

リアルワールドデータを用いた医薬品有害事象の
早期検出に関する研究

2022

山本 浩貴

目次

緒言.....	1
第1章 実臨床データベースに基づくゴールドスタンダードの作成.....	6
方法・結果.....	7
考察.....	41
第2章 レセプトの相関ルールマイニングによる副作用シグナルの早期検出...43	
方法.....	44
結果・考察.....	47
本研究の貢献と限界.....	67
総括.....	69
将来の展望.....	70
謝辞.....	72
発表論文目録.....	73
参考文献.....	74

緒言

医薬品使用に伴う好ましくない作用は医薬品有害事象（副作用）とされ、医薬品の使用には常に副作用のリスクが伴う。米国の調査において副作用は死因の第4位と推計され [1, 2]、また副作用の発見が遅れた場合、たった1種類の医薬品で約8億ドル以上の医療費につながる健康被害を引き起こすとも推計される（例：ロフェコキシブー心筋梗塞） [3]。以前の検討において、米国食品医薬品局（FDA）によって新規に承認された医薬品のうち、承認時に得られていない副作用のために上市された医薬品の約2%が市場から撤退し [4]、さらに約32%が安全性の懸念による警告の追加等のイベントを経験していることが示されている [5]。したがって、副作用を迅速かつ適時に検出できれば、患者の健康リスクを大幅に低減できる可能性がある。

一般に、臨床試験により医薬品の安全性の評価が行われるが、サンプルサイズが小さい、試験期間が短いなど多くの制約が挙げられ [6]、あらゆる副作用を捉えることは非常に困難である [7]。そのため、自発報告システム（SRS：Spontaneous Reporting System）を含めた市販後調査が最も重要であり、ファーマコビジランスでの必須のシステムである。主要なSRSの1つに米国FDAが収集・管理する有害事象自発報告システム（FAERS: FDA Adverse Event Reporting System）が挙げられ、2019年までに約1,100万人の患者が経験した様々な医薬品有害事象に関する情報が含まれる。SRSの利用は重篤な副作用を検出する最も効果的な方法であることが近年の研究で示されている [8-10]。

しかし、SRSには副作用の過小報告や医薬品を使用した母集団が不明などの限界がある [11, 12]。これまでの研究で、重篤な副作用のうち約6%しかSRSに報告されていないと推計されている [13, 14]。過小報告の背景として、医薬品による副作用であるかを評価する有用なツールはいくつか確立されているもの [15, 16]、患者が経験した症状の変化が医薬品に起因するものか、原疾患に起因するものかを区別することが困難であることが挙げられる [13]。これらの限界は、副作用を認識・検出するためのSRSによるデータ解析の質を低下させる可能性がある。以上の背景から、副作用を効果的に検出するためにSRSを補完できる方法が早急に求められている。

近年、大規模ヘルスケアデータを扱うための基盤が構築され、それらを活用した副作用研究への期待が高まっている。SRSとは異なり、電子カルテ情報（EMR: Electronic Medical Record）や診療報酬明細書（レセプト）には、副作用の疑いがあるか否かに関わらず、患者の症状の変化や処方情報が登録される。この特徴を活かし、EMRを活用して副作用シグナルを検出する取り組みなど多数報告されている [17-19]。しかし、EMRは患者が他施設に転院する等のイベントを経験

すると症状の追跡が難しく、より広い範囲の患者をカバーした解析が困難となる場合がある [20]。一方、レセプトデータは患者が転院した場合でも症状の変化や処方薬の情報を追跡することが可能なため、他の臨床データと比較して情報の損失が比較的少ないと考えられる。情報は月単位で纏められるデメリットはあるものの、数ヶ月後には匿名データとして利活用可能といった特徴も挙げられる。

Sequence Symmetry Analysis (SSA) はレセプトデータによる副作用シグナルの検出に最も活用される解析手法の 1 種である [21, 22]。本手法を用いた副作用研究は多数報告されているが、既報の大部分は長期データ（例：17 年間の蓄積データ）を使用し、また特定の医薬品クラス（糖尿病治療薬や睡眠薬など）とそのアウトカムに関する仮説を検討したもので、その有用性は限定的である [23–25]。2018 年には処方箋データベースを用いて、特定の医薬品ではなく数種の医薬品とそれに関係する副作用を評価するための解析研究が報告されているが [26–29]、用いるデータは長期のデータに基づいており、早期検出の視点では検討されていない。また、処方箋データベースには患者の症状が登録されていないため、治療薬が存在しない副作用を解析対象とすることは困難であった。

以上の背景から本研究の目的は、早期に、広範な副作用シグナルの検出を可能とする手法をレセプトデータで見出すことである。レセプトに登録される病名を有害事象の発生と考えることで、広範な副作用シグナルを「能動的」に検出する、この種の研究としては初めてとなる計算科学的な検討である。本研究では、第 1 章で広範な医薬品とその副作用を対象に、データマイニングの性能を評価するためのゴールドスタンダードを構築した。その上で第 2 章では大規模レセプトデータを用いて広範な副作用シグナルを早期に検出する方法を見出した。

これらの研究成果について、以下に論述する。

本文中および図中で使用した略語は以下の通りである。

ADR:	Adverse Drug Reaction
ARM:	Association Rule Mining
ATC:	Anatomical Therapeutic Chemical
AUC:	Area Under the Curve
EMR:	Electronic Medical Record
FAERS:	FDA Adverse Event Reporting System
FDA:	Food and Drug Administration
FDR:	False Discovery Rate
ICD10:	International Classification of Disease, 10th Revision
JADER:	Japanese Adverse Drug–Event Report
MedDRA:	Medical Dictionary for Regulatory Activities
PMDA:	Pharmaceuticals and Medical Devices Agency
PR:	Precision Recall
PT:	Preferred Terms
ROC:	Receiver Operating Characteristic
ROR:	Reporting Odds Ratio
SMQ:	Standardized MedDRA Queries
SNOMED-CT:	Systematized Nomenclature of Medicine-Clinical Terms
SRS :	Spontaneous Reporting System
SSA:	Sequence Symmetry Analysis
UMLS:	Unified Medical Language System

本研究で用いたデータベースは以下の通りである。

1. FDA Adverse Event Reporting System (FAERS) database

本データベースは第 1 章で活用した。FAERS は、米国食品医薬品局 (FDA: Food and Drug Administration) が管理する世界最大の有害事象自発報告データベースである。2004 年から 2019 年までの世界各国から報告された有害事象は、FDA のウェブサイト (<https://www.fda.gov/drugs/drug-approvals-and-databases/fda-adverse-event-reporting-system-faers>) から入手した。

既報を参考に重複報告を排除し [30]、残りの 11,438,031 件の報告を解析した。FAERS に登録されている医薬品名は任意記載形式であるため、ライフサイエンス辞書シソーラスを駆使したテキストマイニングにより医薬品名を一般名に統一した。医薬品有害事象は、Medical Dictionary for Regulatory Activities (MedDRA, <http://www.meddra.org/>; version 23.0) が推奨する用語集に従ってコード化された。FAERS の解析には 226 種類の用語から成る標準化された Standardized MedDRA Queries (SMQ) を使用した。

2. Japanese Adverse Drug–Event Report (JADER) database

本データベースは第 1 章で活用した。JADER は、医薬品医療機器総合機構 (PMDA: Pharmaceuticals and Medical Devices Agency) が管理する国内での有害事象自発報告を集積したデータベースである。2004 年から 2019 年までの自発報告は PMDA のウェブサイト (www.pmda.go.jp) から入手した。FAERS と異なり、各医薬品の初回投与日や各有害事象の発症日など詳細な情報に富む 611,336 件の有害事象報告が収録されている。JADER の解析においても 226 種類の用語から成る SMQ を使用した。

投与日や有害事象発症日のレコードには様々な桁数の数値が入力されていたため (例: 2011, 201102, 20110505, 20110506099, NULL など)、本検討では次の前処理を行った。日付の桁数が 8 桁のレコードのみを抽出し、さらに「有害事象の発現日」の日付が「投与開始日」の日付より後の日付となるレコードのみを抽出し、解析を行った。

3. JMDC insurance claims data (JMDC Claims)

本データベースは第 2 章で活用した。JMDC Claims は、7,438,470 人の従業員とその扶養家族の毎月の医療診断と処方箋の請求が含まれる国内レセプトデータベースである。日本の国民健康保険制度と従業員層の特徴のため、患者は 65

歳以下が中心で、75 歳以上の患者は含まれていない。2005 年から 2019 年までのレセプトデータは、株式会社 JMDC（東京、日本）から購入した。このデータセットにおける全ての診断は国際疾病分類第 10 版（ICD10: International Classification of Disease, 10th Revision）コードを用いて符号化され、また全ての医薬品は解剖治療化学分類（ATC: Anatomical Therapeutic Chemical）コードに対応付けられている。

本検討において ICD10 コードの「O00–O99: 妊娠・分娩及び産褥関連」、「Q00–Q99: 先天奇形・変形及び染色体異常関連」、「V01–Y98: 傷病及び死亡の外因関連」「Z00–Z99: 健康状態に影響を及ぼす要因及び保健サービスの利用関連」は医薬品に起因した症状（副作用）とは考えにくいと判断し、これらのレコードは解析から除外した。また医薬品については外用剤、輸液・灌流液、診断補助薬および漢方薬（ATC コード : D, K, R, T, V）に関するレコードは解析から除外した。

本研究で扱った解析コードは以下に共有した。

<https://github.com/HirokiYamamoto0222/ARM-on-Claims-data>

第 1 章

実臨床データベースに基づくゴールドスタンダードの作成

近年、電子カルテやレセプトデータなどのヘルケアデータに対し、機械学習などの解析を当てはめた副作用研究が進んでいる。機械学習や AI は専門家の知識ベースを超える情報を与える可能性があるが、これらデータマイニングによる副作用シグナル検出法の性能を評価するためには、適切な参照基準（ゴールドスタンダード）が必要となる。これまでにランダム化比較試験や添付文書情報、文献情報を活用することで、数種のゴールドスタンダードが作成され [26, 31–34]、またこれらゴールドスタンダードを用いて副作用検出の性能を評価した研究も報告されている [28, 35–37]。Ryan らは「心筋梗塞」、「腎障害」、「肝障害」、「消化管出血」など、ファーマコビジランス活動に不可欠な 4 つの有害事象を対象とした 165 組の陽性対照と 234 組の陰性対照からなるゴールドスタンダードを作成した [32]。Harpaz らは Med Watch を活用することで、2013 年に米国で発令された安全性情報をもとに 62 組の陽性対照と 75 組の陰性対照で構成されるゴールドスタンダードを作成した [33]。このゴールドスタンダードの特徴として医薬品が承認されてから添付文書が改訂されるまでの期間の情報が含まれる。しかし、ファーマコビジランスを加速させるためには、広範囲の医薬品と関連する副作用にわたるゴールドスタンダードを確立することが極めて重要である。また、実用性の視点でデータマイニング手法が有用であるかを評価するためには、副作用の発生タイミングに関する情報も必要と考えるが、これらを満たすゴールドスタンダードは構築されていない。そこで、近年は大規模臨床データを扱うための基盤が構築されていることから、実臨床データを用いることで上記の限界点を補えると考えた。

本章では、世界最大の自発報告データベース FAERS と、医薬品投与日と有害事象発現日の詳細な情報に富む国内自発報告データベース JADER を用いることで、幅広い副作用を対象にし、また副作用の発症時期情報を含む新たなゴールドスタンダードの構築を目的に検討を行った。

方法・結果

1. ゴールドスタンダード構築のフローチャート

初めに FAERS や JADER の前処理を行った。FAERS は世界中の有害事象報告を集積しているため、様々な医薬品名がデータベースに登録される。そのため、当研究室の薬物名辞書を用いて名寄せを行い、746,345 種類の医薬品の表記名を 3,060 種類ほどの成分名へ統一した。次に有害事象名の基本語は 20,000 種を超えるため、MedDRA が推奨する用語集に従って有害事象の種類を 226 種の SMQ にまとめた。その後、FAERS と JADER の両データベースに登録される約 3,000 種の医薬品と 226 種類の SMQ のペアについて、不均衡分析 (Disproportionality analysis) と二項検定 (Binomial test) の 2 種類の統計解析を網羅的に行い、**図 1-1** に示すフローチャートに従って最終的に 92 種類の陽性対照と 88 種類の陰性対照からなるゴールドスタンダードを構築した。

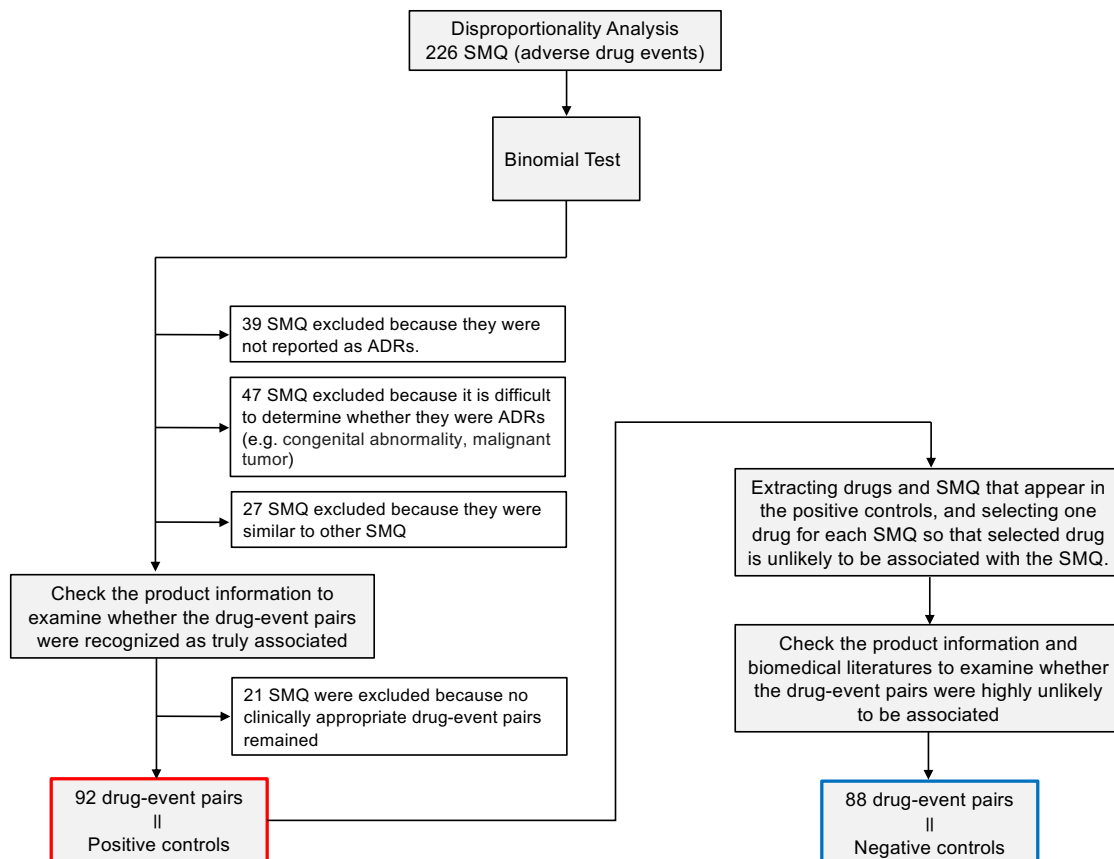


図 1-1 FAERS 及び JADER の網羅的解析による 92 種類の陽性対照と 88 種類の陰性対照構築の全体像

2. 不均衡分析

不均衡分析は、各有害事象について有意に報告される医薬品を特定するために実施した。FAERS や JADER に登録される医薬品には、副作用の寄与として被疑薬、併用薬、相互作用などの情報が登録される。本解析では「第一被疑薬」または「第二被疑薬」とされる報告のみを対象に解析を行った。各医薬品と各有害事象（226 種類の SMQ）との関連について、報告オッズ比（ROR）と有意性の指標となる Z score を網羅的に算出した [8]。具体的には、FAERS 及び JADER に報告される症例を以下の 4 つのグループに分け解析を行った。

- a: 対象の医薬品を投与され、かつ特定の有害事象を発症した症例
- b: 対象の医薬品を投与されたが特定の有害事象を発症しなかった症例
- c: 対象の医薬品を投与されず、特定の有害事象を発症した症例
- d: 対象の医薬品を投与されず、特定の有害事象を発症しなかった症例

ROR と Z score は、以下の式で算出した。

$$\text{ROR} = \frac{a/b}{c/d}$$

$$\text{Z score} = \frac{\log(\text{ROR})}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

ここで a、b、c、d は各群の症例数を示す。4 つのグループのうち 1 つでも 0 となる症例数がある場合、ROR 算出が困難となる。そのため本検討では、Haldane–Anscombe 1/2 correction による補正（各症例数に 0.5 を加算）を行った。「医薬品－有害事象」の陽性対照と陰性対照、さらに副作用の発症時期情報は以下の基準で作成した。

3. 陽性対照

陽性対照は「医薬品－有害事象」の組合せのうち、「医薬品と有害事象間に因果関係が認められる」と考え得る組合せと定義した。ここで、抗がん剤を具体例に挙げた場合、悪心・嘔吐、倦怠感、脱毛、白血球減少症などその他の様々な有害事象が報告されると予想される。本検討では、その中でも他の有害事象よりも有意に報告されているものを特定したいと考えた。そこで、各医薬品について偶然報告よりも高い頻度で報告された有害事象を特定するため、二項検定 (Binomial test) による統計解析を実施した [38]。具体的には、対象となる医薬品について、報告された全ての有害事象発生の p -value を計算し、Benjamini-Hochberg 法による False Discovery Rate (FDR) 補正を行った (R studio 'binom.test'関数を使用)。FDR 補正後の p -value が 0.01 未満の有害事象については、その医薬品との関連が有意に高いと推定される。陽性対照の作成は、二項検定の結果と不均衡度分析で有意 ($ROR > 1$ および $Z \text{ score} > 1.96$) と判定された「医薬品－有害事象」の組合せの中から作成した。

226 種類の SMQ のうち、39 種類は有害事象としての報告数が算出されなかったため、以降の検討から除外された (表 1-1)。一部の SMQ は階層構造になっており、複数の下位の SMQ の組合せからより包括的な上位の SMQ を形成している [39]。今回は最も下位となる SMQ のみを対象にしたため、上位 SMQ は報告数が算出されなかった。また 47 種類の SMQ は有害事象であるかの判定が困難であると判断し、以降の検討では除外した (表 1-2: 例: 先天性異常など)。さらに 27 種類の SMQ は他の SMQ に類似していると判断し以降の検討では除外した (表 1-3)。残った 113 種類の各 SMQ について、FAERS 及び JADER 解析で強い $Z \text{ score}$ を示した組合せから、以下の基準で 1 つの医薬品を選択した。

1. 各 SMQ について FAERS と JADER で $Z \text{ score}$ 値が上位 30 位までの医薬品を抽出
2. そのうち上位 3 組までについて製品情報を確認し、医薬品と有害事象の組合せが妥当な関係であるかを検証。妥当と考えられる医薬品が FAERS/JADER のデータベース間でサポートされていれば、その医薬品を選択
3. 2.で基準を満たす医薬品が存在しなかった場合は、下順位を確認

表 1-4 に陽性対照作成の具体例を 2 つ記した。

表 1-1 有害事象としての報告数が算出されなかった 39 種類の SMQ

SMQ 名	
肝障害	薬物乱用、依存および離脱
薬剤に関連する肝障害-包括的検索	消化管の穿孔、潰瘍、出血あるいは閉塞
薬剤に関連する肝障害-重症事象のみ	口腔咽頭障害
悪性および詳細不明の肝新生物	胆道系障害
造血障害による血球減少症	機能性、炎症性および胆石が関連する胆道系障害
うつ病および自殺/自傷	胆道新生物
出血	悪性および詳細不明の胆道新生物
虚血性心疾患	消化管の非特異的炎症および機能障害
不整脈	悪性および詳細不明の乳房新生物
不整脈用語（徐脈性不整脈および頻脈性不整脈を含む）	悪性および詳細不明の卵巣新生物
徐脈性不整脈（伝導障害および洞結節機能障害を含む）	悪性および詳細不明の前立腺新生物
頻脈性不整脈（上室性および心室性頻脈性不整脈を含む）	悪性および詳細不明の子宮/卵管新生物
中枢神経系血管障害	甲状腺機能障害
中枢神経系出血および脳血管性疾患	聴覚および前庭障害
ショック	悪性および詳細不明の皮膚新生物
塞栓および血栓	妊娠と新生児のトピック
前癌状態	乳汁分泌関連トピック（新生児の母乳を介した曝露を含む）
悪性疾患	悪性腫瘍
悪性または詳細不明の腫瘍	悪性度不明の腫瘍
錐体外路症候群	

表 1-2 有害事象であるかの判定が困難と判断した 47 種類の SMQ

SMQ 名	
肝臓の先天性、家族性、新生児および遺伝性障害	新生児障害
明確にアルコール関連と報告された肝障害	妊娠中絶および流産のリスク
妊娠関連の肝障害	正常妊娠の状態および転帰
効能／効果の欠如	悪性胆道腫瘍
自殺／自傷	悪性度不明の胆道腫瘍
先天性および新生児不整脈	悪性乳房腫瘍
ショック関連の循環あるいは心臓の状態（トルサード ド ポアントを除く）	悪性度不明の乳房腫瘍
血液量減少性ショック状態	悪性卵巣腫瘍
先天性、家族性および遺伝性障害	悪性度不明の卵巣腫瘍
血液の前癌状態	悪性前立腺腫瘍
胃腸の前癌状態	悪性度不明の前立腺腫瘍
生殖器の前癌状態	悪性皮膚腫瘍
皮膚の前癌状態	悪性度不明の皮膚腫瘍
悪性疾患関連状態	悪性子宮／卵管腫瘍
悪性疾患関連の治療と診断の手法	悪性度不明の子宮／卵管腫瘍
腫瘍マーカー	悪性肝臓腫瘍
口腔咽頭の新生物	悪性度不明の肝臓腫瘍
先天性胆道系障害	悪性リンパ腫
事故および損傷	投薬過誤
血管外漏出（注射、注入および埋込み部位）	血液学的悪性腫瘍
全身状態およびその他特定部位の前癌状態	非血液学的悪性腫瘍
妊娠、分娩合併症と危険因子（流産および死産を除く）	悪性度不明の血液学的腫瘍
母乳を介した新生児曝露	悪性度不明の非血液学的腫瘍
胎児障害	

表 1-3 類似と判断して除外した 27 種類の SMQ と残した SMQ

除外された SMQ 名	類似と判断して残した SMQ 名
肝臓関連臨床検査、徴候および症状	肝臓に起因する胆汁うっ滞および黄疸
肝不全、肝線維症、肝硬変およびその他の肝細胞障害	肝臓に起因する胆汁うっ滞および黄疸
肝臓に関連する凝固および出血障害	消化管の出血
出血関連用語（臨床検査用語を除く）	出血性中枢神経系血管障害
出血関連臨床検査用語	消化管の出血、出血性中枢神経系血管障害
不整脈に関連する臨床検査、徴候および症状	上室性頻脈性不整脈、心室性頻脈性不整脈
伝導障害	トルサード ド ポアント/Q T 延長
トルサード ド ポアント、ショック関連状態	トルサード ド ポアント/Q T 延長
中毒性-敗血症性ショック状態	敗血症
アナフィラキシー/アナフィラキシー様ショック状態	アナフィラキシー反応
低血糖性および神経性ショック症状	低血糖
動脈の塞栓および血栓	心筋梗塞
血管タイプ不明あるいは混合型の塞栓および血栓	虚血性中枢神経系血管障害
アカシジア	ジスキネジア、ジストニア
薬剤離脱	薬物乱用および依存
消化管の穿孔、潰瘍、出血あるいは閉塞についての非特異的所見/処置	消化管の出血、消化管の潰瘍、消化管の穿孔、消化管の閉塞
口腔咽頭のアレルギー状態	血管浮腫
歯肉障害	骨壊死
急性中枢性呼吸抑制	呼吸不全
感染性胆道系障害	胆嚢関連障害
胆道系に関連する臨床検査、徴候および症状	胆道障害
虚血性大腸炎	消化管の穿孔
非特異的不整脈用語	上室性頻脈性不整脈、心室性頻脈性不整脈
非特異的徐脈性不整脈用語	上室性頻脈性不整脈、心室性頻脈性不整脈
非特異的頻脈性不整脈用語	上室性頻脈性不整脈、心室性頻脈性不整脈
免疫処置後の全身痙攣発作	痙攣
慢性腎臓病	急性腎不全

表 1-4 陽性対照ペア選択の流れ

<SMQ: 末梢性ニューロパチー>

FAERS				JADER			
No	Drug name	Z score	Class	No	Drug name	Z score	Class
1	ボルテゾミブ	160.84	A,○	1	オキサリプラチン	37.90	A
2	シプロフロキサシン	124.55	A	2	インフルエンザワクチン	31.45	D
3	レボフロキサシン	122.81	A	3	ボルテゾミブ	29.28	A,○

→ボルテゾミブを選択

<SMQ: 肝感染>

FAERS				JADER			
No	Drug name	Z score	Class	No	Drug name	Z score	Class
1	ソホスブビル	195.96	B	1	リツキシマブ	36.02	A,○
2	リバビリン	157.82	B	2	プレドニゾロン	19.03	A
3	レジパスビル	115.68	B	3	ヨード化ケン油 エチルエステル	16.63	C
4	ラミブジン	93.49		4	シクロホスファミド	15.45	
5	シメプレビル	89.84		5	ラミブジン	15.01	
6	リツキシマブ	82.86	A,○	6	ビンクリスチン	14.47	

→リツキシマブを選択

各 SMQ の医薬品について、以下の区分の判定を行った

○: 両データベースで強いシグナルと考えられるペア

(「医薬品－有害事象」に因果関係が認められる、と考え得るペア)

A: 医薬品の製品情報に有害事象が記載されている

B: 医薬品の適応症に関連すると考えられる有害事象

C: 因果関係が認められるかどうかの判断が困難

D: 外用剤・吸入剤・ワクチン・国内未承認薬・詳細不明の医薬品

SMQ: 肝感染について、ソホスブビルやリバビリンは C 型肝炎ウイルス感染症の治療薬であり、肝感染の副作用報告は医薬品の効果不十分による影響が強いと考えられた。統計解析だけでは排除できない部分は個別に確認を行い、妥当と考え得る医薬品を選択した。

大部分は上記 3 つの選択基準により陽性対照となる組合せを作成したが、Z score によるシグナルの強さを基準とした選択では、同じ医薬品が複数の SMQ において選択される可能性があった。幅広い医薬品・ATC 分類を含めたゴールドスタンダードとするために、例外として 8 種類の SMQ（“無顆粒球症”、“喘息／気管支痙攣”、“脂質異常症”、“高血糖／糖尿病の発症”、“静脈の塞栓および血栓”、“胆嚢関連障害”、“高血圧”、“尿細管間質性疾患”）については、より強いシグナルを示す医薬品はあるものの、下位の順位となる医薬品を選択した。12 種類の SMQ（“肝臓に起因する胆汁うっ滞および黄疸”、“溶血性障害”、“造血障害による 2 種以上の血球減少症”、“心室性頻脈性不整脈”、“ジスキネジア”、“消化管の閉塞”、“胆道障害”、“消化管の非特異的機能障害”、“消化管の非特異的症候および処置”、“結膜障害”、“リポジストロフィー”、“骨粗鬆症／骨減少症”）については、JADER での Z score が 30 位までに含まれていないが、FAERS においては報告数が多くかつ Z score が高い医薬品を選択した。2 種類の SMQ（“上室性頻脈性不整脈”、“薬物乱用および依存”）については、FAERS での Z score が 30 位までに含まれていないが、JADER で Z score が高く、国内での使用例数が多いと考えられる医薬品（JMDC Claims で十分な症例数を持つ医薬品）を選択した。“味覚および嗅覚障害”の SMQ については、医薬品自体に苦味が少なく、国内での使用例数が多いと考えられる医薬品を選択した。“網膜障害”の SMQ については、Z score が高い医薬品はその他にあるものの、SMQ に関する副作用名が製品情報の警告欄に記載されていた医薬品を選択した。“その他の虚血性心疾患”の SMQ については、FAERS での Z score が 30 位までに含まれていないが、SMQ に関する副作用が頻度の高い重篤な副作用とされる医薬品を選択した。

最後に、不均衡分析と二項検定から抽出された医薬品のうち、臨床的に適切な医薬品が残らず「医薬品－有害事象」の陽性対照を作ることができなかった合計 21 種類の SMQ は除外された（表 1-5）。上記までの過程により、最終的に 78 種類の医薬品を用いた合計 92 種類の「医薬品－有害事象」の陽性対照が作成された。

表 1-5 陽性対照として妥当と考え得る「医薬品－有害事象」の組合せを作成できなかった 21 種類の SMQ

SMQ 名	
良性肝新生物（嚢胞およびポリープを含む）	前庭障害
抗コリン作動性症候群	涙器障害
後腹膜線維症	眼窩周囲および眼瞼障害
認知症	腎血管障害
口腔咽頭の感染	強膜障害
良性胆道新生物（嚢胞およびポリープを含む）	機能的乳汁分泌障害
ギラン・バレー症候群	生殖能障害
脱髄	筋緊張低下-反応性低下発作
角膜障害	関節炎
出血性あるいは虚血性と特定されない中枢神経系血管障害	骨髄異形成症候群
中枢神経系出血および脳血管性損傷に伴う状態	

4. 陰性対照

陰性対照は「医薬品と有害事象間の関連性が極めて低い」と考え得る組合せと定義した。本検討では陽性対照で用いた医薬品（78種類）とSMQ（92種類）でペア（78×92種類）を作り、以下の基準を満たす「医薬品－有害事象」から陰性対照を作成した。

1. 各SMQについて、JADERで「第一被疑薬」または「第二被疑薬」として1例も有害事象として報告されず、かつFAERS解析でROR < 1となる医薬品
2. SMQ毎に無作為に1医薬品を選択
3. 製品情報及びPubMedを用いた文献検索を行い「医薬品－有害事象」の組合せの妥当性を検証

3つ目の基準について、具体的には対象とする医薬品添付文書に有害事象名が記載されていないことを確認し、さらにPubMedの検索には「Drug-induced event」の用語を使用した。PubMed検索において、副作用に関する症例報告がない場合、または副作用を抑制する症例報告がある場合、その医薬品を選択した。陽性対照で用いた92種類のSMQのうち、アレルギーに関するSMQ（“好酸球増加および全身症状を伴う薬物反応症候群”、“好酸球性肺炎”、“喘息／気管支痙攣”、“血管浮腫”、“急性膵炎”、“アナフィラキシー反応”、“重症皮膚副作用”、“過敏症”）、消化器症状に関するSMQ（“非感染性下痢”、“消化管の非特異的機能障害”、“消化管の非特異的炎症”、“消化管の潰瘍”）、電解質異常に関するSMQ（“低カリウム血症”、“脱水”、“乳酸アシドーシス”）、“高血圧”、貧血に関するSMQ（“造血障害による赤血球減少症”）はどの医薬品の使用でも起こりうる副作用と判断し、「医薬品－有害事象」の陰性対照の作成は困難であると判断した。また大部分の医薬品は肝臓で代謝されるため、2種類のSMQ（“非感染性肝炎”と“肝臓に起因する胆汁うっ滞および黄疸”）について陰性対照を作ることは困難であると判断した。3種類のSMQ（“心室性頻脈性不整脈”、“洞結節機能障害”、“トルサードポアント／QT延長”）はどれも不整脈に関するものであり“上室性頻脈性不整脈”のSMQに含めて陰性対照を作成した。また“胆嚢関連障害”のSMQは“胆道障害”のSMQに含め、陰性対照を作成した。4種類のSMQ（“血行動態的浮腫、蓄水および体液過負荷”、“非感染性脳症／譫妄”、“心不全”、“消化管の非特異的症状および処置”）については、上記3つの基準を満たす適切な医薬品が残らなかったため、陰性対照の組合せの作成ができなかった。上記までの過程により計65種類のSMQについて陰性対照が作成された。

最後に陰性対照の数を陽性対照の数（92 個）に揃えるため、数種類の SMQ は追加でもう 1 つの妥当と考え得る医薬品を選択した。最終的に 65 種類の医薬品を用いた合計 88 種類の「医薬品－有害事象」の陰性対照が作成された。

5. 副作用の発現時期に関する情報

92 種類の陽性対照については、JADER に登録される医薬品投与日と有害事象発現日の差分による中央値から、実臨床における副作用発現日に関する情報を推定・追加した。

6. 有害事象名と ICD10 コードの対応付け

自発報告データにおける有害事象は MedDRA 標準検索式である SMQ に従ってコード化されている。一方、レセプトデータの疾患名は ICD10 という国際分類に従ってコード化されている。作成したゴールドスタンダードを他データベースで活用するには有害事象名を疾患名に紐付ける必要があるが、これらを機械的に対応させても 20–30%程しか紐付けることはできない [40–42]。そのため本検討では陽性対照で用いた SMQ について、各医薬品がどのような有害事象名で高頻度に報告されているかを特定し、レセプトでの類似する疾患名と手作業で紐付けることにより対応した。具体的な作業過程は表 1–6 に記した。SMQ は複数の MedDRA 優先用語（PT: Preferred Terms）をグループ化した有害事象カテゴリーである。初めに、各陽性対照の SMQ について、ペア薬剤の FAERS における PT を上位 3 つまで特定した。JMDC Claims には 22,925 の疾患名が含まれており、それらは 1,500 種類の ICD10 コードにリンクされている。そのため上位 3 種類の PT と JMDC Claims での疾患名が類似しているものをペアとすることで、SMQ と ICD10 を対応させた。

以上、大規模自発報告データベースの統計解析により合理的に選択したゴールドスタンダードを構築した（表 1–7, 表 1–8, 図 1–2, 表 1–9）。計 180 組（陽性対照 92 組と陰性対照 88 組）で構成され、広範な医薬品（78 種類）と関連する有害事象（92 種類）を含み、さらに実臨床における副作用発症時期の情報も含む。本ゴールドスタンダードを用いることで幅広い医薬品と副作用を対象に、データマイニング手法の性能を評価することを可能にした。

表 1-6 SMQ の ICD10 へのマッピング

FAERS			JMDC Claims data		
SMQ	pt	Counts	疾患名	ICD10 code	Counts
トルサード ド ポア ント/Q T 延長	心電図 QT 延長	658	薬物性 QT 延長症候群	I45	19,969
	心室性頻脈	30			
	心電図 QT 間隔異常	14			
横紋筋融解症/ ミオパチー	横紋筋融解症	4,794	薬剤誘発性横紋筋融解症	M62	80,102
	ミオパチー	1,060			
	血中ミオグロビン増加	97			
急性腎不全	急性腎障害	1,895	急性腎不全	N17	5,754
	腎不全	648			
	中毒性ネフロパシー	572			
心不全	右室不全	911	右室不全	I50	107,625
	心不全	875			
	うっ血性心不全	853			
肝臓に起因する胆汁 うっ滞および黄疸	黄疸	942	肝内閉塞性黄疸	K71	6,351
	胆汁うっ滞	811			
	薬物性肝障害	677			
非感染性肝炎	肝炎	241	肝障害	K76	424,151
	劇症肝炎	66			
	中毒性肝炎	52			

肝感染	B型肝炎	777	B型肝炎	B16	14,737
	B型肝炎再活性化	272			
	C型肝炎	130			
溶血性障害	溶血性尿毒症症候群	211	溶血性尿毒症症候群	D59	1,899
	溶血性貧血	170			
	溶血	109			
重症皮膚副作用	スティーヴンス・ジョンソン症候群	1,954	スティーブンス・ジョンソン症候群	L51	17,000
	好酸球増加と全身症状を伴う薬物反応	1,276			
	中毒性表皮壊死融解症	768			
アナフィラキシー反応	アナフィラキシー反応	639	アナフィラキシーショック	T78	128,799
	アナフィラキシーショック	426			
	循環虚脱	136			
急性膵炎	膵炎	2,454	急性膵炎	K85	21,377
	急性膵炎	646			
	壊死性膵炎	58			
無顆粒球症	汎血球減少症	787	発熱性好中球減少症	D70	23,755
	無顆粒球症	473			
	発熱性好中球減少症	265			
血管浮腫	血管浮腫	516	喉頭浮腫	J38	62,411
	舌腫脹	143			
	顔面腫脹	98			

喘息／気管支痙攣	気管支痙攣	403	気管支痙攣	J98	9,365
	喘息	23			
	喘息発作重積	5			
脂質異常症	血中トリグリセリド増加	937	高トリグリセライド血症	E78	525,317
	血中コレステロール増加	840			
	血中コレステロール異常	373			
造血障害による2種以上の血球減少症	汎血球減少症	669	汎血球減少症	D61	2,431
	骨髄機能不全	292			
	有熱性骨髄無形成	79			
造血障害による赤血球減少症	赤血球数減少	623	再生不良性貧血	D61	2,431
	大球性貧血	21			
	再生不良性貧血	3			
造血障害による白血球減少症	好中球減少症	2,366	好中球減少症	D70	23,755
	発熱性好中球減少症	978			
	白血球減少症	805			
造血障害による血小板減少症	血小板数減少	3,635	血小板減少症	D69	55,805
	血小板減少症	1,793			
	血小板生成減少	2			
乳酸アシドーシス	乳酸アシドーシス	10,840	乳酸アシドーシス	E87	62,388
	高乳酸血症	447			
	血中乳酸増加	445			

末梢性ニューロパチー	末梢性ニューロパチー	2,950	末梢神経障害	G62	346,314
	多発ニューロパチー	301			
	末梢性感覚ニューロパチー	258			
高血糖/糖尿病の発症	糖尿病	7,367	高血糖症	R73	54,262
	2型糖尿病	3,901			
	高血糖	1,260			
間質性肺疾患	間質性肺疾患	1,146	肺線維症	J84	11,793
	肺線維症	1,000			
	肺毒性	807			
悪性症候群	悪性症候群	1,151	悪性症候群	G21	6,818
	セロトニン症候群	88			
	悪性高熱	9			
全身性エリテマトーデス	全身性エリテマトーデス	2,756	全身性エリテマトーデス	M32	7,616
	ループス様症候群	1,012			
	抗核抗体陽性	346			
味覚および嗅覚障害	味覚不全	728	味覚障害	R43	30,780
	味覚消失	128			
	無嗅覚	90			
心筋梗塞	心筋梗塞	5,705	心筋梗塞	I21	12,518
	急性心筋梗塞	288			
	冠動脈閉塞	191			

洞結節機能障害	洞性徐脈	164	洞不整脈	I49	159,850
	洞停止	28			
	洞結節機能不全	24			
上室性頻脈性不整脈	心房細動	73	心房細動	I48	25,342
	洞性頻脈	19			
	上室性期外収縮	14			
心室性頻脈性不整脈	トルサード ド ポアント	408	トルサードドポアント	I49	159,850
	心室性頻脈	256			
	心室細動	161			
虚血性中枢神経系血管障害	脳血管発作	1,159	脳梗塞	I63	42,671
	一過性脳虚血発作	289			
	脳梗塞	249			
出血性中枢神経系血管障害	脳血管発作	3,813	脳出血	I61	7,164
	脳出血	2,777			
	頭蓋内出血	1,829			
痙攣	てんかん	193	てんかん様発作	R56	54,052
	痙攣発作	66			
	てんかん重積状態	41			
偽膜性大腸炎	クロストリジウム・ディフィシレ感染	275	クロストリジウム・ディフィシル腸炎	A04	115,574
	クロストリジウム・ディフィシレ大腸炎	173			
	クロストリジウム検査陽性	98			

静脈の塞栓および血栓	深部静脈血栓症	3,323	肺塞栓症	I26	3,576
	肺塞栓症	2,575			
	肺血栓症	654			
ジスキネジア	遅発性ジスキネジア	1,973	ジスキネジア	G24	9,290
	ジスキネジア	1,431			
	眼球回転発作	184			
ジストニア	ジストニア	804	ジストニア	G24	9,290
	眼球回転発作	216			
	側反弓	182			
パーキンソン様事象	パーキンソニズム	721	薬剤性パーキンソン症候群	G21	6,818
	筋固縮	371			
	パーキンソン病	182			
薬物乱用および依存	企図的過量投与	840	薬物乱用	F19	1,929
	薬物乱用	804			
	企図的製品誤用	148			
消化管の閉塞	腸閉塞	4,482	イレウス	K56	28,412
	腸管狭窄	1,643			
	小腸閉塞	1,388			
消化管の潰瘍	胃潰瘍	1,481	胃潰瘍	K25	486,818
	出血性胃潰瘍	917			
	十二指腸潰瘍	750			

消化管の穿孔	消化管穿孔	860	腸穿孔	K63	199,559
	腸管穿孔	687			
	大腸穿孔	486			
消化管の出血	胃腸出血	19,991	消化管出血	K92	113,195
	直腸出血	4,319			
	上部消化管出血	3,428			
口腔咽頭の状態 (新生物、感染および アレルギーを除く)	口内炎	2,568	口内炎	K12	363,218
	口腔内痛	588			
	口腔内潰瘍形成	477			
血栓性静脈炎	表在性血栓性静脈炎	1,181	血栓性静脈炎	I80	27,406
	骨盤静脈血栓症	571			
	血栓性静脈炎	263			
精神病および精神病 性障害	幻覚	741	幻覚	R44	69,011
	幻視	111			
	幻聴	93			
胆嚢関連障害	胆石症	397	急性胆のう炎	K81	17,409
	胆嚢障害	92			
	胆嚢炎	58			
胆道障害	黄疸	352	胆道疾患	K83	12,932
	胆汁うっ滞	98			
	高ビリルビン血症	38			

胆石が関連する障害	胆石症	426	胆のう胆管結石症	K80	61,709
	胆管結石	50			
肺高血圧症	肺高血圧症	191	肺高血圧症	I27	1,922
	肺動脈性肺高血圧症	90			
	右室不全	13			
非感染性脳炎	脳炎	58	脳炎	G04	1,636
	自己免疫性脳炎	17			
	免疫性脳炎	4			
非感染性脳症／譫妄	脳症	366	せん妄	F05	2,986
	中毒性脳症	249			
	譫妄	209			
非感染性髄膜炎	無菌性髄膜炎	824	無菌性髄膜炎	G03	3,541
	羞明	180			
	髄膜炎	140			
消化管の非特異的炎症	大腸炎	1,642	大腸炎	A09	2,050,515
	胃炎	93			
	小腸炎	79			
消化管の非特異的機能障害	胃食道逆流性疾患	2,542	逆流性食道炎	K21	581,274
	消化不良	852			
	胃食道括約筋機能不全	1			

消化管の非特異的症 状および処置	悪心	14,307	悪心	R11	894,417
	嘔吐	4,365			
	便秘	2,158			
低ナトリウム血症/ S I A D H	低ナトリウム血症	1,059	低ナトリウム血症	E87	62,388
	抗利尿ホルモン不適合分泌	112			
	血中ナトリウム減少	60			
敵意／攻撃性	攻撃性	1,914	攻撃性気分異常	F38	118
	怒り	906			
	殺人念慮	201			
血行動態的浮腫、蓄水 および体液過負荷	胸水	2,046	浮腫	R60	85,137
	肺水腫	513			
	体液貯留	403			
緑内障	眼圧上昇	128	薬物誘発性緑内障	H40	251,279
	緑内障	91			
	高眼圧症	45			
高血圧	高血圧	946	高血圧症	I10	336,332
	血圧上昇	791			
	拡張期血圧上昇	22			
視神経障害	視神経炎	87	視神経炎	H46	5,787
	視神経症	87			
	視神経萎縮	22			

心筋症	駆出率減少	1,147	心筋症	I42	6,059
	心毒性	451			
	心筋症	253			
水晶体障害	白内障	279	薬物性白内障	H26	151,761
	白内障手術	120			
	中毒性前眼部症候群	20			
好酸球性肺炎	肺臓炎	1,135	好酸球性肺炎	J82	764
	好酸球性肺炎	23			
	好酸球性多発血管炎性肉芽腫症	12			
網膜障害	網膜毒性	182	クロロキン網膜症	H35	171,134
	網膜症	154			
	黄斑症	101			
甲状腺機能低下症	甲状腺機能低下症	842	薬剤性甲状腺機能低下症	E03	59,750
	原発性甲状腺機能低下症	17			
	自己免疫性甲状腺機能低下症	13			
甲状腺機能亢進症	甲状腺機能亢進症	1,169	甲状腺機能亢進症	E05	48,305
	甲状腺中毒クリーゼ	75			
	バセドウ病	18			
うつ病 (自殺/自傷を除く)	うつ病	8,982	うつ病	F32	211,934
	抑うつ気分	1,329			
	大うつ病	234			

その他の虚血性心疾患	ストレス心筋症	18	狭心症	I20	103,091
	狭心症	17			
	プリンツメタル狭心症	12			
聴覚障害	聴器毒性	87	聴覚障害	H91	96,090
	耳鳴	63			
	難聴	59			
血管炎	抗好中球細胞質抗体陽性血管炎	54	ANCA 関連血管炎	M31	2,647
	血管炎	49			
	壊死性血管炎	7			
結膜障害	結膜炎	268	薬物性結膜炎	H10	1,750,634
	結膜充血	168			
	ドライアイ	113			
リポジストロフィー	後天性リポジストロフィー	544	脂肪萎縮症	E88	16,351
	脂肪組織萎縮症	116			
	顔のやせ	31			
骨粗鬆症／骨減少症	骨粗鬆症	436	薬物誘発性骨粗鬆症	M81	89,890
	骨減少症	88			
	骨粗鬆症性骨折	85			
骨壊死	顎骨壊死	7,226	下顎骨壊死	K10	1,961
	非定型大腿骨骨折	596			
	骨壊死	388			

眼感染	眼内炎	776	眼内炎	H44	2,331
	硝子体炎	211			
	眼感染	142			
眼球運動障害	眼瞼下垂	2,388	眼瞼下垂症	H02	113,921
	眼瞼痙攣	68			
	斜視	47			
過敏症	薬物過敏症	4,103	発疹	R21	21,240
	発疹	2,837			
	蕁麻疹	2,073			
非感染性下痢	下痢	3,806	非感染性下痢	K52	81,092
	血性下痢	30			
	処置後下痢	10			
腫瘍崩壊症候群	腫瘍崩壊症候群	186	腫瘍崩壊症候群	E88	16,351
蛋白尿	蛋白尿	1,610	尿細管性蛋白尿	N39	113,070
	尿中蛋白陽性	189			
	尿蛋白	100			
尿細管間質性疾患	中毒性ネフロパシー	572	尿細管間質性腎炎	N12	32,009
	尿細管間質性腎炎	365			
	腎尿細管障害	64			

呼吸不全	呼吸停止	1,194	呼吸不全	J96	116,491
	呼吸抑制	538			
	呼吸不全	227			
腱障害および 靭帯障害	腱炎	2,735	腱炎	M77	190,079
	腱断裂	2,644			
	腱痛	1,618			
好酸球増加および全 身症状を伴う薬物反 応症候群	好酸球増加と全身症状を伴う薬物反応	1,948	好酸球増加症	D72	30,596
	偽リンパ腫	6			
低血糖	低血糖	4,480	インスリン低血糖	E16	37,136
	血中ブドウ糖減少	3,915			
	低血糖昏睡	351			
感染性肺炎	肺炎	11,899	肺炎	J18	289,347
	細菌性肺炎	114			
	ウイルス性肺炎	100			
脱水	脱水	2,237	脱水症	E86	720,775
	血液量減少症	260			
	血液量減少性ショック	93			
低カリウム血症	低カリウム血症	1,897	低カリウム血症	E87	62,388
	血中カリウム減少	327			
	低マグネシウム血症	294			

敗血症	敗血症	1,735	敗血症	A41	14,858
	敗血症性ショック	1,068			
	好中球減少性敗血症	266			
日和見感染	サイトメガロウイルス感染	2,714	サイトメガロウイルス 感染症	B25	3,974
	ポリオーマウイルス関連腎症	701			
	BK ウイルス感染	635			

FAERS のカラムに含まれる Counts は有害事象の報告例数を示す。

JMDC Claims data のカラムに含まれる Counts は、JMDC Claims 内での ICD10 登録例数を示す。これにより、データベース内に症例が存在する ICD10 コードに対応付けられていることが確認された。

表 1-7 ゴールドスタンダード：92 種類の陽性対照

ICD10	Event	ATC code	Drug name	Log ₁₀ (ROR)	Onset	Class
A04	偽膜性大腸炎	J01CR05	タゾバクタム・ピペ ラシリン	1.47	14	short
A09	消化管の非特異的炎症	L01XC11	イピリムマブ	1.35	42	short
A41	敗血症	L01BC01	シタラビン	1.19	13	short
B16	肝感染	L01XC02	リツキシマブ	0.98	165	long
B25	日和見感染	L04AA06	ミコフェノール酸	1.54	67	short
D59	溶血性障害	L04AD02	タクロリムス	0.91	29	short
D61	造血障害による2種 以上の血球減少症	L01BA04	ペメトレキセド	1.16	14	short
D61	造血障害による赤血 球減少症	L01XE33	パルボシクリブ	0.94	16	short
D69	造血障害による血小 板減少症	J05AP01	リバビリン	0.84	9	short
D70	造血障害による白血 球減少症	L01XA03	オキサリプラチン	0.91	20	short
D70	無顆粒球症	J01EE01	スルファメトキサ ゾール・トリメトプ リム	0.94	33	short
D72	好酸球増加および全 身症状を伴う薬物反 応症候群	M04AA01	アロプリノール	2.27	29	short
E03	甲状腺機能低下症	L01XC17	ニボルマブ	1.17	81	short
E05	甲状腺機能亢進症	C01BD01	アミオダロン	1.83	939	long
E16	低血糖	A10AE04	インスリン グラル ギン	1.44	101	long
E78	脂質異常症	N05AH03	オランザピン	0.97	164	long
E86	脱水	C03CA01	フロセミド	0.97	29	short
E87	乳酸アシドーシス	A10BA02	メトホルミン	2.29	209	long
E87	低カリウム血症	C03CA01	フロセミド	1.21	23	short
E87	低ナトリウム血症/ S I A D H	C03BA11	インダパミド	1.89	32	short

E88	リボジストロフィー	J05AF05	ラミブジン	2.03	1,228	long
E88	腫瘍崩壊症候群	L01AA09	ベンダムスチン	1.59	3	short
F05	非感染性脳症／譫妄	J05AB11	バラシクロビル	1.16	3	short
F19	薬物乱用および依存	N06AX11	ミルタザピン	0.82	5	short
F32	うつ病（自殺／自傷を除く）	N07BA03	バレニクリン	0.97	15	short
F38	敵意／攻撃性	N06AB05	パロキセチン	1.12	97	long
G03	非感染性髄膜炎	J06BA01 J06BA02	人免疫グロブリン	1.27	3	short
G04	非感染性脳炎	L01XC32	アテゾリズマブ	1.49	14	short
G21	悪性症候群	N05AD01	ハロペリドール	1.71	5	short
G21	パーキンソン様事象	N05AX12	アリピプラゾール	0.99	28	short
G24	ジストニア	N05AX08	リスペリドン	1.04	80	short
G24	ジスキネジア	N05AX12	アリピプラゾール	1.20	87	short
G62	末梢性ニューロパチー	L01XX32	ボルテゾミブ	1.23	32	short
H02	眼球運動障害	M03AX01	A型ボツリヌス毒素	1.55	9	short
H10	結膜障害	N03AX09	ラモトリギン	0.50	26	short
H26	水晶体障害	S01LA05	アフリベルセプト	0.88	428	long
H35	網膜障害	P01BA02	ヒドロキシクロロキン	0.75	195	long
H40	緑内障	H02AB08 S01BA05	トリアムシノロン アセトニド	1.08	27	short
H44	眼感染	S01LA05	アフリベルセプト	1.65	69	short
H46	視神経障害	J04AK02	エタンブトール	1.58	153	long
H91	聴覚障害	J01GB03	ゲンタマイシン	0.95	9	short
I10	高血圧	L01XE29	レンバチニブ	1.19	4	short
I20	その他の虚血性心疾患	L01XX35	アナグレリド	0.92	152	long
I21	心筋梗塞	M01AH01	セレコキシブ	1.01	120	long
I26	静脈の塞栓および血栓	L04AX04	レナリドミド	0.48	71	short
I27	肺高血圧症	L01XE06	ダサチニブ	0.71	910	long
I42	心筋症	L01XC03	トラスツズマブ	1.42	236	long
I45	トルサード ド ポアン ト/QT 延長	L01XE08	ニロチニブ	1.11	37	short
I48	上室性頻脈性不整脈	B01AC23	シロスタゾール	0.76	28	short
I49	洞結節機能障害	N06DA02	ドネペジル	1.51	255	long

I49	心室性頻脈性不整脈	N06AB10	エスシタロプラム (シタロプラム)	0.79	46	short
I50	心不全	C02KX01	ボセンタン	0.85	38	short
I61	出血性中枢神経系血管障害	B01AF01	リバーロキサバン	0.91	212	long
I63	虚血性中枢神経系血管障害	S01LA04	ラニビズマブ	0.80	67	short
I80	血栓性静脈炎	G03FA17	ドロスピレノン・エチニルエストラジオール	1.93	132	long
J18	感染性肺炎	L04AX04	レナリドミド	0.57	91	long
J38	血管浮腫	C09AA02	エナラプリル	0.68	50	short
J82	好酸球性肺炎	L01XC17	ニボルマブ	1.34	43	short
J84	間質性肺疾患	C01BD01	アミオダロン	1.46	176	long
J96	呼吸不全	N05BA01	ジアゼパム	0.94	1	short
J98	喘息／気管支痙攣	N01AX10	プロポフォール	0.85	1	short
K10	骨壊死	M05BX04	デノスマブ	1.42	447	long
K12	口腔咽頭の状態(新生物、感染およびアレルギーを除く)	L01XE10 L04AA18	エベロリムス	0.73	15	short
K21	消化管の非特異的機能障害	M05BA04	アレンドロン酸	1.03	214	long
K25	消化管の潰瘍	B01AC06 N02BA01	アスピリン	0.96	232	long
K52	非感染性下痢	L01XX19	イリノテカン	0.81	13	short
K56	消化管の閉塞	L04AB04	アダリムマブ	0.74	114	long
K63	消化管の穿孔	L01XC07	ベバシズマブ	1.06	72	short
K71	肝臓に起因する胆汁うっ滞および黄疸	J01CR02	アモキシシリン・クラブラン酸	1.41	33	short
K76	非感染性肝炎	J04AC01	イソニアジド	1.36	37	short
K80	胆石が関連する障害	J01DD04	セフトリアキソン	1.18	18	short
K81	胆嚢関連障害	H01CB02	オクトレオチド	0.82	207	long
K83	胆道障害	J04AC01	イソニアジド	1.22	56	short
K85	急性膵炎	A10BH01	シタグリプチン	1.38	135	long
K92	消化管の出血	B01AF01	リバーロキサバン	1.47	121	long

L51	重症皮膚副作用	N03AX09	ラモトリギン	1.22	25	short
M31	血管炎	H03BA02	プロピルチオウラシル	2.07	748	long
M32	全身性エリテマトーデス	L04AB04	アダリムマブ	0.74	324	long
M62	横紋筋融解症／ミオパチー	C10AA01	シンバスタチン	1.72	597	long
M77	腱障害および靭帯障害	J01MA12	レボフロキサシン	1.89	6	short
M81	骨粗鬆症／骨減少症	H02AB06	プレドニゾロン	0.67	169	long
N12	尿細管間質性疾患	J01XA01	バンコマイシン	1.41	9	short
N17	急性腎不全	J01XA01	バンコマイシン	1.02	8	short
N39	蛋白尿	L01XC07	ベバシズマブ	1.41	169	long
R11	消化管の非特異的症状および処置	N07BA03	バレニクリン	0.51	10	short
R21	過敏症	J01CA04	アモキシシリン	0.94	7	short
R43	味覚および嗅覚障害	J01FA09	クラリスロマイシン	0.88	7	short
R44	精神病および精神病性障害	J05AH02	オセルタミビル	1.00	1	short
R56	痙攣	J01DH51	イミペネム・シラス タチン	1.06	5	short
R60	血行動態的浮腫、蓄水 および体液過負荷	L01XE06	ダサチニブ	0.70	47	short
R73	高血糖／糖尿病の発症	N05AH04	クエチアピン	0.90	293	long
T78	アナフィラキシー反応	M03AC09	ロクロニウム	1.80	1	short

表 1-8 ゴールドスタンダード：88 種類の陰性対照

ICD10	Event	ATC code	Drugname	Log ₁₀ (ROR)
A04	偽膜性大腸炎	N05AH03	オランザピン	-0.41
A04	偽膜性大腸炎	C01BD01	アミオダロン	-0.09
A41	敗血症	N06DA02	ドネペジル	-0.36
A41	敗血症	N05BA01	ジアゼパム	-0.09
B16	肝感染	N05AX12	アリピプラゾール	-1.09
B16	肝感染	C03CA01	フロセミド	-0.32
B25	日和見感染	N06AB05	パロキセチン	-0.92
B25	日和見感染	B01AC23	シロスタゾール	-0.19
D59	溶血性障害	N07BA03	バレニクリン	-0.95
D61	造血障害による2種以上の血球減少症	N07BA03	バレニクリン	-1.92
D69	造血障害による血小板減少症	S01LA05	アフリベルセプト	-0.22
D70	無顆粒球症	M03AX01	A型ボツリヌス毒素	-2.00
D70	造血障害による白血球減少症	S01LA04	ラニビズマブ	-1.22
D70	無顆粒球症	A10AE04	インスリン グラルギン	-1.15
E03	甲状腺機能低下症	L01XX19	イリノテカン	-0.52
E03	甲状腺機能低下症	J01MA12	レボフロキサシン	-0.01
E05	甲状腺機能亢進症	M01AH01	セレコキシブ	-0.47
E05	甲状腺機能亢進症	L01XA03	オキサリプラチン	-0.42
E16	低血糖	L01BA04	ペメトレキセド	-0.58
E16	低血糖	L01XE33	パルボシクリブ	-0.55
E78	脂質異常症	J01DD04	セフトリアキソン	-0.44
E87	低ナトリウム血症/S I A D H	S01LA05	アフリベルセプト	-0.54
E88	腫瘍崩壊症候群	N03AX09	ラモトリギン	-1.66
E88	腫瘍崩壊症候群	J05AP01	リバビリン	-0.88
E88	リポジストロフィー	A10BH01	シタグリプチン	-0.81
E88	リポジストロフィー	N05AH04	クエチアピン	-0.60
F19	薬物乱用および依存	L01AA09	ベンダムスチン	-1.57
F19	薬物乱用および依存	C10AA01	シンバスタチン	-0.30
F32	うつ病（自殺／自傷を除く）	B01AF01	リバーロキサバン	-0.76
F38	敵意／攻撃性	B01AC06 N02BA01	アスピリン	-0.63

F38	敵意／攻撃性	M05BA04	アレンドロン酸	-0.52
G03	非感染性髄膜炎	L01XA03	オキサリプラチン	-0.19
G04	非感染性脳炎	A10AE04	インスリン グラルギン	-0.56
G21	悪性症候群	L04AB04	アダリムマブ	-1.60
G21	パーキンソン様事象	L01XX32	ボルテゾミブ	-0.54
G24	ジスキネジア	L01XC02	リツキシマブ	-1.12
G24	ジスキネジア	B01AF01	リバーロキサバン	-1.08
G24	ジストニア	J01XA01	バンコマイシン	-0.45
G62	末梢性ニューロパチー	C02KX01	ボセンタン	-0.53
G62	末梢性ニューロパチー	C09AA02	エナラプリル	-0.11
H02	眼球運動障害	L04AX04	レナリドミド	-0.73
H10	結膜障害	A10AE04	インスリン グラルギン	-0.40
H26	水晶体障害	J01EE01	スルファメトキサゾール・トリメトプリム	-0.34
H26	水晶体障害	L01XE10 L04AA18	エベロリムス	-0.26
H35	網膜障害	J01CR05	タゾバクタム・ピペラシリン	-1.00
H40	緑内障	L01XX19	イリノテカン	-1.12
H40	緑内障	L04AA06	ミコフェノール酸	-0.07
H44	眼感染	M04AA01	アロプリノール	-0.14
H46	視神経障害	A10BA02	メトホルミン	-0.34
H91	聴覚障害	A10BA02	メトホルミン	-0.31
I20	その他の虚血性心疾患	H03BA02	プロピルチオウラシル	-0.50
I21	心筋梗塞	J04AC01	イソニアジド	-0.61
I26	静脈の塞栓および血栓	J05AB11	バラシクロビル	-0.59
I27	肺高血圧症	J01MA12	レボフロキサシン	-0.49
I42	心筋症	J01DH51	イミペネム・シラスタチン	-0.63
I48	上室性頻脈性不整脈	J04AK02	エタンブトール	-0.42
I61	出血性中枢神経系血管障害	N05AD01	ハロペリドール	-0.36
I63	虚血性中枢神経系血管障害	J05AH02	オセルタミビル	-0.62
I80	血栓性静脈炎	J01MA12	レボフロキサシン	-0.24
J18	感染性肺炎	C03BA11	インダパミド	-0.25
J84	間質性肺疾患	G03FA17	ドロスピレノン・エチニルエストラジオール	-0.43

J96	呼吸不全	G03FA17	ドロスピレノン・エチニルエストラジオール	-0.60
K10	骨壊死	N05AX12	アリピプラゾール	-1.27
K12	口腔咽頭の状態（新生物、感染およびアレルギーを除く）	C02KX01	ボセンタン	-0.37
K56	消化管の閉塞	M05BX04	デノスマブ	-0.55
K56	消化管の閉塞	J05AF05	ラミブジン	-0.06
K63	消化管の穿孔	N05AX08	リスペリドン	-0.83
K80	胆石が関連する障害	B01AF01	リバーロキサバン	-0.54
K80	胆石が関連する障害	J01FA09	クラリスロマイシン	-0.28
K83	胆道障害	H02AB08 S01BA05	トリアムシノロンアセトニド	-1.04
K92	消化管の出血	M03AX01	A型ボツリヌス毒素	-1.30
K92	消化管の出血	P01BA02	ヒドロキシクロロキン	-0.33
M31	血管炎	N06DA02	ドネペジル	-0.31
M32	全身性エリテマトーデス	N06AX11	ミルタザピン	-0.69
M32	全身性エリテマトーデス	J01GB03	ゲンタマイシン	-0.48
M62	横紋筋融解症／ミオパチー	S01LA04	ラニビズマブ	-1.38
M62	横紋筋融解症／ミオパチー	G03FA17	ドロスピレノン・エチニルエストラジオール	-1.35
M77	腱障害および靭帯障害	L01XX35	アナグレリド	-1.07
M77	腱障害および靭帯障害	L01XC03	トラスツズマブ	-0.73
M81	骨粗鬆症／骨減少症	L01BC01	シタラビン	-0.39
N12	尿細管間質性疾患	N05AX12	アリピプラゾール	-0.70
N17	急性腎不全	N06AB10	エスシタロプラム (シタロプラム)	-0.14
N39	蛋白尿	N05BA01	ジアゼパム	-0.28
R43	味覚および嗅覚障害	H02AB08 S01BA05	トリアムシノロンアセトニド	-0.24
R44	精神病および精神病性障害	S01LA05	アフリベルセプト	-1.03
R44	精神病および精神病性障害	H01CB02	オクトレオチド	-0.37
R56	痙攣	L01XX35	アナグレリド	-0.18
R73	高血糖／糖尿病の発症	J01CA04	アモキシシリン	-0.52

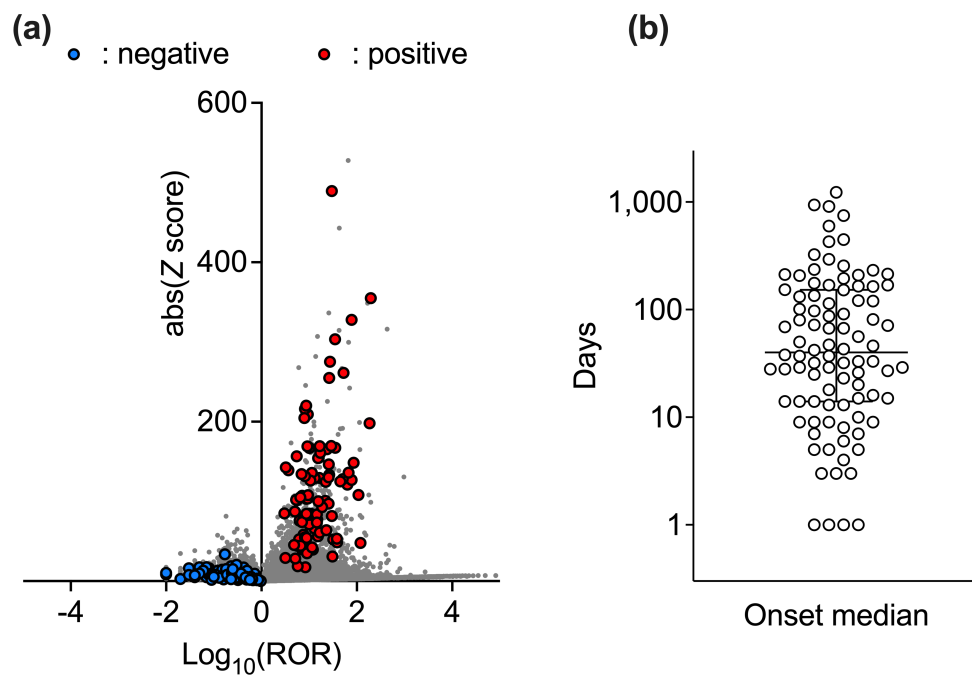


図 1-2 大規模有害事象データに基づくゴールドスタンダードの特徴

(a) FAERS データにおける対数スケールの ROR とその統計的有意性 (Z score) を視覚化した volcano plot (563,805 plots)。陽性対照は赤丸で、陰性対照は青丸で示した。(b) 92 種の陽性対照における副作用発症時期の情報

表 1-9 ゴールドスタンダードに含まれる ICD10 分類の種類と例数

ICD10 category	ICD10 detail	No. of categories
A	寄生虫感染症	7
B	感染性疾患	6
C	悪性新生物	0
D	良性新生物	13
E	代謝疾患	22
F	精神疾患	9
G	神経疾患	16
H	眼科および耳鼻咽喉科疾患	18
I	循環器系疾患	23
J	呼吸器系疾患	9
K	消化器系疾患	24
L	皮膚疾患	1
M	筋骨格系疾患	13
N	腎疾患	6
R, T	その他	13

考察

本研究では、大規模臨床データの統計解析により合理的に選択した 92 組の陽性対照及び 88 組の陰性対照から成るゴールドスタンダードを構築した。特定の副作用を対象にしたこれまでのゴールドスタンダードと異なり、本ゴールドスタンダードには解剖学的、治療学的、化学的に多様な特性を持つ医薬品が含まれ、また複数の副作用が含まれる。副作用の種類については表 1-9 の結果より、精神疾患や代謝疾患、腎疾患など多様な臓器を対象にしていることが確認された。

ゴールドスタンダードは作成の情報源が異なれば、陽性対照や陰性対照の包括性や正確性も異なる可能性がある。ランダム化比較試験は医薬品の有効性の面では高いエビデンスレベルを示すが、安全性の面では特定の集団に限定されており、これらを情報源にした場合は稀なイベントを対象にすることは困難となる可能性がある。SRS の統計解析に基づき作成された本ゴールドスタンダードの陽性対照は、臨床現場で特に注意が必要で注目すべき「医薬品－有害事象」の組合せが含まれると考えられる。実際に、PMDA より緊急安全性情報・安全性速報が発令された「クエチアピリン－高血糖」、「ラモトリギン－重症皮膚副作用」、「ドロスピレノン・エチニルエストラジオール ベータデクス－血栓性静脈炎」、「オセルタミビル－精神病および精神病性障害」、「リツキシマブ－肝感染」などのペアが陽性対照に含まれることが確認された [43]。陰性対照について、必ず起こり得ない副作用ペアを作成することは困難と判断し、本検討では「関連性が極めて低いと考え得るペア」と定義した。陰性対照作成の条件において、FAERS では被疑薬として 1 例も報告されていないペアは大部分の SMQ で見られなかった。そのため本検討では「 $ROR < 1$ となるペア」と JADER とは異なる基準を設定した。この背景として FAERS は報告例数の規模の大きさや、また報告者の傾向が JADER と異なることも要因として考えられた。JADER は医療従事者からの報告が大部分を占める一方 [44]、FAERS は約 40%もの比率が消費者からの報告を占める [45]。しかし、上記の基準で陰性対照を作成したが、今後新たに認知される副作用が出現する可能性はあり、従ってゴールドスタンダードは動的であると考えられるべきで、定期的な再評価が必要と考えられる。

副作用の発症時期は、JADER の各レコードに含まれる投与開始日と有害事象発現日の差分による中央値から推定・追加した。「ロクロニウム－アナフィラキシー：期間 1 日」、「バンコマイシン－急性腎障害：期間 8 日」など早期性のものや、「アリピプラゾール－ジスキネジア：期間 87 日」のように早期から遅発にかけて発生するもの、そして「アミオダロン－甲状腺機能障害：期間 939 日」のようにアミオダロンに含まれるヨウ素の蓄積により発症する遅発性のものなど、薬理的にも妥当と考え得る発症時期情報が推定されていることが確認された。

しかし、妥当であるかの判定が困難なものも含まれる。具体的には「インスリンー低血糖：期間 101 日」と算出されたが、これは医薬品の投与量や使用方法（誤った使用など）に大きく依存する可能性がある。副作用の発現パターンについてはワイブル分布を活用することで、薬剤誘発性の消化管閉塞の発現時期を推定した研究などが報告されている [46–48]。有用性は限定的ではあるものの、ゴールドスタンダードにおける副作用発症時期パターンについてはさらなる検討が必要と考えられた。

本検討では手作業により 92 種類の SMQ を ICD10 に対応付けた。医学用語は標準化されていないため、その用語は医療データベース間で異なる。請求データに診断を記録するための ICD10 コードや、EHR ベンダーに積極的に使用される SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) [49]、臨床試験や自発報告データにおける有害事象の記録に用いられる MedDRA などいくつかの語彙が存在する。そのため異なるデータからの変換を可能とする体系的なプロセスの開発が求められており、医学用語のマッピングを可能にするメタシンソーラス (UMLS: Unified Medical Language System) も開発されている [50]。本検討においても MedDRA の SMQ と ICD10 を UMLS 経由でマッピングさせることを試みたが大部分の紐付けが困難であった。機械的な対応付けが困難な原因として語彙の表記揺れと階層の違いのほか、正確な対応には医学的な解釈が求められる可能性が考えられた。そのためマッピングは複数人の独立した専門家により識別され、不一致が生じた場合はさらに別の専門家による検討を行うことで最終決定を行う、などの対応付けプロセスが実施されている [51]。適切な対応付けには診断名や処方薬の情報だけでなく、処置情報および検査値情報を活用し、さらに複数人の専門家による臨床的な意味合いも含めた対応付けのプロセスが不可欠であると考えられる。本検討では、上記のプロセスを経ていないため、対応付けについてはある程度の主観的な部分が含まれる点は避けられない。また ICD10 コードは階層構造となっており、ICD10 大分類コードは 22 種類、中分類コードは 224 種類、小分類コードは 1,500 種類、細分類コードは 6562 種類存在する。対応付けにおいては、意味合いに幅を持たせること、そして解析における計算量を考慮して今回は ICD10 小分類コードと対応させており、必然的に不完全なマッピングによる情報の損失に関する懸念も生じる。さらに MedDRA は毎年 2 回の更新が行われるため、対応付けられた結果も定期的な更新が求められると考える。

上記の限界点があるものの、本ゴールドスタンダードにより、データマイニング手法で既知の副作用がどの程度検出されるかを定量的に評価することが可能となった。レセプトデータをはじめ電子カルテデータ等の大規模ヘルスケアデータを活用したシグナル検出法の性能評価の基盤となることが期待される。

第 2 章

レセプトの相関ルールマイニングによる

副作用シグナルの早期検出

相関ルールマイニング (ARM: Association Rule Mining) は膨大なデータから共起確率が高いアイテムを効率的に見出す解析手法である [52]。SRS や電子カルテなどの医療データ解析において、投薬・診断・臨床転帰の間の未発見の関連性を特定する手法として ARM の有用性は報告されており、ファーマコビジランスや薬理学研究のためのアプローチとしても提案されている [9, 53–55]。そこで、レセプトデータに登録される処方薬や症状に対して ARM を適用し、着目する医薬品と共起しやすい症状を抽出することで広範な副作用シグナルを能動的かつ早期に検出できるのではないかと考えた。

本章では、民間レセプトデータベース JMDC Claims に ARM を適応させ、第 1 章で構築したゴールドスタンダードを用いて、副作用シグナル検出における ARM の有用性を既存手法 (後述) との比較により評価した。

方法

1. 相関ルールマイニング (ARM: Association Rule Mining)

トランザクションのセット（各トランザクションはアイテムのセットを含む）が与えられたとき、ARM は $X \rightarrow Y$ と表現される。ここで X と Y はアイテムのセットを表す。Support はそのルールがトランザクション内でどの程度頻繁に発現するかを示し、以下のように算出された。

$$\text{Support}(X, Y) = P(X \cap Y)$$

Confidence は条件付き確率 $P(Y|X)$ に相当し、次のように算出された。

$$\text{Confidence}(X \rightarrow Y) = \frac{P(X \cap Y)}{P(X)}$$

Lift は X の発現によって Y の発現がどの程度増加したかを表したものである。Lift は以下のように算出された。

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Confidence}(X \rightarrow Y)}{\text{Support}(Y)}$$

Conviction は X の発現によって Y の発現しない確率がどの程度減少したかを表したものである。Conviction は以下のように算出された。

$$\text{Conviction}(X \rightarrow Y) = \frac{1 - \text{Support}(Y)}{1 - \text{Confidence}(X \rightarrow Y)}$$

Conviction は lift と異なり、結果的に発現しない情報も利用するため、ルールの方向に対して敏感である。lift ($X \rightarrow Y$) = lift ($Y \rightarrow X$) となるが、conviction ($X \rightarrow Y$) \neq conviction ($Y \rightarrow X$) となる。一般に ARM の検出基準として lift > 1 が用いられるが、本検討では conviction > 1 も用いた [56]。「医薬品－有害事象」ペアの関連性の強さは lift と conviction を算出することによって評価し、統計的有意性はカイ二乗検定により推定した。カイ二乗の値は以下のように計算された。

$$\text{Chi square} = n(\text{Lift} - 1)^2 \frac{\text{Support} * \text{Confidence}}{(\text{Confidence} - \text{Support})(\text{Lift} - \text{Confidence})}$$

ここで、10未満の値があった場合はカイ二乗検定の代わりにフィッシャーの正確確率検定を使用した。ARMはR version 4.0.2 (2020-06-22)のarulesライブラリのApriori関数を用いて行った。本検討では、レセプトデータに登録される医薬品と疾患（ここでは有害事象）にARMを適応させることで、副作用シグナルのスクリーニングに有効であるかを検討した。

2. 前処理

レセプトデータに登録される各症例の加入開始日の情報、医薬品の情報とその処方月、ICD10コードとその登録月の記録を抽出した。本研究では各医薬品と各疾患の初回発生のみを解析に用い、さらにRun-in periodを設けた。Run-in periodは保険加入前に既に医薬品を処方されていた症例や病気と診断された症例を除外するためのウォッシュアウト期間である [57]。ARMは事前にsupportの閾値を設定し、その閾値を超える全てのパターンを算出する。第1章で構築したゴールドスタンダードには非常に稀な頻度で発生する副作用が含まれる可能性があるため、本検討におけるsupportの閾値は 1×10^{-10} に設定した。レセプトデータへのARMの有用性は、5種類の検討により評価を行った。ARMは、医薬品と有害事象の2項目のみの関連性を評価するために使用された。

3. 感度 (sensitivity)、特異度 (specificity)、適合率 (precision)、F-measure

ARMの有用性を評価するために感度 (sensitivity)、特異度 (specificity)、適合率 (precision) を算出した。F-measureは、感度と適合率の調和平均から算出した。

4. Sequence Symmetry Analysis (SSA)

本検討では既存手法であるSSAとの比較でARMの有用性を評価した。SSAは「医薬品の処方」と「イベントの登録」が発生する順番に着目して、人数分布の偏りから副作用シグナルを検出する解析手法である [21, 22]。医薬品を使った後にイベントが登録された人数が、医薬品を使う前にイベントが登録された人

数より多ければ、その医薬品はイベントを引き起こす可能性があると考えられ、これまでに数種の副作用シグナルが同定されている [23–25]。1996 年に SSA の解析概念が発表されて以降、本手法を用いた研究論文数は増加傾向にあり、疫学研究で最も活用されている [58, 59]。SSA は注目するイベントが発生した症例のみを解析対象とした自己対照デザインのため時間非依存性 (交絡) 因子は未知のものを含め、解析において自動的に調整される特徴を有す。以上の背景から、SSA を ARM の性能を比較するための最良のベースライン手法とみなした。

結果・考察

1. ARM の性能の算出

はじめに、2005年から2019年までの15年間のレセプトデータにARMを適応させた（加入者数: 6,072,316人、レコード数: 162,454,898件）。十分な期間と考えられる6ヶ月のRun-in periodを設定し、ARMとSSAの性能を算出した（表2-1, 2-2）。カイ二乗検定の p -valueは極端に小さい値に対応するため、対数スケール $-\text{Log}_{10}(p\text{-value})$ で算出した。ARMの副作用シグナルの検出基準を $\text{lift} > 1$ 、 $\text{conviction} > 1$ 、 $p\text{-value} < 0.05$ とした場合、156組のシグナルが同定され、そのうち90組が真の陽性対照であった（感度: 0.98、特異度: 0.25、適合率: 0.58、F-measure: 0.73）。SSAでは59組のシグナルが同定され、そのうち42組が真の陽性対照であった（感度: 0.46、特異度: 0.81、適合率: 0.46、F-measure: 0.56）。両手法ともでシグナルとして検出されたのは57組であり、そのうち42組が真の陽性対照であった。ARMは高い感度を示すが特異度が低く、一方、SSAはARMに比べて低い感度を示すが特異度は高い傾向がみられた。ARMの検出基準としてlift値を指標とした既報が大部分であり、また今回算出された陽性対照の「医薬品－有害事象」ペアのlift値は、陰性対照のlift値より高い値を示す傾向が見られた。そのためliftを閾値としてARMの性能をROC (Receiver-Operating-Characteristic) 曲線とPR (Precision-Recall) 曲線により視覚化した（図2-1 a, b）。その結果、ROC-AUCは0.80、PR-AUCは0.83を示し（いずれも0.8以上）、ARMは感度を維持したまま適度な特異度と適合率を達成できる可能性が示された。lift値を変化させた場合の具体的な感度、特異度、適合率、F-measureの推移を算出した（表2-3）。conviction値を含めてlift値を変化させた場合、 $\text{lift} > 2.301$ かつ $\text{conviction} > 1.0025$ のとき、F-measureは最大の0.78を示した（感度: 0.85、特異度: 0.66、適合率: 0.72）。最適なliftとconviction値は活用するデータベースによって異なると考えられる。副作用シグナルを迅速に検出することが重要と考えた場合、感度のような陽性対照から計算される性能指標が重視される。しかし感度を犠牲にしてliftやconviction値を上げることで、ARMは感度と特異度のバランスを保ちながらファーマコビジランスを強化できる可能性が示された。

表 2-1 2005 年から 2019 年までのレセプトデータを使用した ARM と SSA の性能：陽性対照

ICD10	Event	Drug name	ARM				SSA		
			Lift	Count	Conviction	-Log ₁₀ (p-value)	ASR	ASR 95%CI Lower	ASR 95%CI Upper
A04	偽膜性大腸炎	タゾバクタム・ピペラシリン	3.70	1,164	1.0553	510.89	1.46	1.28	1.67
A09	消化管の非特異的炎症	イピリムマブ	1.57	18	1.4074	1.74	3.12	1.19	8.20
A41	敗血症	シタラビン	242.79	1,042	2.4575	54636.26	3.57	3.04	4.20
B16	肝感染	リツキシマブ	43.45	217	1.1152	1961.60	0.93	0.68	1.27
B25	日和見感染	ミコフェノール酸	530.15	399	1.5303	45801.71	3.05	2.39	3.88
D59	溶血性障害	タクロリムス	129.68	158	1.0419	4387.25	0.80	0.55	1.15
D61	造血障害による赤血球減少症	パルボシクリブ	15.66	2	1.0059	2.13	4.80	0.30	76.68
D61	造血障害による2種以上の血球減少症	ペメトレキセド	18.45	8	1.0070	7.69			
D69	造血障害による血小板減少症	リバビリン	6.91	70	1.0580	78.87	0.36	0.22	0.59
D70	造血障害による白血球減少症	オキサリプラチン	54.58	805	1.2665	9239.08	4.95	4.10	5.99
D70	無顆粒球症	スルファメトキサゾール・ トリメトプリム	36.51	3,208	1.1620	24243.12	2.43	2.24	2.65
D72	好酸球増加および全身症状を伴う薬物反応症候群	アロプリノール	3.00	565	1.0102	167.21	0.78	0.65	0.93
E03	甲状腺機能低下症	ニボルマブ	26.14	197	1.3330	1046.51	2.72	1.96	3.76

E05	甲状腺機能亢進症	アミオダロン	7.77	171	1.0574	222.34	1.57	1.14	2.17
E16	低血糖	インスリン グラルギン	9.18	630	1.0530	1007.48	1.75	1.46	2.09
E78	脂質異常症	オランザピン	2.37	4,351	1.1496	827.32	0.95	0.89	1.00
E86	脱水	フロセミド	2.35	13,189	1.2227	2547.78	0.76	0.73	0.79
E87	乳酸アシドーシス	メトホルミン	2.97	1,590	1.0209	462.32	1.52	1.37	1.69
E87	低カリウム血症	フロセミド	14.57	7,072	1.1640	19767.89	1.50	1.42	1.59
E87	低ナトリウム血症 / SIADH	インダパミド	4.81	397	1.0412	264.67	1.32	1.06	1.63
E88	リポジストロフィー	ラミブジン	54.61	10	1.1693	116.06	0.64	0.12	3.48
E88	腫瘍崩壊症候群	ベンダムスチン	27.70	27	1.0777	152.82	2.62	1.03	6.64
F05	非感染性脳症／譫妄	バラシクロビル	1.57	209	1.0003	10.82	1.34	1.01	1.77
F19	薬物乱用および依存	ミルタザピン	11.55	130	1.0034	275.24	1.22	0.85	1.76
F32	うつ病（自殺／自傷を除く）	バレニクリン	3.10	3,427	1.0822	1104.64	0.85	0.80	0.91
F38	敵意／攻撃性	パロキセチン	20.30	12	1.0004	49.32	2.70	0.56	12.99
G03	非感染性髄膜炎	人免疫グロブリン	21.98	108	1.0124	472.33	1.47	0.87	2.47
G04	非感染性脳炎	アテゾリズマブ		0					
G21	パーキンソン様事象	アリピプラザール	70.54	3,165	1.0848	47484.63	2.73	2.51	2.98
G21	悪性症候群	ハロペリドール	64.74	1,085	1.0772	14841.48	1.47	1.26	1.71
G24	ジスキネジア	アリピプラザール	6.18	378	1.0080	361.25	2.20	1.77	2.73
G24	ジストニア	リスペリドン	8.67	237	1.0119	352.66	2.09	1.55	2.80
G62	末梢性ニューロパチー	ボルテゾミブ	6.31	121	1.4737	126.11	1.83	1.25	2.68

H02	眼球運動障害	A型ボツリヌス毒素	4.20	257	1.0652	140.42	0.62	0.47	0.81
H10	結膜障害	ラモトリギン	1.15	2,880	1.0647	18.26	0.63	0.59	0.68
H26	水晶体障害	アフリベルセプト	13.57	1,095	1.4753	2842.93	0.45	0.39	0.53
H35	網膜障害	ヒドロキシクロロキン	12.01	293	1.4688	662.42	1.76	1.34	2.31
H40	緑内障	トリアムシノロンアセトニド	2.64	12,802	1.0763	3021.96	0.95	0.92	0.99
H44	眼感染	アフリベルセプト	17.75	22	1.0065	76.94	4.20	1.50	11.79
H46	視神経障害	エタンブトール	159.26	311	1.1778	10637.11	7.87	4.93	12.57
H91	聴覚障害	ゲンタマイシン	2.03	901	1.0168	105.89	0.89	0.77	1.02
I10	高血圧	レンバチニブ	7.66	53	1.6400	71.85	1.75	0.86	3.57
I20	その他の虚血性心疾患	アナグレリド	21.46	43	1.5466	186.90	1.44	0.77	2.72
I21	心筋梗塞	セレコキシブ	2.36	1,835	1.0028	335.90	1.19	1.08	1.30
I26	静脈の塞栓および血栓	レナリドミド	48.36	9	1.0287	12.26	0.54	0.10	2.96
I27	肺高血圧症	ダサチニブ	29.81	3	1.0092	3.81			
I42	心筋症	トラスツズマブ	10.83	22	1.0099	43.95	2.71	1.05	6.98
I45	トルサード ド ポアント/ Q T延長	ニロチニブ	14.57	8	1.0469	6.96	2.71	0.53	13.98
I48	上室性頻脈性不整脈	シロスタゾール	14.29	523	1.0590	1413.75	1.07	0.87	1.30
I49	心室性頻脈性不整脈	エスシタロプラム	2.71	3,422	1.0484	830.15	0.79	0.73	0.85
I49	洞結節機能障害	ドネペジル	4.19	164	1.0944	90.24	0.71	0.51	0.98
I50	心不全	ボセンタン	19.56	61	1.5033	239.04	0.20	0.10	0.40

I61	出血性中枢神経系血管障害	リバーロキサバン	10.14	76	1.0109	137.68	1.16	0.73	1.83
I63	虚血性中枢神経系血管障害	ラニビズマブ	8.58	128	1.0567	189.19	0.81	0.57	1.17
I80	血栓性静脈炎	ドロスピレノン・エチニル エストラジオール	3.92	291	1.0134	140.16	3.46	2.63	4.56
J18	感染性肺炎	レナリドミド	6.91	104	1.4195	121.27	2.24	1.50	3.35
J38	血管浮腫	エナラプリル	1.94	205	1.0098	21.51	0.84	0.64	1.12
J82	好酸球性肺炎	ニボルマブ	31.13	3	1.0038	3.86	1.04	0.09	11.46
J84	間質性肺疾患	アミオダロン	46.13	248	1.0963	2385.70	6.41	4.38	9.40
J96	呼吸不全	ジアゼパム	3.56	10,939	1.0527	4585.02	1.05	1.01	1.10
J98	喘息／気管支痙攣	プロポフォール	6.71	2,682	1.0089	2960.24	1.64	1.48	1.83
K10	骨壊死	デノスマブ	11.47	17	1.0034	36.53	3.75	1.32	10.65
K12	口腔咽頭の状態（新生物、感 染およびアレルギーを除く）	エベロリムス	9.91	290	2.3103	538.71	1.60	1.20	2.14
K21	消化管の非特異的機能障害	アレンドロン酸	4.56	5,433	1.6042	3633.18	0.62	0.58	0.65
K25	消化管の潰瘍	アスピリン	3.54	13,486	1.2848	5860.83	1.19	1.14	1.23
K52	非感染性下痢	イリノテカン	5.07	182	1.0583	132.42	0.54	0.39	0.74
K56	消化管の閉塞	アダリムマブ	18.95	152	1.0922	565.91	1.04	0.75	1.46
K63	消化管の穿孔	ベバシズマブ	4.46	484	1.1332	293.39	0.18	0.14	0.24
K71	肝臓に起因する胆汁うっ滞 および黄疸	アモキシシリン・クラブラ ン酸	1.52	390	1.0005	16.64	1.25	1.02	1.54
K76	非感染性肝炎	イソニアジド	3.04	539	1.1813	174.33	1.07	0.89	1.28

K80	胆石が関連する障害	セフトリアキソン	2.41	6,337	1.0147	1198.30	0.94	0.89	0.99
K81	胆嚢関連障害	オクトレオチド	21.66	85	1.0631	366.51	0.43	0.26	0.72
K83	胆道障害	イソニアジド	4.63	25	1.0078	16.52	1.77	0.73	4.31
K85	急性膵炎	シタグリプチン	4.57	610	1.0128	374.27	0.77	0.66	0.91
K92	消化管の出血	リバーロキサバン	3.22	381	1.0439	130.32	1.25	1.02	1.54
L51	重症皮膚副作用	ラモトリギン	2.55	62	1.0044	13.72	1.40	0.82	2.38
M31	血管炎	プロピルチオウラシル	18.47	19	1.0077	69.58	3.99	1.16	13.79
M32	全身性エリテマトーデス	アダリムマブ	28.84	62	1.0362	364.10	0.81	0.48	1.36
M62	横紋筋融解症／ミオパチー	シンバスタチン	3.04	258	1.0280	79.07	0.72	0.56	0.92
M77	腱障害および靭帯障害	レボフロキサシン	1.71	57,088	1.0235	4592.38	1.18	1.16	1.20
M81	骨粗鬆症／骨減少症	プレドニゾロン	3.40	16,747	1.0373	6604.06	2.25	2.17	2.34
N12	尿細管間質性疾患	バンコマイシン	10.08	379	1.0506	679.44	1.34	1.08	1.66
N17	急性腎不全	バンコマイシン	104.91	709	1.1093	15881.53	0.85	0.70	1.03
N39	蛋白尿	ベバシズマブ	6.55	403	1.1178	421.52	1.93	1.57	2.36
R11	消化管の非特異的症候および処置	バレニクリン	1.43	6,654	1.0795	218.59	1.10	1.04	1.16
R21	過敏症	アモキシシリン	2.79	8,917	1.0063	2637.58	1.27	1.22	1.33
R43	味覚および嗅覚障害	クラリスロマイシン	1.82	17,991	1.0042	2142.80	1.92	1.85	1.99
R44	精神病および精神病性障害	オセルタミビル	1.95	18,536	1.0110	2179.70	0.52	0.50	0.54
R56	痙攣	イミペネム・シラスタチン	0.99	52	0.9999	0.02	0.57	0.31	1.02

R60	血行動態的浮腫、蓄水および体液過負荷	ダサチニブ	11.66	52	1.1787	113.11	1.78	0.99	3.18
R73	高血糖／糖尿病の発症	クエチアピン	2.90	399	1.0175	110.83	0.80	0.66	0.98
T78	アナフィラキシー反応	ロクロニウム	1.30	6,339	1.0066	104.00	0.72	0.69	0.76

表 2-2 2005 年から 2019 年までのレセプトデータを使用した ARM と SSA の性能：陰性対照

ICD10	Event	Drug name	ARM				SSA		
			Lift	Count	Conviction	-Log ₁₀ (p-value)	ASR	ASR 95%CI Lower	ASR 95%CI Upper
A04	偽膜性大腸炎	アミオダロン	1.21	64	1.0042	0.94	1.02	0.60	1.72
A04	偽膜性大腸炎	オランザピン	1.75	706	1.0148	51.79	0.73	0.63	0.85
A41	敗血症	ジアゼパム	5.23	2,050	1.0105	1571.40	1.50	1.35	1.66
A41	敗血症	ドネペジル	14.57	53	1.0344	147.32	2.11	1.16	3.85
B16	肝感染	アリピプラゾール	2.00	194	1.0024	22.37	0.95	0.71	1.28
B16	肝感染	フロセミド	11.20	1,284	1.0254	2618.68	0.67	0.60	0.76
B25	日和見感染	シロスタゾール	7.14	41	1.0040	48.41	1.18	0.60	2.34
B25	日和見感染	パロキセチン	1.86	37	1.0006	3.91	0.67	0.35	1.28
D59	溶血性障害	バレニクリン	1.82	18	1.0003	2.00	1.14	0.44	2.93
D61	造血障害による 2 種以上の 血球減少症	バレニクリン	1.42	18	1.0002	0.87	0.97	0.38	2.45
D69	造血障害による血小板減少症	アフリベルセプト	2.63	78	1.0153	18.31	1.14	0.72	1.80
D70	無顆粒球症	A型ボツリヌス毒素	2.12	27	1.0044	4.20	1.97	0.90	4.31
D70	無顆粒球症	インスリン グラルギン	7.66	336	1.0268	426.41	1.07	0.85	1.34
D70	造血障害による白血球減少症	ラニビズマブ	1.93	16	1.0037	2.13	0.86	0.32	2.31

E03	甲状腺機能低下症	イリノテカン	3.63	96	1.0268	41.36	0.97	0.65	1.45
E03	甲状腺機能低下症	レボフロキサシン	1.79	18,725	1.0079	1724.19	1.14	1.11	1.18
E05	甲状腺機能亢進症	オキサリプラチン	1.90	57	1.0073	6.14	0.43	0.23	0.79
E05	甲状腺機能亢進症	セレコキシブ	2.00	5,987	1.0081	697.06	0.97	0.92	1.02
E16	低血糖	パルボシクリブ	1.54	3	1.0033	0.51	0		
E16	低血糖	ペメトレキセド	1.81	12	1.0050	1.44	1.56	0.46	5.34
E78	脂質異常症	セフトリアキソン	1.61	36,028	1.0612	2062.27	0.95	0.93	0.97
E87	低ナトリウム血症/ SIADH	アフリベルセプト	5.61	186	1.0502	156.07	1.58	1.18	2.12
E88	リポジストロフィー	クエチアピン	6.81	282	1.0159	306.81	0.75	0.58	0.98
E88	リポジストロフィー	シタグリプチン	6.49	663	1.0150	676.23	1.10	0.93	1.29
E88	腫瘍崩壊症候群	ラモトリギン	3.98	93	1.0081	46.43	0.92	0.61	1.38
E88	腫瘍崩壊症候群	リバビリン	13.13	39	1.0339	96.60	0.48	0.25	0.92
F19	薬物乱用および依存	シンバスタチン	2.44	5	1.0005	1.25	1.12	0.19	6.70
F19	薬物乱用および依存	ベンダムスチン		0					
F32	うつ病（自殺／自傷を除く）	リバーロキサバン	1.76	391	1.0284	30.22	1.05	0.86	1.29
F38	敵意／攻撃性	アスピリン	1.08	1	1.0000	0.22	0		
F38	敵意／攻撃性	アレンドロン酸		0					
G03	非感染性髄膜炎	オキサリプラチン	0.91	2	0.9999	0.19	1.68	0.11	26.94
G04	非感染性脳炎	インスリン グラルギン	8.93	27	1.0021	42.63	0.73	0.31	1.73
G21	パーキンソン様事象	ボルテゾミブ		0					

G21	悪性症候群	アダリムマブ	2.08	4	1.0012	0.89	0		
G24	ジスキネジア	リツキシマブ	1.91	6	1.0014	1.00	0.60	0.11	3.28
G24	ジスキネジア	リバーロキサバン	2.57	25	1.0024	6.03	1.23	0.56	2.69
G24	ジストニア	バンコマイシン	2.11	23	1.0017	3.61	0.74	0.32	1.72
G62	末梢性ニューロパチー	エナラプリル	2.30	1,351	1.0854	230.73	0.79	0.71	0.88
G62	末梢性ニューロパチー	ボセンタン	1.59	16	1.0373	1.28	0.32	0.10	1.01
H02	眼球運動障害	レナリドミド	2.70	16	1.0336	4.53	0.62	0.22	1.80
H10	結膜障害	インスリン グラルギン	1.00	3,248	1.0017	0.11	0.91	0.84	0.97
H26	水晶体障害	エベロリムス	5.40	66	1.1271	54.00	1.07	0.65	1.75
H26	水晶体障害	スルファメトキサゾール・トリメトプリム	3.55	1,995	1.0700	820.23	1.21	1.11	1.33
H35	網膜障害	タゾバクタム・ピペラシリン	2.69	1,255	1.0517	301.25	0.85	0.76	0.96
H40	緑内障	イリノテカン	1.27	141	1.0117	2.41	0.33	0.23	0.48
H40	緑内障	ミコフェノール酸	4.92	234	1.2035	166.98	0.87	0.66	1.14
H44	眼感染	アロプリノール	2.86	41	1.0007	11.79	0.70	0.38	1.30
H46	視神経障害	メトホルミン	2.64	131	1.0016	30.43	1.22	0.85	1.74
H91	聴覚障害	メトホルミン	1.48	1,222	1.0078	44.07	0.92	0.82	1.04
I20	その他の虚血性心疾患	プロピルチオウラシル	2.90	116	1.0338	32.99	0.78	0.53	1.16
I21	心筋梗塞	イソニアジド	3.06	16	1.0043	5.63	1.38	0.50	3.79
I26	静脈の塞栓および血栓	バラシクロビル	1.51	240	1.0003	10.24	1.10	0.85	1.42

I27	肺高血圧症	レボフロキサシン	1.49	502	1.0002	22.29	1.29	1.07	1.54
I42	心筋症	イミペネム・シラスタチン	3.41	20	1.0024	8.28	1.24	0.50	3.11
I48	上室性頻脈性不整脈	エタンブトール	3.16	27	1.0091	9.59	0.78	0.35	1.78
I61	出血性中枢神経系血管障害	ハロペリドール	16.92	298	1.0192	974.94	0.49	0.36	0.67
I63	虚血性中枢神経系血管障害	オセルタミビル	0.80	4,722	0.9986	59.58	1.23	1.16	1.31
I80	血栓性静脈炎	レボフロキサシン	1.98	9,526	1.0045	1225.67	1.19	1.14	1.25
J18	感染性肺炎	インダパミド	1.05	403	1.0026	0.53	1.07	0.88	1.30
J84	間質性肺疾患	ドロスピレノン・エチニルエストラジオール	0.72	23	0.9995	0.94	3.50	1.44	8.52
J96	呼吸不全	ドロスピレノン・エチニルエストラジオール	1.42	449	1.0084	13.58	1.35	1.12	1.63
K10	骨壊死	アリピプラゾール	2.32	30	1.0004	5.75	1.45	0.70	3.01
K12	口腔咽頭の状態（新生物、感染およびアレルギーを除く）	ボセンタン	2.28	24	1.0886	4.73	1.17	0.52	2.60
K56	消化管の閉塞	デノスマブ	5.40	116	1.0211	92.27	1.87	1.29	2.71
K56	消化管の閉塞	ラミブジン	3.14	1	1.0102	0.56			
K63	消化管の穿孔	リスペリドン	1.24	727	1.0082	8.41	0.72	0.62	0.83
K80	胆石が関連する障害	クラリスロマイシン	1.25	24,737	1.0026	398.24	1.34	1.30	1.37
K80	胆石が関連する障害	リバーロキサバン	4.46	288	1.0368	171.24	1.02	0.80	1.30
K83	胆道障害	トリアムシノロンアセトニド	2.29	570	1.0028	92.96	1.44	1.22	1.70
K92	消化管の出血	A型ボツリヌス毒素	2.22	135	1.0238	21.18	0.84	0.60	1.19

K92	消化管の出血	ヒドロキシクロロキン	1.30	21	1.0057	0.65	0.51	0.15	1.75
M31	血管炎	ドネペジル	6.17	4	1.0023	2.36	2.08	0.22	20.04
M32	全身性エリテマトーデス	ゲンタマイシン	2.10	74	1.0014	10.29	0.80	0.50	1.29
M32	全身性エリテマトーデス	ミルタザピン	3.35	149	1.0030	55.06	0.67	0.48	0.93
M62	横紋筋融解症／ミオパチー	ラニビズマブ	2.97	83	1.0270	24.92	1.00	0.65	1.54
M62	横紋筋融解症／ミオパチー	ドロスピレノン・エチニルエストラジオール	2.45	531	1.0198	101.93	1.28	1.08	1.52
M77	腱障害および靭帯障害	アナグレリド	1.35	5	1.0116	0.51	2.65	0.44	15.88
M77	腱障害および靭帯障害	トラスツズマブ	1.63	104	1.0209	6.54	0.77	0.52	1.14
M81	骨粗鬆症／骨減少症	シタラビン	11.09	288	1.1787	584.84	2.45	1.89	3.18
N12	尿細管間質性疾患	アリピプラゾール	2.05	432	1.0056	52.40	1.09	0.90	1.32
N17	急性腎不全	エスシタロプラム	1.87	85	1.0008	8.39	1.36	0.88	2.11
N39	蛋白尿	ジアゼパム	2.36	7,031	1.0265	1252.26	1.16	1.10	1.22
R43	味覚および嗅覚障害	トリアムシノロンアセトニド	2.38	1,411	1.0071	252.38	0.95	0.86	1.06
R44	精神病および精神病性障害	アフリベルセプト	0.63	23	0.9957	1.64	0.86	0.37	1.98
R44	精神病および精神病性障害	オクトレオチド	0.84	13	0.9981	0.29	0.14	0.03	0.63
R56	痙攣	アナグレリド		0					
R73	高血糖／糖尿病の発症	アモキシシリン	1.08	8,798	1.0007	14.23	1.05	1.00	1.09

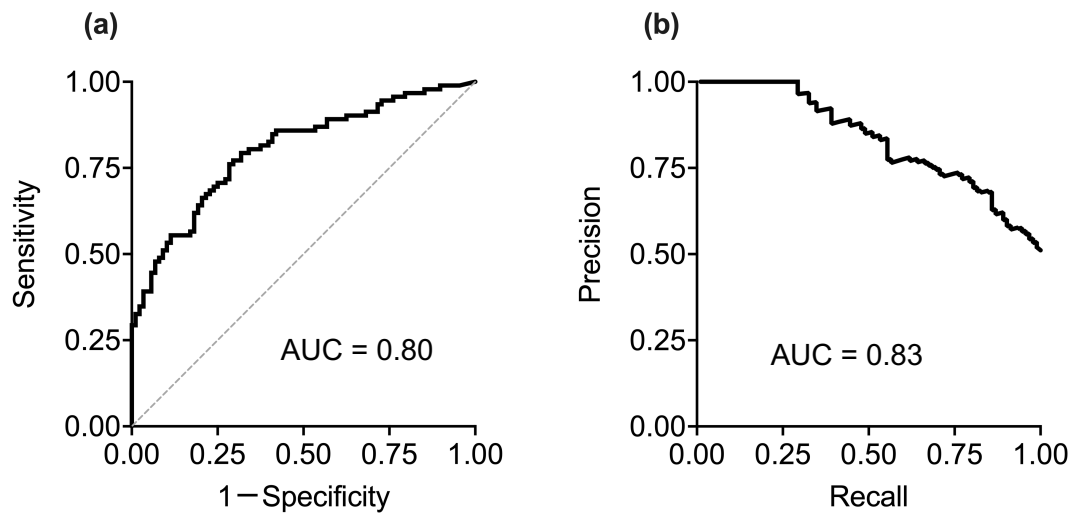


図 2-1 副作用シグナル検出における ARM の性能の視覚化

2005 年から 2019 年までの 15 年間のデータに ARM を適応させ、lift 値を用いて (a) ROC 曲線および (b) PR 曲線により性能を視覚化した。

表 2-3 lift の閾値を変化させた場合の ARM による副作用シグナルの検出性能

Lift	Sensitivity	Specificity	Precision	F-measure
1	0.98	0.25	0.58	0.73
2	0.87	0.49	0.64	0.74
3	0.75	0.73	0.74	0.75
4	0.64	0.81	0.78	0.70
5	0.57	0.83	0.78	0.65

2. ARM の性能の再現性

次に ARM の性能の再現性を検証するため、レセプトデータを加入者の登録開始時期に応じて分割することで異なる母集団を想定したデータセットを準備した (表 2-4)。各データセットに対し ARM を適応させ、lift 値を閾値として性能を視覚化したところ、ROC-AUC は 0.71–0.77、PR-AUC は 0.74–0.81 を示した (図 2-2 a, b)。全てのデータセットで ROC-、PR-AUC とも 0.7 を超えており、ARM は適度な性能を再現可能に達成できる可能性が示された。

表 2-4 レセプトデータを各個人の加入時期に応じて分割したデータセットの症例数およびレコード数

No	Dataset name	Enrollment Start	No. of patients	No. of records
1	2008	from Jan. 2008 to Dec. 2008	122,609	5,288,581
2	2009	from Jan. 2009 to Dec. 2009	209,257	8,288,921
3	2010	from Jan. 2010 to Dec. 2010	428,086	16,888,695
4	2011	from Jan. 2011 to Dec. 2011	330,718	12,466,478
5	2012	from Jan. 2012 to Dec. 2012	486,751	17,095,830
6	2013	from Jan. 2013 to Dec. 2013	1,038,966	33,786,660
7	2014	from Jan. 2014 to Dec. 2014	297,147	8,717,316
8	2015	from Jan. 2015 to Dec. 2015	1,131,638	30,602,401
9	2016	from Jan. 2016 to Dec. 2016	738,270	15,924,137
10	2017	from Jan. 2017 to Dec. 2017	851,522	14,879,488

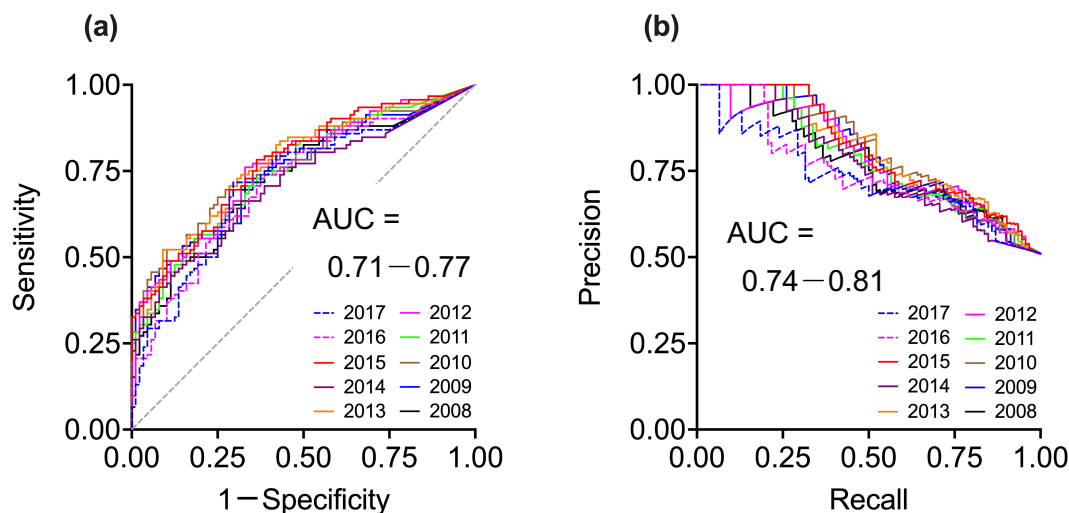


図 2-2 ARM の副作用シグナル検出性能の再現性

15 年間のレセプトデータを 10 分割したデータセットを準備し、各データセットに ARM を適応させて (a) ROC 曲線および (b) PR 曲線により性能を視覚化した。各データセットの加入者数およびレコード数は表 2-4 に示した。

3. 副作用シグナルの早期検出

1.と 2.の検討において ARM のモデルの性能について評価を行ったので、早期検出条件下 (短期間のデータ) における ARM の有用性を評価した。2018 年 1 月の記録を開始点として 1 ヶ月分のデータ、3 ヶ月分のデータ、6 ヶ月分のデータと解析期間を徐々に延長させたデータセットを準備し ARM と SSA を適応させた。今回の検討においては、途中から加入する患者の影響を排除するため、1 ヶ月のデータセットで対象にした加入者 (1,337,370 人) のみを 6 ヶ月間追跡した。レコード数の減少を可能な限り抑えるため、Run-in period は 3 ヶ月とした。本検討では、ARM のシグナル検出基準を $lift > 1$ 、 $conviction > 1$ 、 $p\text{-value} < 0.05$ とした。

各データセットにおける ARM と SSA のシグナル検出数を図 2-2 に示した。1 ヶ月のデータセットにおいて ARM は感度 0.38 を示したが (35 組の陽性対照を検出)、SSA は適用されなかった。これは JMDC Claims が一月単位でレコードが纏められているため、順序を考慮できなかったためである (図 2-2 a)。3 ヶ月のデータセットにおいて、ARM は感度 0.57 を示し (追加で 17 組の陽性対照を検出)、SSA は感度 0.05 を示した (5 組の陽性対照を検出)。SSA で検出された計 5 組の陽性対照は ARM で検出された 52 組に含まれていた。6 ヶ月のデータ

セットにおいて、ARM は感度 0.71 を示し（追加で 14 組の陽性対照を検出したが、1 組が検出されなくなった）、SSA は感度 0.1 を示した（追加で 5 組の陽性対照を検出したが、1 組が検出されなくなった）。SSA で検出された計 9 組の陽性対照は、ARM で検出された 65 組に含まれていた。以上から、ARM はベースラインの手法である SSA に比べてより多くの副作用シグナルを早期に検出できることが確認された。

また、前章で構築した副作用の発症時期情報を用いて新規シグナルの検出数を比較した。発現時期を 2 つのカテゴリー（short: Onset \leq 90、long: Onset $>$ 90）に分類したところ、ARM は 1 ヶ月のデータセットで short カテゴリーの陽性対照を複数検出し、short および long どちらのカテゴリーにおいても SSA より多くのペアを検出していることが確認された（図 2-2 b）。個別の組合せを確認すると「バンコマイシンー急性腎障害：8 日」など早期に発生すると考え得るペアは、1 ヶ月のデータセットでシグナルが生成され、遅発的に発生しやすいと考えられる「リツキシマブー肝感染：165 日」も早期に有害事象を発生した症例から関連性が算出され、1 ヶ月のデータでシグナルが検出されていた。一方で、「ロクロニウムーアナフィラキシー：1 日」はどの解析期間においてもシグナルは生成されなかった。アナフィラキシーはどの医薬品でも起きうる副作用のためシグナルとして強く検出されなかった可能性や、その他にもアナフィラキシーを起こしうる医薬品と月単位で情報が纏められるレセプトの限界点も考えられた。

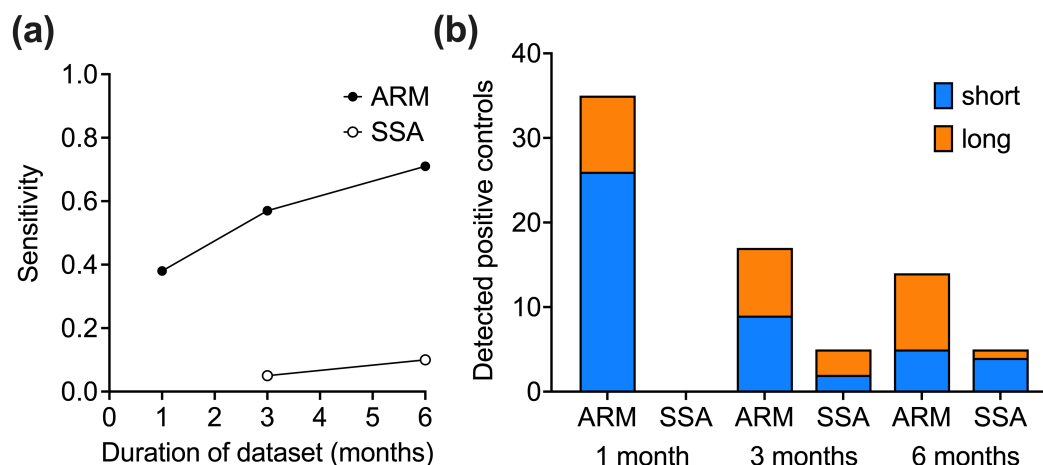


図 2-2 早期検出における ARM と SSA の副作用シグナル検出能の比較

(a) 1 ヶ月、3 ヶ月、6 ヶ月の各期間における ARM と SSA の感度の比較 (b) 副作用の発症時期情報を用いた、新規シグナルの検出数の比較 (short: Onset \leq 90、long: Onset $>$ 90)

SSA は自己対照デザインの一つであり、医薬品の使用とイベント発生の順序を考慮し、かつ時間非依存の交絡を最小にする点で ARM より優れた手法である [60]。しかし、SSA は医薬品とイベントの両方が登録される症例のみを解析対象とするため、データ数が少ない場合など短期間でのシグナル検出は困難になると考えられ、この傾向は既報と一致していた [26]。一方、ARM は順序を考慮せず、また患者背景などの交絡因子も考慮していないため、偽陽性シグナルが多くなる可能性が高い [61]。そこで感度以外の指標である特異度や適合率を算出し、ARM と SSA の性能を比較した (図 2-3) その結果、ARM は SSA に比べ、特異度や適合率は低い値を示すものの、0.7 程度の値を維持したまま高い感度を示すことが明らかになった。短期間のデータでは、ゴールドスタンダードに含まれるいくつかの「医薬品-有害事象」のペアをもつ症例が含まれず、共起確率が算出できなかったため特異度や適合率が維持された可能性がある。解析期間が長くなればなるほど、余分なペアが作られ偽陽性の発生率は上昇すると予想される。以上の欠点はあるものの、ARM はより短いデータセットにおいて副作用シグナルを早期に検出できる可能性があることが示された。

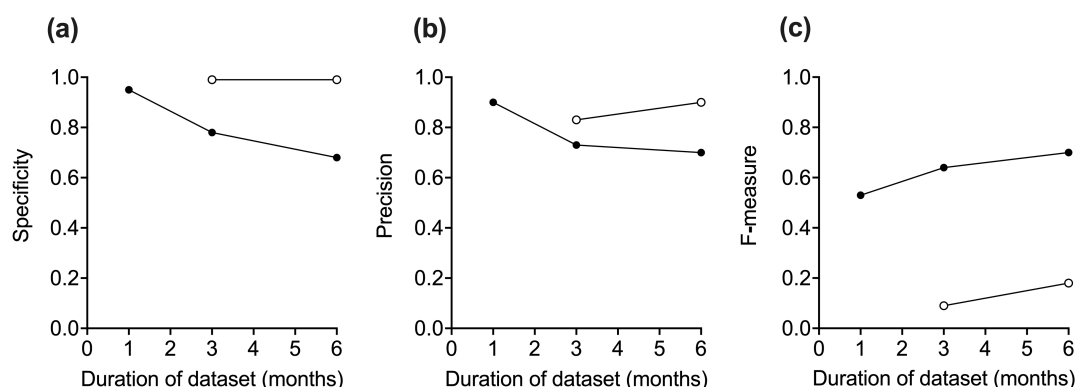


図 2-3 早期検出における ARM と SSA の特異度、適合率、F-measure

1ヶ月、3ヶ月、6ヶ月の期間における ARM と SSA の (a) 特異度、(b) 適合率、(c) F-measure

4. 安全性ラベルの変更

次に、より現実的なシミュレーションを仮定して ARM が新薬の副作用シグナル早期に検出できるか検討を行った。前章で構築したゴールドスタンダードのうち「ドロスピレノン・エチニルエストラジオール-血栓性静脈炎」の陽性対照に着目した。月経困難症治療剤のドロスピレノン・エチニルエストラジオールは販売開始後 3 年ほどで PMDA から血栓症の有害事象に関する安全性速報が発令されている。本検討では、2010 年 11 月（販売開始月）から 2013 年 11 月（安全性速報発令の前月）までのデータセットを準備し、ARM が事前に安全性シグナルを検出するか検討を行った。ARM のシグナル検出基準を $\text{lift} > 1$ 、 $\text{conviction} > 1$ 、 $p\text{-value} < 0.05$ とした。

その結果、販売 39 ヶ月後に発令された安全性速報より 25 ヶ月前の時点（承認から 14 ヶ月後）に血栓症シグナルが検出された（図 2-3）。解析では Run-in period を 3 ヶ月に設定しており、新規処方患者とイベント発症者を対象とした。ARM は順序配列を考慮していないが、短期間のデータセットにおいては安全性情報発令を補完し、新薬の安全性を検討する上で非常に強力なツールとなり得ることが示された。

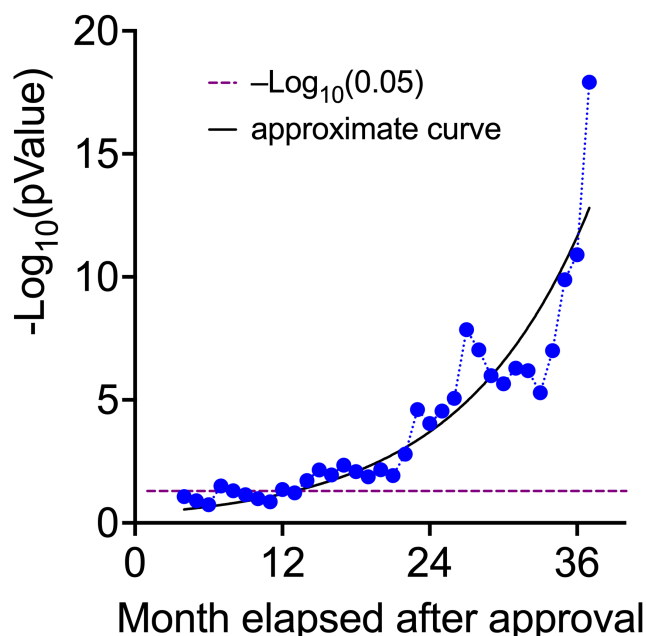


図 2-3 承認直後の医薬品を用いた ARM による安全性シグナルの検出

「ドロスピレノン・エチニルエストラジオール-血栓性静脈炎」に着目。lift や conviction の値は早期から 1 を超えていたため、 $p\text{-value}$ の揭示変化をプロットした。近似曲線式は非線形回帰により算出した。

5. 既存のゴールドスタンダードを用いた検証

これまで得られた知見を検証するために、既存のゴールドスタンダードを用いて同様の検討を行った。Ryan らが提案したゴールドスタンダードは、ファーマコビジランス活動に不可欠な4つの副作用（急性心筋梗塞、急性肝障害、消化管出血、急性腎障害）を含む [32]。このうち国内で承認される医薬品で、かつエビデンスレベルが高いとされる「医薬品－有害事象」ペア（陽性対照 61 組と陰性対照 39 組）のみを抽出し、ARM の性能を評価した。4つの有害事象は以下の ICD10 コードに対応させた（急性心筋梗塞：I21、急性肝障害：K71、消化管出血：K92、急性腎障害：N17）。

2005 年から 2019 年までの全データセットにおいて $lift > 1$ 、 $conviction > 1$ 、 $p\text{-value} < 0.05$ を検出基準とした場合、ARM では 85 組のシグナルが同定され、そのうち 54 組が真の陽性対照であった（感度：0.89、特異度：0.21、精度：0.64、F-measure：0.74）。SSA では 23 組のシグナルが同定され、そのうち 15 組が真の陽性対照であった（感度：0.25、特異度：0.79、精度：0.65、F-measure：0.36）。両手法ともでシグナルとして検出されたのは 21 組で、そのうち 14 組が真の陽性対照であった。lift 値を閾値として ARM の性能を視覚化したところ、ROC-AUC は 0.65、PR-AUC は 0.78 を示した（図 2-4）。既存のゴールドスタンダードを用いた場合、第 1 章で構築したゴールドスタンダードで評価した場合に比べて AUC 値は低い傾向にあった。この背景として 4 種類の有害事象に対応させた ICD10 コードをもつ症例数がレセプトデータにおいて相対的に少ない傾向があり、モデルの性能に影響を与えた可能性があった。しかし、lift 値が上位のペアは陽性対照が大部分を占めており、閾値を調整することでバランスのとれた性能を得られる可能性が示された。

早期検出における解析デザインでは、ARM は 1 ヶ月のデータセットにおいて感度 0.13 を示した（8 組の陽性対照を検出）。3 ヶ月のデータセットでは感度 0.28 を示した（追加で 10 組の陽性対照を検出したが、1 組が検出されなくなった）。6 ヶ月のデータセットでは感度 0.43 を示した（追加で 10 組の陽性対照を検出したが、1 組が検出されなくなった）。一方、SSA は 1 ヶ月、3 ヶ月、6 ヶ月のどの期間においても、いずれの陽性対照もシグナルとして検出できなかった。蓄積期間の短いデータにおいて、ARM は SSA よりも多くの陽性対照シグナルを検出していた。以上、既存のゴールドスタンダードを用いた検討により、これまで得られた知見をより強固なものにできたと考えられる。

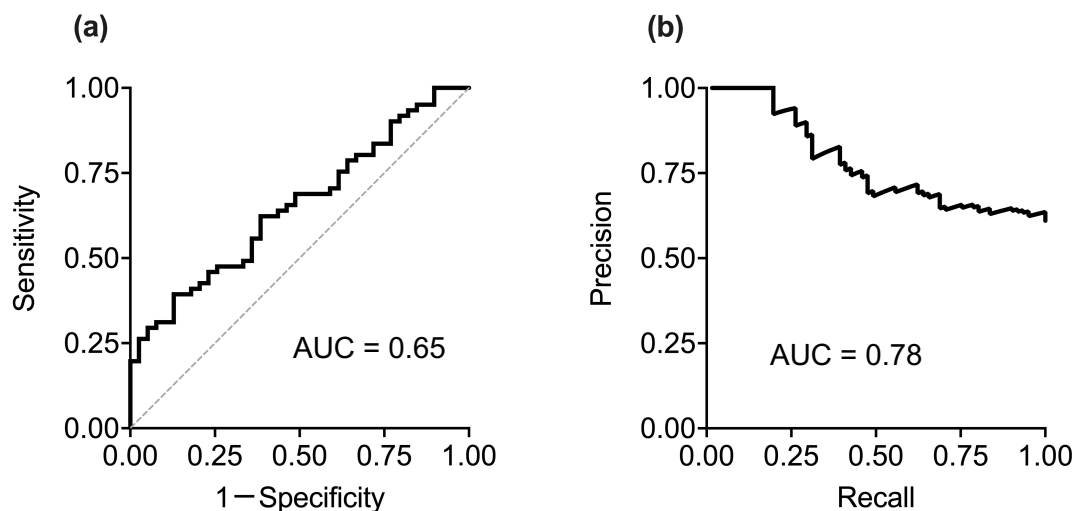


図 2-4 既存のゴールドスタンダードによる ARM の性能評価

2005 年から 2019 年までの 15 年間のデータに ARM を適応させ、lift 値を用いて (a) ROC 曲線および (b) PR 曲線により性能を視覚化した。

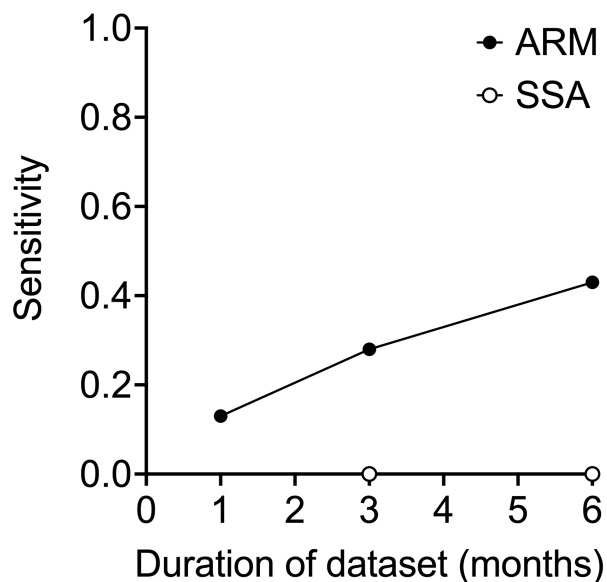


図 2-5 既存のゴールドスタンダードを用いた早期検出における ARM と SSA の副作用シグナル検出能の比較

1 ヶ月、3 ヶ月、6 ヶ月の各期間での ARM と SSA の感度を比較した。

本研究の貢献と限界

本研究は、大規模レセプトデータを用いて広範な副作用シグナルを早期に検出するアプローチを見出した初めての検討である。第1章で臨床エビデンスに基づくゴールドスタンダードを構築し、第2章ではそれを用いて副作用シグナルの早期検出にレセプトへのARMが有用であることを見出すことで、既存のファーマコビジランスを強化できることを示した。レセプトや電子カルテなどの電子診療情報を活用して副作用シグナルを検出する研究はこれまでも行われている。しかし、大部分は特定の副作用を対象としていたため、本検討では包括的かつ定量的に安全性シグナルを評価した。近年、薬理学的標的などの情報を用いて、新薬の副作用シグナルを検出する検討などが報告されている [62, 63]。これらのモデルは高い性能を示す結果が得られているものの、**first-in-class**の医薬品を対象にできない限界点があった。レセプトへのARMは上記の制限はなく、実際、我々が構築したゴールドスタンダードには**first-in-class**の医薬品として国内で2017年に承認されたCDK4/6阻害剤パルボシクリブが含まれており、ARMはこの医薬品の副作用シグナルを検出していた（表2-1）。

レセプトデータへのARMの適応は市販後の安全性に関する懸念事項を早期に特定することを可能にし、安全性に関するコミュニケーションを加速させる可能性を持っている。しかし、ARMによって検出される副作用シグナルは専門家による臨床レビューに取って代わることはできない。本手法の実用性の点については複数の限界点が考えられ、レセプトデータの活用とARM適応の2点に分けて論じる。第一に、レセプトデータの症状名には対象とする医薬品の適応症や、医薬品使用と関係は無くとも原疾患に併発・付随する症状も含まれるため、必ずしも有害事象に対応しない症状が解析される。この影響を減らすためには、少なくともATCコードとICD10コードから同じ臓器に属することが示唆される医薬品と疾患（有害事象）のペアは解析前に除外する必要がある。また、レセプトデータは管理・請求の目的で構築されるため、正確な情報を有していない可能性があり（レセプト病名など）、この点の影響を機械的に減らすことは困難な可能性がある。第二に、レセプトデータを副作用シグナル検出に用いる場合、有害事象と病名との語彙の違いに対処する作業が必要となる。有害事象名とICD10コードのマッピングは、UMLSによりある程度確立されている。しかしUMLSを用いた場合でも、有害事象名をICD10へ体系的にマッピングすることは困難な点がある [40-42]。本検討では各SMQを手作業によりICD10へマッピングを行ったが、マッピングが不完全であるために一部の患者情報が失われる可能性がある。これらは、本研究におけるシグナル検出の性能に影響を与えた可能性がある。第三に、日本では診療報酬は翌月10日

までに請求される。さらに、請求データがデータベースとして活用できるまでに数ヶ月を要する。今後、データの利活用が円滑になった場合でも、最大1ヶ月のタイムラグがあり、承認後1ヶ月未満の副作用シグナルをレセプトデータのみを用いて検出することは困難である可能性がある。

その他、ARMの解析手法自体への限界点も挙げられる。ARMは順序分析ではないため、本検討では関連性の強さに基づいて因果関係を推定した。ARMにおける偽陽性シグナルは以下のような状況で発生すると思われる。逆順のペア（医薬品→イベントの登録患者、イベント→医薬品の患者）が存在しない場合、SSAの解析値は0またはnullとなり、評価できないため、シグナルとして検出されない。一方、ARMでは必ずしも医薬品→イベントのペアでなくても共起確率を計算してしまうため、偽陽性を招いてしまう。第二に、因果の逆転、時間依存交絡、相互作用などの交絡因子の影響が避けられない [22, 64]。特にARMでは複数の医薬品が併用されている場合、どの医薬品が被疑薬であるかを区別することは困難である。さらに、2つの症状との間に強い併存性がある場合、どちらの症状が有害事象であるかの区別は困難である。第三に、偽陽性を減らすためにlift値の閾値を高く設定した場合でも、必ずしも副作用シグナルの可能性である、とは考察できない点である。医薬品の使用頻度が低く、また有害事象も稀にしか発生しない場合、対象とする医薬品と有害事象のペアが偶然にも共起した場合、liftは非常に高くなる [65]。このような極端に高い「医薬品-有害事象」ペアの共起を排除するためには、ARMのsupportの閾値を高くする検討等が考えられる。しかし、実際には、ARMだけでなく、他の分析手法を組み合わせ、最終的には偽陽性シグナルを臨床家などの手によって減らすプロセスは必要である。

上記の限界点があるものの、本検討で得られた知見は重要な副作用検出の遅れを減らすことを示しており、SRSの補完ツールとして有用である可能性を示している。

総括

本研究では、複数の大規模臨床データを活用することにより、以下の新発見を得た。

第1章では、臨床エビデンスに基づく新たなゴールドスタンダードを構築した。本ゴールドスタンダードの特徴を以下に示す。

- 1) 180組（陽性対照：92組、陰性対照：88組）で構成される
- 2) 広範な医薬品（78種類）と関連する有害事象（92種類）を含む
- 3) 実臨床における副作用発症時期の情報を含む

本ゴールドスタンダードを活用することで、広範な医薬品とその副作用を対象にデータマイニング手法の性能を評価することを可能にした。

第2章では以下3つの検討結果より、副作用シグナルの早期検出にレセプトデータへのARMが有用であることを示した。

- 1) 適度な性能とその再現性
- 2) 既存手法との比較による早期検出における有用性
- 3) 新薬を用いた安全性シグナルの検出

レセプトデータへのARMはSRSを補完し、既存のファーマコビジランスを強化できることを見出した。

将来の展望

本アプローチにより既存のリファレンスがない「医薬品－有害事象」が強いシグナル（高い lift や conviction 値）として生成された場合、それらのルールは十分に知られていない副作用として有望である可能性がある。また本検討は医薬品と有害事象の 2 項目のみの関連性に留まったものであるが、将来的には複数要素のルール（前提条件または結果のいくつかの項目）を作成することで患者背景などを考慮した安全性の評価を加速させることが可能と考える。しかし、本アプローチのみでは様々な限界が含まれるため、生成されるノイズを減らし分析精度を向上させるものとして以下の 3 点について論じる。

1. データ基盤の構築

複数の臨床データを組み合わせることで安全性シグナルの検出性能は向上することが報告され [17, 18, 66]、患者プライバシーに配慮した臨床データの共有化も報告されている [67]。請求目的のレセプトの解析においても、電子カルテ情報に含まれる検査値や処置データを活用することで医薬品有害事象に対応しない症状を削ぎ落とせる可能性があり、ARM を含めた解析手法による偽陽性シグナルの低減化につながる可能性がある。

2. ゴールドスタンダードの構築

本検討ではレセプトデータのみを用いたが、電子カルテやその他、インターネットの検索履歴 [68] など様々なデータに複雑な解析法を当てはめた研究が報告される。ARM を含めた機械学習や AI は専門家の知識ベースを越える情報を与え得る可能性があるが、データソースに関係なく解析法の性能を適切に評価するための信頼できる十分に大きな参照基準は必要である。本検討では計 180 組のゴールドスタンダードを構築したが、例えば構築プロセス自体を人工知能に学習させることで、より多くのペアで構成されるゴールドスタンダードを作成することも可能であり、それにより性能の高い解析手法を決定することができれば、安全性情報の発信や警告の追加などのタイムリーな措置に繋げることができると考える。

3. 医療用語の標準化および対応付け

電子カルテ情報を例にしても、これらのデータは研究目的ではないため、副作用情報の記載は統一されていない部分があり複雑である。医療用語の統一は大きな課題であり、情報の損失はあっても最小限に抑えながら、異なるデータソースからの情報を共通のモデルに変換するための体系的なプロセスの開発が求められる。

将来的には複数の臨床データをオンサイトで機械的に解析を行い、安全性シグナルを適時に検出することで医薬品副作用の早期発見に繋げたい。本研究成果は、今後利活用が進む大規模臨床データを用いた副作用予測システムや未知の副作用検出の基盤となることが期待される。

謝辞

本研究に際しまして、終始懇切丁寧な御指導および御鞭撻を賜りました京都大学大学院薬学研究科教授 金子周司 先生に謹んで感謝の意を表します。また、直接御指導を賜り、多くの有益な御助言をいただきました京都大学大学院薬学研究科准教授 白川久志 先生、助教 永安一樹 先生に心から感謝の意を表します。

本研究の遂行に多大なる御協力をいただきました長島卓也 博士、Soni Siswanto 博士、酒井幸 博士、長岡巧樹 博士、澁井紀宏 学士、古田晴香 学士、戸田千尋 学士、栢沼玄 学士、岩井信太郎 君、鈴木崇英 君に心から感謝いたします。そして、日々の研究生活において互いに切磋琢磨し合えた京都大学大学院薬学研究科生体機能解析学分野の諸氏に深く感謝いたします。

また、本研究の遂行に関しまして藤多仁生奨学金をいただきました故藤多哲郎 先生、およびご遺族・関係者に深く感謝いたします。

最後に、著者の良き理解者であり博士号取得に向けて背中を押し、常に支えてくださいました家族に深く感謝します。

発表論文目録

Early detection of adverse drug reaction signals by association rule mining using large-scale administrative claims data

(大規模レセプトデータを用いた相関ルールマイニングによる医薬品副作用シグナルの早期検出)

Hiroki Yamamoto, Gen Kayanuma, Takuya Nagashima, Chihiro Toda, Kazuki Nagayasu, Shuji Kaneko

Drug Safety. **in press** 2023 年 2 月

参考文献

1. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA*. 1998; 279: 1200–1205.
2. FDA (2018). Preventable Adverse Drug Reactions: A Focus on Drug Interactions. <https://www.fda.gov/drugs/drug-interactions-labeling/preventable-adverse-drug-reactions-focus-drug-interactions>. Accessed 22 August 2022.
3. Huybrechts KF, Desai RJ, Park M, Gagne JJ, Najafzadeh M, Avorn J. The Potential Return on Public Investment in Detecting Adverse Drug Effects. *Med Care*. 2017; 55: 545–551.
4. Wysowski DK, Swartz L. Adverse drug event surveillance and drug withdrawals in the United States, 1969-2002: the importance of reporting suspected reactions. *Arch Intern Med*. 2005; 165: 1363–1369.
5. Downing NS, Shah ND, Aminawung JA, Pease AM, Zeitoun JD, Krumholz HM, et al. Postmarket Safety Events Among Novel Therapeutics Approved by the US Food and Drug Administration Between 2001 and 2010. *JAMA*. 2017; 317: 1854–1863.
6. Rogers AS. Adverse drug events: identification and attribution. *Drug Intell Clin Pharm*. 1987; 21: 915–920.
7. Berlin JA, Glasser SC, Ellenberg SS. Adverse event detection in drug development: recommendations and obligations beyond phase 3. *Am J Public Health*. 2008; 98:1366–1371.
8. Sakaeda T, Tamon A, Kadoyama K, Okuno Y. Data mining of the public version of the FDA Adverse Event Reporting System. *Int J Med Sci*. 2013; 10: 796–803.
9. Chasioti D, Yao X, Zhang P, Lerner S, Quinney SK, Ning X, et al. Mining Directional Drug Interaction Effects on Myopathy Using the FAERS Database. *IEEE J Biomed Health Inform*. 2019; 23: 2156–2163.
10. Sarangdhar M, Tabar S, Schmidt C, Kushwaha A, Shah K, Dahlquist JE, et al. Data mining differential clinical outcomes associated with drug regimens using adverse event reporting data. *Nat Biotechnol*. 2016; 34: 697–700.
11. Gibbons RD, Amatya AK, Brown CH, Hur K, Marcus SM, Bhaumik DK, et al. Post-approval drug safety surveillance. *Annu Rev Public Health*. 2010; 31: 419–437.
12. Kaneko S, Nagashima T. Drug Repositioning and Target Finding Based on Clinical Evidence. *Biol Pharm Bull*. 2020; 43: 362–365.
13. Hazell L, Shakir SA. Under-reporting of adverse drug reactions. a systematic review. *Drug Saf*. 2006; 29: 385–396.

14. Alatawi YM, Hansen RA. Empirical estimation of under-reporting in the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). *Expert Opin Drug Saf.* 2017 ; 16: 761–767.
15. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther.* 1981; 30: 239–245.
16. Gallagher RM, Kirkham JJ, Mason JR, Bird KA, Williamson PR, Nunn AJ, et al. Development and inter-rater reliability of the Liverpool adverse drug reaction causality assessment tool. *PLoS One.* 2011; 6: e28096.
17. Harpaz R, Vilar S, Dumouchel W, Salmasian H, Haerian K, Shah NH, et al. Combining signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc.* 2013; 20: 413–419.
18. Li Y, Ryan PB, Wei Y, Friedman CA. Method to Combine Signals from Spontaneous Reporting Systems and Observational Healthcare Data to Detect Adverse Drug Reactions. *Drug Saf.* 2015; 38: 895–908.
19. Wang L, Rastegar-Mojarad M, Ji Z, Liu S, Liu K, Moon S, et al. Detecting Pharmacovigilance Signals Combining Electronic Medical Records With Spontaneous Reports: A Case Study of Conventional Disease-Modifying Antirheumatic Drugs for Rheumatoid Arthritis. 2018; *Front Pharmacol.* 9: 875.
20. West SL, Johnson W, Visscher W, Kluckman M, Qin Y, Larsen A. The challenges of linking health insurer claims with electronic medical records. *Health Informatics J.* 2014; 20: 22–34.
21. Lai EC, Pratt N, Hsieh CY, Lin SJ, Pottegård A, Roughead EE, et al. Sequence symmetry analysis in pharmacovigilance and pharmacoepidemiologic studies. *Eur J Epidemiol.* 2017; 32: 567–582.
22. Hallas J, Wang SV, Gagne JJ, Schneeweiss S, Pratt N, Pottegård A. Hypothesis-free screening of large administrative databases for unsuspected drug-outcome associations. *Eur J Epidemiol.* 2018; 33: 545–555.
23. Takada M, Fujimoto M, Hosomi K. Association between Benzodiazepine Use and Dementia: Data Mining of Different Medical Databases. *Int J Med Sci.* 2016; 13: 825–834.
24. Adimadhyam S, Schumock GT, Calip GS, Smith Marsh DE, Layden BT, Lee TA. Increased risk of mycotic infections associated with sodium-glucose co-transporter 2 inhibitors: a prescription sequence symmetry analysis. *Br J Clin Pharmacol.* 2019; 85: 160–168.
25. Li XX, Cheng YC, Zhai SD, Yao P, Zhan SY, Shi LW. Risk of Liver Injury Associated

- with Intravenous Lipid Emulsions: A Prescription Sequence Symmetry Analysis. *Front Pharmacol.* 2021; 12: 589091.
26. Wahab IA, Pratt NL, Wiese MD, Kalisch LM, Roughead EE. The validity of sequence symmetry analysis (SSA) for adverse drug reaction signal detection. *Pharmacoepidemiol Drug Saf.* 2013; 22: 496–502.
 27. Zhan C, Roughead E, Liu L, Pratt N, Li J. A data-driven method to detect adverse drug events from prescription data. *J Biomed Inform.* 85: 10–20 (2018).
 28. Hoang T, Liu J, Roughead E, Pratt N, Li J. Supervised signal detection for adverse drug reactions in medication dispensing data. *Comput Methods Programs Biomed.* 2018; 161: 25–38.
 29. Zhan C, Roughead E, Liu L, Pratt N, Li J. Detecting potential signals of adverse drug events from prescription data. *Artif Intell Med.* 2020; 104: 101839.
 30. Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A Curated and Standardized Adverse Drug Event Resource to Accelerate Drug Safety Research. *Sci Data.* 2016; 3: 160026.
 31. Coloma PM, Avillach P, Salvo F, Schuemie MJ, Ferrajolo C, Pariente A, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. *Drug Saf.* 2013; 36: 13–23.
 32. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. *Drug Saf.* 2013; 36: Suppl 1: S33–47.
 33. Harpaz R, Odgers D, Gaskin G, DuMouchel W, Winnenburg R, Bodenreider O, et al. A time-indexed reference standard of adverse drug reactions. *Sci Data.* 2014; 1: 140043.
 34. Kontsioti E, Maskell S, Dutta B, Pirmohamed M. A reference set of clinically relevant adverse drug-drug interactions. *Sci Data.* 2022; 9: 72.
 35. Patadia VK, Schuemie MJ, Coloma P, Herings R, van der Lei J, Straus S, et al. Evaluating performance of electronic healthcare records and spontaneous reporting data in drug safety signal detection. *Int J Clin Pharm.* 2015; 37: 94–104.
 36. Ryan PB, Stang PE, Overhage JM, Suchard MA, Hartzema AG, DuMouchel W, et al. A comparison of the empirical performance of methods for a risk identification system. *Drug Saf.* 2013; 36: 143–158.
 37. Suchard MA, Zorych I, Simpson SE, Schuemie MJ, Ryan PB, Madigan D. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. *Drug Saf.* 2013 ; 36: 83–93.
 38. Ietswaart R, Arat S, Chen AX, Farahmand S, Kim B, DuMouchel W, et al. Machine

- learning guided association of adverse drug reactions with in vitro target-based pharmacology. *EbioMedicine*. 2020; 57: 102837.
39. MedDRA. Support Documentation. <https://www.meddra.org/how-to-use/support-documentation/japanese>. Accessed 6 January 2023.
 40. Lee S, Han J, Park RW, Kim GJ, Rim JH, Cho J, et al. Development of a Controlled Vocabulary-Based Adverse Drug Reaction Signal Dictionary for Multicenter Electronic Health Record-Based Pharmacovigilance. *Drug Saf*. 2019; 42: 657–670.
 41. Koutkias V. From Data Silos to Standardized, Linked, and FAIR Data for Pharmacovigilance: Current Advances and Challenges with Observational Healthcare Data. *Drug Saf*. 2019; 42: 583–586.
 42. Declerck G, Souvignet J, Rodrigues JM, Jaulent MC. Automatic annotation of ICD-to-MedDRA mappings with SKOS predicates. *Stud Health Technol Inform*. 2014; 205: 1013–1017.
 43. PMDA. The Yellow Letter / Blue Letter. <https://www.pmda.go.jp/english/safety/info-services/drugs/esc-rsc/0001.html>. Accessed 1 January 2023.
 44. Wakao R, Taavola H, Sandberg L, Iwasa E, Soejima S, Chandler R, et al. Data-Driven Identification of Adverse Event Reporting Patterns for Japan in VigiBase, the WHO Global Database of Individual Case Safety Reports. *Drug Saf*. 2019; 42: 1487–1498.
 45. Sonawane KB, Cheng N, Hansen RA. Serious Adverse Drug Events Reported to the FDA: Analysis of the FDA Adverse Event Reporting System 2006-2014 Database. *J Manag Care Spec Pharm*. 2018; 24: 682–690.
 46. Satake R, Matsumoto K, Tanaka M, Mukai R, Shimada K, Yoshida Y, et al. Analysis of Drug-Induced Gastrointestinal Obstruction and Perforation Using the Japanese Adverse Drug Event Report Database. *Front Pharmacol*. 2021; 12: 692292.
 47. Hasegawa S, Ikesue H, Nakao S, Shimada K, Mukai R, Tanaka M, et al. Analysis of immune-related adverse events caused by immune checkpoint inhibitors using the Japanese Adverse Drug Event Report database. *Pharmacoepidemiol Drug Saf*. 2020; 29: 1279–1294.
 48. Sauzet O, Carvajal A, Escudero A, Molokhia M, Cornelius VR. Illustration of the weibull shape parameter signal detection tool using electronic healthcare record data. *Drug Saf*. 2013; 36: 995–1006.
 49. Giannangelo K, Fenton SH. SNOMED CT survey: an assessment of implementation in EMR/EHR applications. *Perspect Health Inf Manag*. 2008; 5: 7.
 50. National Library of Medicine, US National Institutes of Health. Unified Medical

Language System. <https://www.nlm.nih.gov/research/umls/index.html>. Accessed 5 February 2023.

51. Reich C, Ryan PB, Stang PE, Rocca M. Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. *J Biomed Inform.* 2012; 45: 689–696.
52. Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules. In J. B. Bocca, M. Jarke, and C. Zaniolo, editors, Proc. 20th Int. Conf. *Very Large Data Bases, VLDB*, 1994; pages 487–499.
53. Harpaz R, Chase HS, Friedman C. Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC Bioinformatics.* 2010; 11: Suppl 9: S7.
54. Fujiwara M, Kawasaki Y, Yamada H. A Pharmacovigilance Approach for Post-Marketing in Japan Using the Japanese Adverse Drug Event Report (JADER) Database and Association Analysis. *PLoS One.* 2016; 11: e0154425.
55. Yildirim P. Association Patterns in Open Data to Explore Ciprofloxacin Adverse Events. *Appl Clin Inform.* 2015; 6: 728–747.
56. Jorge AM, Azevedo PJ. An Experiment with Association Rules and Classification: Post-Bagging and Conviction. Proceedings of the 8th international conference on Discovery Science. 2005; Pages 137–149.
57. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol.* 2005; 58: 323–337.
58. Idema DL, Wang Y, Biehl M, Horvatovich PL, Hak E. Effect estimate comparison between the prescription sequence symmetry analysis (PSSA) and parallel group study designs: A systematic review. *PLoS One.* 2018; 13: e0208389.
59. Morris EJ, Hollmann J, Hofer AK, Bhagwandass H, Oueini R, Adkins LE, et al. Evaluating the use of prescription sequence symmetry analysis as a pharmacovigilance tool: A scoping review. *Res Social Adm Pharm.* 2022; 18: 3079–3093.
60. Takeuchi Y, Shinozaki T, Matsuyama Y. A comparison of estimators from self-controlled case series, case-crossover design, and sequence symmetry analysis for pharmacoepidemiological studies. *BMC Med Res Methodol.* 2018; 18: 4.
61. Wang CH, Nguyen PA, Jack Li YC, Islam MM, Poly TN, Tran QV, et al. Improved diagnosis-medication association mining to reduce pseudo-associations. *Comput Methods Programs Biomed.* 2021; 207: 1066181.
62. Schotland P, Racz R, Jackson D, Levin R, Strauss DG, Burkhart K. Target-Adverse Event Profiles to Augment Pharmacovigilance: A Pilot Study With Six New

- Molecular Entities. *CPT Pharmacometrics Syst Pharmacol*. 2018; 7: 809–817.
63. Schotland P, Racz R, Jackson DB, Soldatos TG, Levin R, Strauss DG, et al. Target Adverse Event Profiles for Predictive Safety in the Postmarket Setting. *Clin Pharmacol Ther*. 2021; 109: 1232–1243.
 64. Arfè A, Corrao G. The lag-time approach improved drug-outcome association estimates in presence of protopathic bias. *J Clin Epidemiol*. 2016; 78: 101–107.
 65. Hahsler M, Hornik K. New probabilistic interest measures for association rules. *Intelligent Data Analysis*. 2008; 11: 437–455.
 66. Zhou X, Douglas IJ, Shen R, Bate A. Signal Detection for Recently Approved Products: Adapting and Evaluating Self-Controlled Case Series Method Using a US Claims and UK Electronic Medical Records Database. *Drug Saf*. 2018; 41: 523–536.
 67. Ta CN, Dumontier M, Hripcsak G, Tatonetti NP, Weng C. Columbia Open Health Data, clinical concept prevalence and co-occurrence from electronic health records. *Sci Data*. 2018; 5: 180273.
 68. White RW, Wang S, Pant A, Harpaz R, Shukla P, Sun W, et al. Early identification of adverse drug reactions from search log data. *J Biomed Inform*. 2016; 59: 42–48.