

言語AIによるOTC医薬品の
安全性情報抽出とリスク数値化

国立大学法人 東京大学大学院 薬学系研究科 助教

みずの ただはや
水野 忠快

言語 AI による OTC 医薬品の安全性情報抽出とリスク数値化

東京大学薬学部 調査・研究 実施者氏名 水野 忠快

【要旨】

調査・研究は、セルフメディケーションの普及に伴い重要性が高まっている医薬品使用時の有害事象リスクについて、臨床試験データベースと言語 AI を用いて構造化し、定量的に評価するための基盤技術を構築することを目的とした。当初は OTC 医薬品全体を対象とした網羅的な解析を計画していたが、限られた研究期間・助成規模・計算資源の中で、まずは解析パイプラインの実現可能性を検証することを優先した。

具体的には、ClinicalTrials.gov の情報を整理した AACT データベースを対象に、臨床試験における介入群、対照群、有害事象発現数を抽出する処理系を構築した。その過程で、介入情報、試験アーム情報、有害事象情報が必ずしも一貫した ID で接続されていないことが明らかとなり、自然言語として記載された title、および description を用いた情報抽出・突合が必要であることが示された。そこで大規模言語モデルを用いて医薬品名、投与群、プラセボ群、有害事象情報を構造化する検討を行った。

本年度は、OTC 医薬品全体への展開に先立つケーススタディとして、AACT 上で臨床試験情報、および有害事象情報が取得可能であり、予備解析により感染・炎症関連有害事象の偏りが示唆された navarixin を対象とした。比較対象として、同じ呼吸器領域で臨床試験情報が得られる albuterol、および formoterol を参照薬として加えた。Nasopharyngitis, Rhinitis, Diverticulitis, Otitis media chronic, Gastroenteritis を対象に、各試験内および試験間の不確実性を考慮した nested bootstrap 解析を実施した。

その結果、navarixin では一部の感染・炎症関連有害事象についてログリスク比が正方向に推定された。一方、試験数や症例数が限られる項目では信頼区間が広く、本結果は確定的な安全性評価ではなく、臨床試験データに基づく探索的シグナルとして解釈すべきであると考えられた。

本年度の成果は、OTC 医薬品全体の網羅解析には至らなかったものの、臨床試験データベースから有害事象リスクを抽出・数値化するためのデータ処理、言語 AI による情報構造化、および統計的リスク推定の一連の基盤を構築した点に意義がある。今後は、本パイプラインを OTC 医薬品成分、および ATC 分類に基づく薬効群へ拡張し、薬剤師・消費者が理解しやすい安全性情報の可視化へ展開する。

1、調査・研究目的

本調査・研究の最終的な目的は、OTC 医薬品使用時の有害事象リスクを臨床試験データに基づいて定量的に評価し、セルフメディケーションにおける安全な医薬品選択を支援することである。セルフメディケーションでは、消費者が医師の直接的な管理下ではなく医薬品を選択・使用するため、有害事象リスクを専門家以外にも理解しやすい形式で提示することが重要である。

従来、有害事象リスクの把握には自発報告データベースが広く用いられてきた。しかし、自発報告データは報告バイアス、処方数・使用者数の不明確さ、重複報告などの課題を持つ。そこで本研究では、臨床試験データベースに記載された投与群ごとの有害事象発現数に着目した。臨床試験では、少なくとも試験ごとに投与群、被験者数、および有害事象発現数が整理されているため、発現割合やリスク比を推定できる可能性がある。

当初計画では、OTC 医薬品を広く対象とし、医薬品ごとの有害事象リスクを網羅的に数値化することを目指した。一方、実際に解析を進めたところ、臨床試験データベース内の介入情報、試験アーム情報、および有害事象情報の接続には構造上の制約があり、単純なテーブル結合のみでは医薬品ごとの有害事象リスクを自動抽出できないことが明らかとなった。また全 OTC 医薬品を対象とするには、言語 AI の利用料、手動検証、計算資源の面で大きなコストを要することが判明した。

そこで本年度は、研究全体の第一段階として、対象薬剤を限定し、臨床試験データベースから有害事象リスクを抽出・数値化するためのパイプライン構築とパイロット解析を実施した。対象薬剤としては、予備解析で既報のない有害事象が推定された navarixin を中心とし、比較対象として albuterol、並びに formoterol を加えた。これにより、OTC 医薬品全体への展開に先立ち、臨床試験データを用いた有害事象リスク定量化の実現可能性と課題を検証した。

2、調査・研究方法

2-1 臨床試験データベースの収集と統合

ClinicalTrials.gov [1] に登録された臨床試験情報を整理した AACT データベース [2] を対象とした。AACT に含まれる介入情報、試験アーム情報、および有害事象情報を取得し、各臨床試験における薬剤投与群、対照群、被験者数、および有害事象発現数を統合するためのデータ処理系を構築した。

具体的には、各臨床試験の試験 ID、介入名、アーム名、アームの説明文、有害事象名、有害事象発現数、および被験者数を抽出した。試験 ID や群 ID により直接対応付け可能な情報についてはテーブル結合により統合し、直接対応付けが困難な情報については、後述する言語 AI による情報抽出の対象とした。

本処理により、試験単位、群単位、有害事象単位の情報を整理し、医薬品ごとの有害事象リスク推定に利用可能な中間データセットを作成することを目指した。

2-2 言語 AI による非構造化テキストの構造化

AACT データベース中には、介入名、アーム名、title、description など、自然言語として記載された項目が含まれる。そこで本研究では、大規模言語モデルを用いて、これらの非構造化テキストから、薬剤名、投与群、対照群、プラセボ群、および群間の対応関係を抽出する処理を行った [3]。

抽出対象とした主な情報は、①対象薬剤が含まれるか、②当該アームが実薬群か対照群か、③プラセボ群に該当するか、④複数薬剤併用試験か、⑤有害事象発現数をどの群に対応付けるべきか、である。モデル出力は構造化形式で取得し、必要に応じて人手確認により妥当性を確認した。

当初は計算コストを抑えるため、オープンウェイトの大規模言語モデルを用いる方針とした。一方で、解析対象を限定した検証では、高性能な商用言語モデル API も併用し、非構造化テキストから有害事象リスク推定に必要な構造化データを作成した。

2-3 方法論検証用ケーススタディの設定

本研究では、OTC 医薬品全体への展開に先立ち、臨床試験データベースから薬剤群、有害事象名、発現数、および被験者数を抽出し、リスクを数値化する解析パイプラインの実現可能性を検証した。

ケーススタディとして、AACT 上で臨床試験情報および有害事象情報が取得可能であり、予備的なデータ確認において感染・炎症関連有害事象の偏りが示唆された navarixin を対象とした。navarixin 自体は OTC 医薬品ではないが、本研究で扱うデータ抽出、群情報の構造化、および risk ratio 推定という技術的課題は、OTC 医薬品の安全性情報解析にも共通する。そのため、本研究では navarixin を方法論検証用の対象薬剤として位置付けた。

AACT データベースから、title に navarixin を含む臨床試験を抽出し、対応する臨床試験 ID を取得した。次に、それらの試験に紐づく有害事象情報のうち、organ system が「Infections and infestations」に分類される有害事象語を抽出した。得られた有害事象語のうち、炎症との関連が想定される Nasopharyngitis, Rhinitis, Diverticulitis, Otitis media chronic, および Gastroenteritis を解析対象とした。

比較対象として、同じ呼吸器領域で臨床試験情報が取得可能であった albuterol, および formoterol を参照薬として加えた。

2-4. nested bootstrap によるリスク推定

各薬剤・有害事象の組み合わせについて、試験ごとの発現数と被験者数から risk ratio を算出した。臨床試験ごとの症例数やイベント数にはばらつきがあり、特に希少な有害事象では

推定値が不安定になりやすい。そのため、本研究では nested bootstrap を用いた。

具体的には、まず各臨床試験内で bootstrap を行い、試験内の risk ratio の不確実性を評価した。次に、得られた試験ごとの risk ratio を試験間で統合し、試験間のばらつきを考慮した信頼区間を算出した。bootstrap の反復回数は 5,000 回とした。解析は Python 3.10 を用いて実施した。

3、調査・研究成果

3-1 データ収集と抽出

まず ClinicalTrials.gov の情報を整理した AACT データベースから、臨床試験情報、介入情報、試験アーム情報、および有害事象情報を取得した。その結果、最終的に 77,455 件の臨床試験情報を解析対象候補として取得した。

取得したデータを確認したところ、臨床試験データベースを医薬品安全性評価に用いる上で、構造上の課題が明らかとなった。具体的には、試験アーム情報と有害事象情報は対応付け可能である一方、介入情報とアーム情報の対応関係は必ずしも明示的な ID で接続されていなかった。そのため単純なテーブル結合のみでは、特定の医薬品を投与された群における有害事象発現数を自動的に集計することは困難であった。

特に介入名やアーム名は自然言語として記載されており、同一薬剤であっても表記が揺れること、併用薬や投与量が複雑に記載されること、プラセボ群と実薬群の区別が title や description に依存することが課題であった。これは、OTC 医薬品全体を対象とした網羅的な自動解析を行う上で、非構造化テキストの解釈が不可欠であることを示す結果である。

この課題に対して、大規模言語モデルを用いた情報抽出を試みた。llama 3.1 などのオープンウェイトモデルを用いた検討では、薬剤名、併用薬、投与量、そしてプラセボ群の判定などにおいて、人による判断と一致しない例が認められた。そのため、本年度の助成規模と研究期間の中で OTC 医薬品全体を対象とした網羅解析を行うことは困難であると判断した。

一方、対象薬剤を限定し、高性能な商用言語モデル API を用いた場合には、非構造化テキストから薬剤名、投与群、プラセボ群、および有害事象情報を比較的妥当に抽出できることが確認された。これにより、対象薬剤を限定した条件では、臨床試験データベース中の非構造化情報を統計解析に利用可能な形式へ変換できる可能性が示された。

以上より、本年度は OTC 医薬品全体を対象とした網羅解析には至らなかったものの、臨床試験データベースから有害事象リスクを抽出・数値化するための基盤的な処理手順の構築に成功した。

3-2 統計解析・リスク評価

上記の構造化処理により作成したデータセットを用いて、navarixin を対象としたケーススタ

ディを実施した。AACT データベースから navarixin を含む臨床試験を抽出し、それらの試験に紐づく有害事象情報を整理したところ、感染・炎症関連の有害事象として Nasopharyngitis, Rhinitis, Diverticulitis, Otitis media chronic, および Gastroenteritis を解析対象として選定された。

比較対象として、同じ呼吸器領域で臨床試験情報が取得可能であった albuterol および formoterol を参照薬として加えた。これにより、navarixin 単独の結果ではなく、呼吸器領域の参照薬と比較しながら有害事象発現傾向を評価する解析系を構築した。

nested bootstrap により、各薬剤・有害事象の組み合わせについてログリスク比と信頼区間を推定した。その結果、navarixin では一部の感染・炎症関連有害事象についてログリスク比が正方向に推定された。特に Nasopharyngitis では、forest plot 上で正方向の点推定値が得られた (図 1)。

一方、信頼区間は広く、試験数や症例数が限られる項目では推定値の不確実性が大きいことも確認された (表 1)。また Diverticulitis や Otitis media chronic のようにイベント数が限られる有害事象では、個別試験や少数イベントの影響を受けやすいと考えられた。以上より、本解析結果は確定的な安全性評価ではなく、臨床試験データを横断的に用いた探索的シグナル検出として位置付けるのが妥当と考えられる。

本研究により、臨床試験データベースから薬剤ごとの有害事象発現リスクを抽出し、統計的に数値化し、forest plot として可視化する一連の処理手順を実装した。本パイプラインは、今後 OTC 医薬品成分リストに展開することで、セルフメディケーションに関連する医薬品の有害事象リスク評価へ応用可能である。

4、考察

本研究では、臨床試験データベースと言語 AI を組み合わせることで、医薬品ごとの有害事象リスクを定量化する枠組みの構築を試みた。当初計画では OTC 医薬品全体を対象とすることを想定していたが、実際に解析を進める過程で、臨床試験データベースの構造、言語 AI の利用コスト、および手動検証コストの面で複数の課題が存在することが明らかとなった。そのため本研究では、対象薬剤を限定し、臨床試験データから有害事象情報を抽出・構造化し、統計的にリスクを推定するための方法論検証を優先した。

本研究で明らかとなった重要な点は、臨床試験データベースにおいても、医薬品安全性評価に必要な情報が完全に構造化されているわけではないという点である。有害事象名、発現数、および被験者数は比較的整備されている一方、それらがどの薬剤投与群に対応するかを正確に判断するには、アーム名、介入名、そして description を統合的に解釈する必要がある。これは、OTC 医薬品全体を対象とした自動解析を行う上で避けて通れない課題である。

また言語 AI の利用可能性についても重要な知見が得られた。オープンウェイトモデルは費用

面で有利である一方、臨床試験記載のような専門的かつ複雑なテキストから、薬剤投与群と対照群を安定して抽出するには十分でない場面があった。一方、高性能な商用言語モデルを対象薬剤を限定して用いることで、有害事象リスク推定に必要な構造化データを生成できる可能性が示された。今後は、高精度モデルによる抽出結果を教師データとして蓄積し、軽量モデルやルールベース処理と組み合わせることで、費用対効果の高い解析系を構築する必要がある。

本年度のケーススタディでは、AACT上で臨床試験情報と有害事象情報が取得可能であり、予備解析により感染・炎症関連有害事象の偏りが示唆された navarixin を対象とした。navarixin 自体は OTC 医薬品ではないため、本解析結果を OTC 医薬品の安全性評価そのものの結論とは位置付けることはできない。しかしながら本研究の中心課題である臨床試験データベースから薬剤群、有害事象名、発現数、並びに被験者数を抽出し、リスクを数値化するという技術的課題は、OTC 医薬品にも共通する点を強調する。

一方、本研究には複数の限界がある。第一に、本研究では対象薬剤を限定したため、OTC 医薬品全体のリスク評価には至っていない。第二に、希少な有害事象では発現数が少なく、risk ratio の推定値が不安定になりやすいことが挙げられる。第三に、臨床試験の対象疾患、被験者背景、投与量、併用薬が試験ごとに異なるため、単純な risk ratio の比較には交絡が残る。最後に、本年度は探索的解析に留まったため、in vivo 試験による検証は実施しなかった点も課題である。特に動物試験を実施するには、臨床試験データから得られたシグナルが十分に頑健であるだけでなく、ヒト-動物モデルの外挿性を考慮する必要がある。本研究の結果は、これらの前段階に位置付けられる。

今後は、本研究で構築したパイプラインを実際の OTC 医薬品成分へ拡張する。具体的には、対象薬剤リストを OTC 医薬品成分に置き換え、ATC 分類や薬効分類に基づいて同系統薬間の有害事象リスクを比較する。また risk ratio だけでなく、ベイズ階層モデルや希少イベントに対応したメタ解析手法を導入し、試験数が少ない薬剤・有害事象の組み合わせでも安定した推定が可能な方法を検討する。さらに得られたリスク情報を薬剤師や消費者が理解しやすい形で可視化し、セルフメディケーション支援に資する情報提示へ発展させたい。

5、まとめ

本年度は、OTC 医薬品全体を対象とした網羅的な有害事象リスク評価には至らなかったものの、臨床試験データベースと言語 AI を組み合わせて、医薬品ごとの有害事象情報を抽出し、統計的にリスクを数値化するための基盤を構築した。特に AACT データベースの構造上の課題を明らかにし、非構造化テキストを含む臨床試験情報の構造化に言語 AI が有用である可能性を示した点は重要である。

また予備解析により感染・炎症関連有害事象の偏りが示唆された navarixin をケーススタディとして、薬剤・有害事象ごとの log risk ratio を推定し、フォレストプロットなどとして可視化する

一連の解析手順を実装した。navarixin 自体は OTC 医薬品ではないものの、本解析で確立したデータ抽出・構造化・リスク推定の手順は、OTC 医薬品の安全性情報解析にも展開可能である。

今後、本解析系を OTC 医薬品成分へ拡張することで、セルフメディケーションにおける医薬品選択の安全性向上に資する情報基盤の構築が期待される。

6、調査・研究発表（口頭又は誌上発表）

ありません。

7、引用文献

- [1] ClinicalTrials.gov: <https://clinicaltrials.gov/>
- [2] Tasneem A, Aberle L, Ananth H, Chakraborty S, Chiswell K, McCourt BJ, Pietrobon R. The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. PLoS One. 2012;7(3):e33677. doi: 10.1371/journal.pone.0033677. Epub 2012 Mar 16. PMID: 22438982; PMCID: PMC3306288.
- [3] Gisladottir U, Zietz M, Kivelson S, Tanaka Y, Sirdeshmukh G, Brown KL, Tatonetti NP. Leveraging Large Language Models in Extracting Drug Safety Information from Prescription Drug Labels. Drug Saf. 2026 Feb;49(2):177-193. doi: 10.1007/s40264-025-01594-x. Epub 2025 Sep 2. PMID: 40892374; PMCID: PMC12860809.

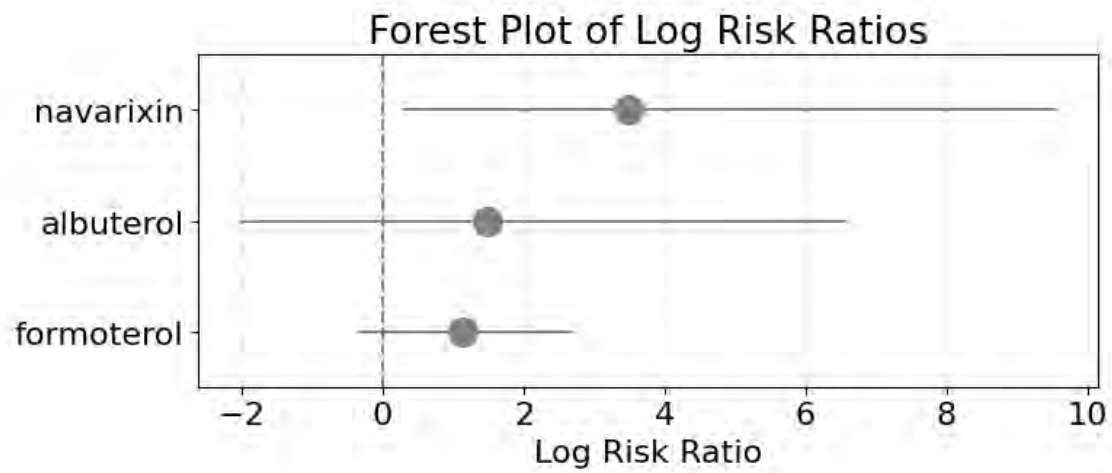


図 1. ログリスク比のフォレストプロット

表 1. 解析結果詳細

| 薬剤 | 有害事象 | log risk ratio | 95% CI |
|------------|-----------------|----------------|------------|
| navarixin | Nasopharyngitis | 3.48 | 0.27–9.57 |
| navarixin | Rhinitis | -0.25 | -1.14–0.62 |
| navarixin | Gastroenteritis | -0.10 | -0.10–0.10 |
| albuterol | Nasopharyngitis | 1.49 | -2.03–6.58 |
| formoterol | Nasopharyngitis | 1.14 | -0.34–2.71 |
| formoterol | Rhinitis | 3.11 | 2.30–3.91 |