでなく、画像・病理診断の結果やフリーテキスト情報を用いている。また、abstractorsによって不適切症例は除外されている。
- **死亡日がより正確:** Social Security Death Index(SSDI)およびobituary data（死亡記事）とのレコードリンケージによって死亡の複合エンドポイントの正確性が高い（18のがん種において感度および特異度が80%以上 [1]）。
- **がんのサブタイプが分かる:** ER, PR, HER2, PD-L1, BRCA, PIK3CAの検査日、検査結果、検体組織、等が分かる。※CGDBにはより多くの遺伝情報が格納されている。
- **治療ライン番号の付与:** 非構造化データも踏まえて1stラインと最後のラインを定義した上で、その間のライン番号も付与されている。
- **全施設の臨床検査値情報:** 施設や種類による制限はない。

3. データベース(EDM)の特徴

- **臨床検査値の正常範囲情報:** 各ラボで入力された正常範囲情報を用いて、施設依存の測定値のばらつきを考慮しうる。
- **病期やECOG PS:** 患者背景の調整に活用できる。PSと治療ライン番号との組み合わせによって、1stラインでPSが3の症例といった特定が可能。
- **転移部位:** 転移の部位、診断日などの情報がある。
- **EDM単体には初期乳がんの前治療/後治療情報がない:** 周術期治療に該当する前治療（転移前の薬剤治療、放射線治療、手術等）や後治療の情報が含まれていない。 ※Spotlight Projectを利用すれば取得可能。
- **重篤な有害事象の捕捉率が低い可能性:** オンコロジークリニックでの診療情報が中心であるため、重篤な有害事象の情報が含まれていない懸念。

4. データ活用の状況

- 2021年8月16日時点でFlatiron Health社のパブリケーションページに掲載されていた研究全515件（論文と発表）のうち150件を，閲覧可能な範囲の情報に基づき，研究目的や活用情報等の軸で集計した（表1）．
- OSやPFSを評価する研究が多く(52.0%)，安全性評価に活用された事例は多くない(3.3%)．事例としてはSanglier et al.(2019)[2]やSpillane et al.(2020)[3]がある．
- Flatiron Health社データの特徴でもある遺伝子変異情報や病期・PSといった情報が多くの研究で活用されており，それぞれの活用研究割合は54.7%と48.0%であった．

4. データ活用の状況

表1: Flatiron Health社のデータを利用した研究の集計

| 分類軸 | 分類 | 件数 | 割合 |
|-------|-----------------------------|----|-------|
| 目的 | 有効性評価(OS/PFS) | 78 | 52.0% |
| | 治療実態把握 ¹ | 55 | 36.7% |
| | 遺伝子検査実態把握 ² | 27 | 18.0% |
| | 安全性評価 | 5 | 3.3% |
| | 予後予測/リスクスコア構築 | 5 | 3.3% |
| | バリデーション/DB活用促進 ³ | 4 | 2.7% |
| | 診断名/日等の予測 ⁴ | 4 | 2.7% |
| | 臨床検査実態把握 ² | 3 | 2.0% |
| 活用情報 | 遺伝子変異情報 | 82 | 54.7% |
| | 病期やPS等 | 72 | 48.0% |
| | 臨床検査値 | 22 | 14.7% |
| がん種 | 転移性乳がん | 22 | 14.7% |
| | 乳がん | 10 | 6.7% |
| 使用データ | CGDB | 29 | 19.3% |

1:治療実態には、治療パスウェイを把握する研究やtime to treatment/discontinuationといった指標を求める研究が含まれる

2:臨床検査実態/遺伝子検査実態把握には、検査数を見る研究と検査結果を見る研究が含まれる

3:バリデーション/DB活用促進には、バリデーション研究やrwOSなどFlatironを活用するにあたり利用できる指標の整備を行った研究などが含まれる

4:診断名/日等の予測には、転移性有無と診断日の予測等、既存コーディングでは判定の難しい疾患を非構造化データを用いて予測する研究などが含まれる

※本文を確認できなかった文献も多数存在するため、実態とは乖離の可能性がある。

5. 簡易的な既報の再現

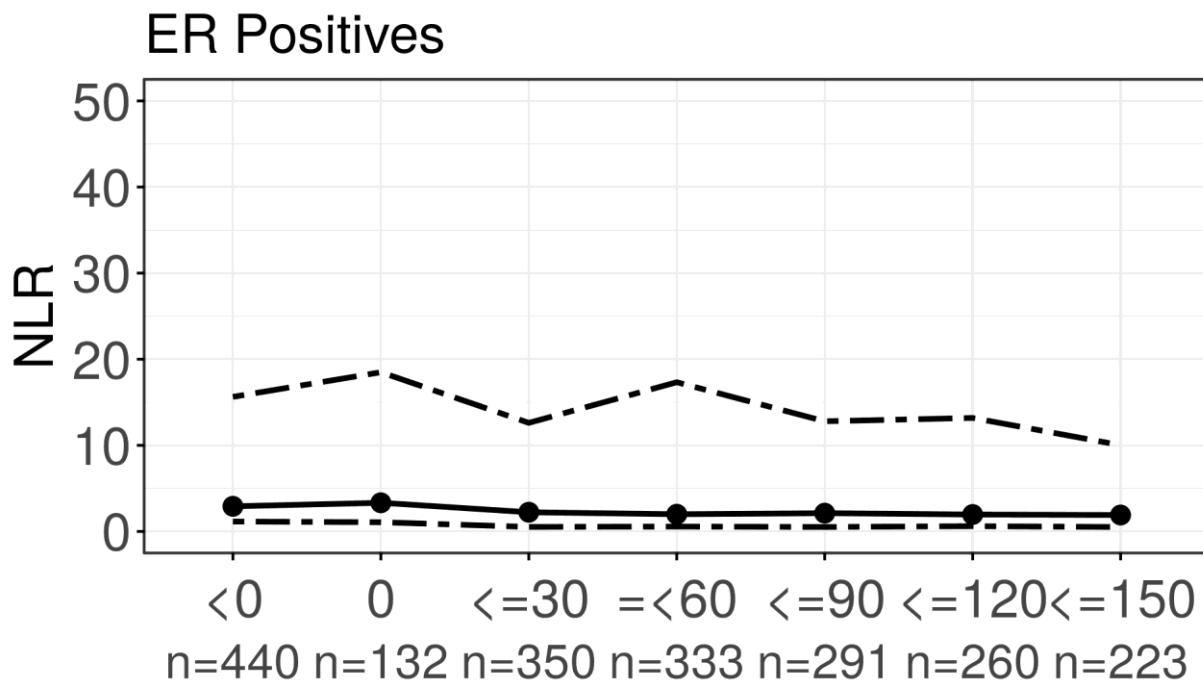
- 以上の特性を活用しながら，検査値やOverall Survival(OS)をアウトカム，遺伝子変異を曝露とした簡易的なデータ解析を行った.

5-1. Neutrophil-Lymphocyte Ratio(NLR)とERの関係

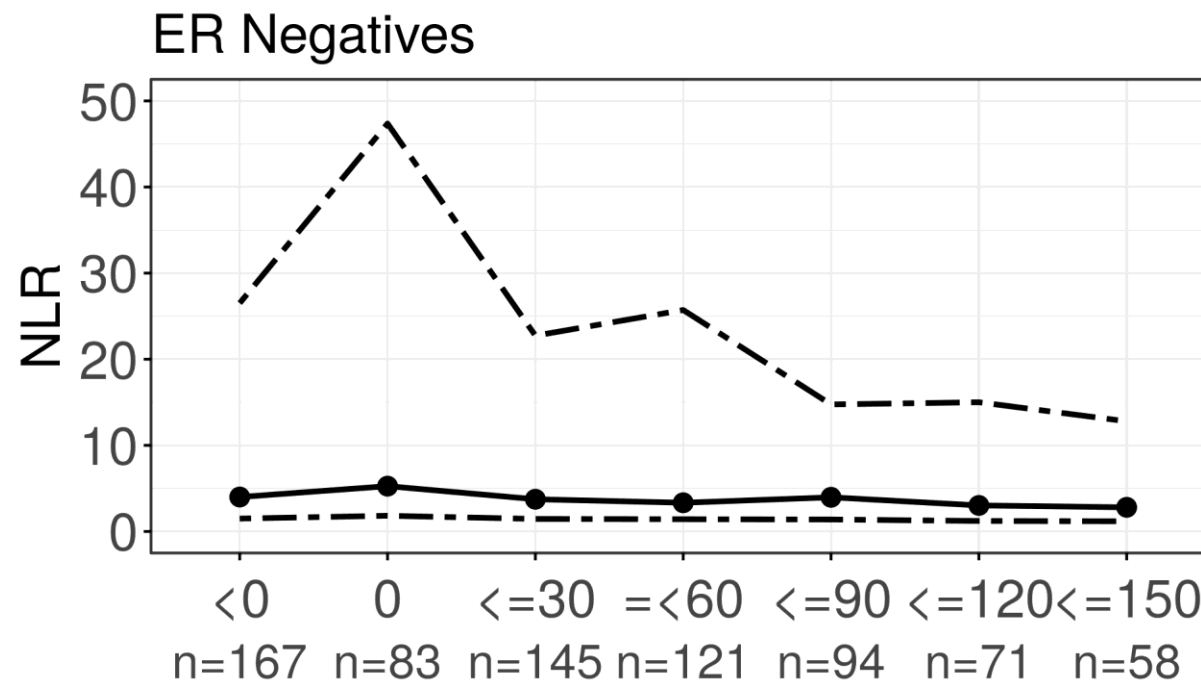


- NLRはがんの予後との関連が示唆されており，転移性乳がん症例においてもNLRが高いほどOSが短い[4,5]. また，ER陽性と比較してER陰性の方が予後が悪い傾向にある[5,6,7].
- 次の条件を満たす症例コホートにおいて、1stライン中のNLR推移をER変異別に確認した（図1） .
 - Estrogen receptor(ER)バイオマーカーの結果判明日が存在して，かつ，結果が陽性か陰性であり，1stライン開始日より前である，
 - 1stライン開始1か月前からNLRを算出できた，
 - 転移診断日から90日以内に1stラインが開始されている，等.
- 1stライン投与日をピークとして、投与開始30日前から投与後90日前までの間，97.5%点がER陰性群において大きく上昇しており，既報とも整合する結果が得られたと考える.

5-1. Neutrophil-Lymphocyte Ratio(NLR)とERの関係



Categories of days from 1st line start date



Categories of days from 1st line start date

図1: ERステータス別（左: 陽性, 右: 陰性）, 1stライン開始30日前から終了までにおける, 30日間ごとのNLR中央値（黒線）と95%パーセンタイル区間（破線）

5-2. ER+転移性乳がん症例におけるPgRとOSの関係

- 死亡日の正確性と遺伝子変異情報を活かした分析として、1stラインでアロマターゼ阻害薬単剤による治療を受けたER陽性転移性乳がん症例に絞り、Progesterone Receptor(PgR)陽性/陰性別の生存曲線を Kaplan-Meier 法でグラフ化した (図2) .
- 解析対象となるコホートの作成方法は5-1と同様である.
- PgR陽性の方がOSの中央値が大きく、Ravdin et al.(1992)[8]とも一致した.

5-2. ER+転移性乳がん症例におけるPgRとOSの関係

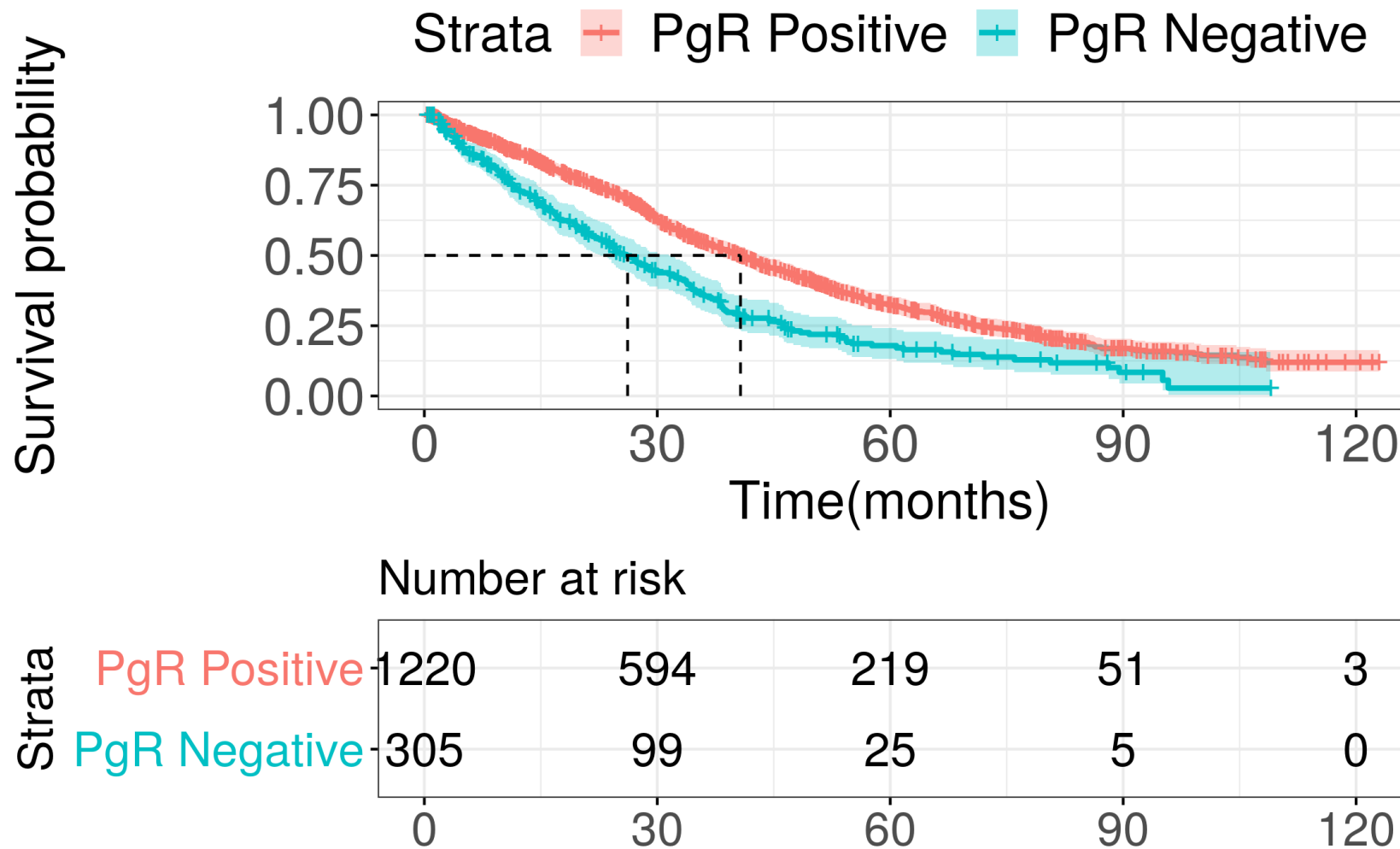


図2: 1stラインでアナストロゾールまたはレトロゾールの単剤治療を受けた
ER陽性転移性乳がん症例における, PgR陽性/陰性別 Kaplan-Meier 生存曲線

6. 結論・考察

- 電子カルテ由来の米国転移性乳がんデータベースを用いて、遺伝子サブタイプと臨床検査値やOSの関係を可視化し、概ね既報を再現できた。
- 非構造化データ（フリーテキスト及び画像診断や病理診断のレポート）や外部データとのリンケージを用いることで、より正確な疾患の特定や死亡日情報の整備がされていることが、この再現性に貢献していると考える。
- このようなデータベースを整備するためには、非構造化データの構造化、データの日視確認や外部データとの突合による正確なデータの構築、等が必要である。本邦でも、更なる電子カルテの活用が期待される。

7. 参考文献

1. Zhang, Q., et al., *Validation analysis of a composite real-world mortality endpoint for patients with cancer in the United States*. Health Serv Res, 2021.
2. Sanglier, T., et al. *Cardiac events in patients with HER2-positive metastatic breast cancer who have low left ventricular ejection fraction prior to initiating treatment with ado-trastuzumab emtansine: A retrospective cohort study using electronic health record data*. in *Cancer Research*. 2020. AMER ASSOC CANCER RESEARCH 615 CHESTNUT ST, 17TH FLOOR, PHILADELPHIA, PA ...
3. Spillane, S., et al., *Organ Dysfunction in Patients with Advanced Melanoma Treated with Immune Checkpoint Inhibitors*. *Oncologist*, 2020. 25(11): p. e1753-e1762.
4. de la Peña, F.A., et al., *Neutrophil-lymphocyte ratio (NLR) as a prognostic factor in metastatic breast cancer*. *Annals of Oncology*, 2017. 28: p. v94.
5. Ethier, J.-L., et al., *Prognostic role of neutrophil-to-lymphocyte ratio in breast cancer: a systematic review and meta-analysis*. *Breast Cancer Research*, 2017. 19(1): p. 1-13.
6. Arciero, C.A., et al., *ER+/HER2+ breast cancer has different metastatic patterns and better survival than ER-/HER2+ breast cancer*. *Clinical breast cancer*, 2019. 19(4): p. 236-245.
7. Parl, F.F., et al., *Prognostic significance of estrogen receptor status in breast cancer in relation to tumor stage, axillary node metastasis, and histopathologic grading*. *Cancer*, 1984. 54(10): p. 2237-2242.
8. Ravdin, P.M., et al., *Prognostic significance of progesterone receptor levels in estrogen receptor-positive patients with metastatic breast cancer treated with tamoxifen: results of a prospective Southwest Oncology Group study*. *Journal of clinical oncology*, 1992. 10(8): p. 1284-1291.

創造で、想像を超える。