

RESEARCH PAPER SERIES

学術論文から見る製薬産業における
リアルワールドデータの利用実態

岡田 法大
(医薬産業政策研究所 主任研究員)

リサーチペーパー・シリーズ

No. 85
(2025年3月)

日本製薬工業協会
医薬産業政策研究所

本リサーチペーパーは研究上の討論のために配布するものであり、著者の承諾なしに転載、複写・複製することを禁ずる。

本リサーチペーパーに記された意見や考えは著者の個人的なものであり、日本製薬工業協会及び医薬産業政策研究所の公式な見解ではない。

内容照会先：

日本製薬工業協会 医薬産業政策研究所

〒103-0023 東京都中央区日本橋本町2-3-11 日本橋ライフサイエンスビルディング 7F

TEL：03-5200-2681 FAX：03-5200-2684

E-mail：opir-sp@jpma.or.jp

URL：https://www.jpma.or.jp/opir/

目 次

要旨	4
はじめに	5
1 章 製薬産業が関与する学術研究の整理とリアルワールドデータの利用	7
1.1 製薬産業におけるリアルワールドデータの現状	7
1.2 調査に用いたデータベースと調査方法	8
1.3 製薬産業が関与する学術論文の全体像	8
1.4 製薬産業におけるリアルワールドデータの位置づけ	11
2 章 リアルワールドデータの利用の実態	15
2.1 製薬産業で利用されるリアルワールドデータの特徴	15
2.2 製薬産業によるリアルワールドデータ利用の傾向	16
2.3 リアルワールドデータ関連論文の特徴	20
2.4 被引用数上位の論文から見るデータに求められる要件	25
2.5 国際的なリアルワールドデータ関連論文の実態	26
2.6 小括	31
3 章 モバイルデータの利用の実態	32
3.1 製薬産業で利用されるモバイルデータの特徴	32
3.2 製薬産業によるモバイルデータ利用の傾向	32
3.3 モバイルデータ関連論文の特徴	34
3.4 被引用数上位の論文から見る利用の特徴	37
3.5 国際的なウェアラブル技術関連論文の実態	38
3.6 小括	41
4 章 大規模ゲノムデータの利用の実態	42
4.1 製薬産業で利用される分子プロファイル情報の特徴	42
4.2 製薬産業によるバイオバンクに由来する情報の利用の傾向	42
4.3 バイオバンクに関する研究の特徴	44
4.4 国際的な大規模ゲノムデータ関連論文の実態	47
4.5 小括	51
おわりに	53

要 旨

本稿では、製薬産業におけるリアルワールドデータの学術研究での利用に関して、Web of Science Core Collection に収載されている学術論文を基に実施した調査の結果を紹介する。主にグローバルでの医薬品事業売上高上位10社の製薬企業、または国内に本社を有する医薬品事業売上高上位10社の製薬企業に所属する研究者が著者として関与した論文を調査対象とし、これらの論文におけるリアルワールドデータの利用状況を中心に調査を行った。医療機関で取得されるような電子カルテや、保険請求データ等の学術論文の中で明示的にリアルワールドデータという表現を用いて利用されている医療情報に関する調査（2章）に加えて、ウェアラブルデバイスや生体センサーを搭載したスマートフォンから取得されるモバイルデータに関する調査（3章）、ゲノムデータを中心とする分子プロファイルの情報に関する調査（4章）を実施した。製薬企業において利用されている各データの特徴、国際的な連携の現状、ならびに関連する課題に関して定量的な情報を提示する。

医療情報の活用

電子カルテや請求データ等の医療の過程で作成される情報は、実際の医療現場における患者の健康状態や治療過程を反映する貴重な情報源として、多くの学術研究に利用されている。2014年以降、リアルワールドデータを用いた研究は増加しており、直近のデータでは製薬企業が関与する全ての論文の中で、リアルワールドデータに関連する内容を取り上げた論文は10%を超える。利用頻度の高いデータベースは、詳細な疾患特異的な情報を含む疾患単位のデータベースと、請求データ等に由来する疾患横断的な大規模なデータベースの二つに大別される。これらの研究の多くは米国を中心に実施されており、国内の製薬企業が関与する論文も大半が米国の著者を含む論文となっている。米国の医療実態に則したエビデンスを強固にする傾向は米国での売上が大半を占める国内製薬企業のビジネスモデルとも合致している。

モバイルデータの活用

モバイルデータは、スマートフォンやウェアラブルデバイスを用いて疾患状態をリアルタイムで把握する新たな手段として注目を集めている。センサー技術やアプリケーション開発の進展に伴い、測定可能となる情報は拡大しており、その結果として、応用可能な疾患領域が広がってきている。現状では、慣性センサーやスマートフォンのアプリケーションが多くの研究で利用されており、特に神経系の疾患や循環器系の疾患のように、身体活動が疾患の進行や予防に深く関連する領域において研究が活発に行われている。中でも、ウェアラブル技術に関連する研究は、米国と中国を中心に活発に進められており、国内外の製薬企業においても、米国のヘルステック企業との連携が複数開始されている。

分子プロファイルデータの活用

分子プロファイルデータ、特に大規模なゲノムデータベースは、新薬開発の初期段階におけるターゲット探索やバイオマーカーの同定などに利用され、製薬企業の研究開発において重要な情報源となる。これらのデータの取得を目的とし、製薬企業からバイオバンクへ資金提供を行う事例も増加しており、バイオバンクの情報を利用した論文数も増加している。大規模ゲノムデータベースに関する研究は中国からの論文が急速に増加しており、近年では、米国を超える規模に達している。一方、日本における関連研究の論文数は、欧米の主要国や中国と比較して少なく、遺伝的背景に基づく研究の進展には課題が残る。国際的な研究に日本人データを利用するためには、国内外でのデータ共有の仕組みを確立することが不可欠であり、日本人の情報を意思決定に反映させるための検討が必要とされる。

はじめに

リアルワールドデータは、近年、データ収集技術と分析手法の進展により、様々な分野において従来の意思決定プロセスを革新する可能性を持つ重要な情報源として注目を集めている。日常的に収集される人々の行動や社会機能に関する多様なデータを有効活用することは、現実世界の集団をより正確に反映した洞察を得ることに繋がる。これによって、医療、金融、産業、行政等の分野で、従来の方法論では得られなかった新たな知見が明らかとなり、イノベーションの促進に寄与することが期待されている。より身近な事例を示すと、小売店においては、販売データや顧客の特徴、来店パターン等のデータを利用することで、需要予測の精度を向上させ、品切れや過剰在庫の問題を低減させることが可能となる。また、交通政策分野においても、GPS デバイスや交通カメラから得られるデータを解析することで、交通渋滞のパターンを明らかにし、信号タイミングの最適化や新たな道路建設の計画に繋げることが可能となる。これらの事例のように、従来は経験的な観測に基づいて最適化されていた課題解決のアプローチから、より定量的なデータに基づいたデータ駆動型のアプローチへの移行が進むことは、より合理的で信頼性の高い社会の形成に寄与すると考えられる。医療や製薬の分野も例外ではなく、レセプトの電子化や電子カルテの導入により、医療記録を電子情報として蓄積することが可能となり、これらの情報を利用した、医療資源配置の最適化や、治療効果の評価、新たな知見の創出などに関する研究が進展している。

医療分野におけるリアルワールドデータの利用は、COVID-19パンデミック時に、感染地域の疫学調査やワクチンおよび医薬品の有効性・安全性に関する情報提供などの分野で注目を集めた。現実世界で発生する事象について、大規模なデータを基に迅速かつ包括的に評価を行い、適切な医療介入を提供することは、個人の健康維持に不可欠であり、公衆衛生の観点からも特に重要な役割を果たす。製薬分野においても、リアルワールドデータは臨床研究の限界を補完する情報源として重要視されており、臨床研究では確認できないような、医薬品の効果や副作用に関する情報の把握が可能となりつつある。製薬企業が現実世界における医薬品の使用状況や疾患の予後について正確な情報を把握することは、患者個人に対して最適な治療法を提供するため、また医薬品の安全性や有効性に関する適切な情報を発信していくために極めて重要となる。

製薬産業におけるリアルワールドデータの利用は、発展途上の分野であり、データの収集技術および分析手法の進展に伴い、その価値を最大限に引き出すための継続的な検討と技術的研究が不可欠である。信頼性の高い情報を提供するためには、データの品質と解析結果の信頼性を確保するためのガバナンス体制の整備も重要となる。これらの課題に対しては、製薬企業単独の取り組みだけでは十分でなく、大学や医療機関といった他のステークホルダーを巻き込み、多くの研究が進行中である。そこで、本稿では、今後の医療分野への貢献に繋がると考えられる基礎的・応用的な研究の実態を把握することを目的として、リアルワールドデータに関連する学術研究の現状を調査した。これらの研究がより発展することで、データの多様性や規模が拡大し、より高精度で信頼性の高いエビデンスの構築が進むと同時に、それに基づく意思決定が一層合理的かつ科学的に行われるようになり、新薬の開発や医療の質の向上に貢献することが期待される。また、将来的には、AI や機械学習などの解析技術の進展に伴い、製薬企業がリアルワールドデータを利用し、患者のニーズに対応した医薬品を効率的かつ迅速に市場に送り出すことが可能となり、個別化医療や予防医療の実現に寄与することが期待される。

本レポートで使用する語句の説明

語句	説明
グローバル（上位）製薬企業	2022年度の医薬品事業売上高上位10社（Pfizer、AbbVie、Johnson & Johnson、Merck、Novartis、Roche、Bristol-Myers Squibb、AstraZeneca、Sanofi、GlaxoSmithKline）を指し、Web of Scienceにおいて定義されている各社の子会社の情報も集計に含まれる。 引用元：日本製薬工業協会、DATA BOOK 2024、 https://www.jpma.or.jp/news_room/issue/databook/ja/index.html
国内（上位）製薬企業	2022年度の医薬品事業売上高上位10社（武田薬品工業、アステラス製薬、第一三共、中外製薬、大塚ホールディングス、エーザイ、住友ファーマ、田辺三菱製薬、小野薬品工業、塩野義製薬）を指し、Web of Scienceにおいて定義されている各社の子会社の情報も集計に含まれる。 引用元：日本製薬工業協会、DATA BOOK 2024、 https://www.jpma.or.jp/news_room/issue/databook/ja/index.html
リアルワールドデータ関連論文	タイトル又は要旨で以下のいずれかの語句を使用した Web of Science Core Collection に収載されている原著論文 “Real World”、“Health Record”、“Medical Record”、“Claim Database”、“Claim Data”、“Medical Data”、“Data Mart”、“Research Database”、“Patient Registry”、“Registry Data”
モバイルデータ関連論文	タイトル又は要旨で以下のいずれかの語句を使用した Web of Science Core Collection に収載されている原著論文の中から、実際に臨床研究の中でモバイルデータを取得し、それらの研究結果を報告していることが要旨の記載から判断可能であった論文 “Wearable”、“Device”、“Smartwatch”、“Wristband”、“Fitness-bound”、“Fitness Band”、“Wireless Watch”、“Flexible Band”、“Smart Insole”、“Bracelet”、“Accelerometer”、“Gyroscope”、“Pedometer”、“Barometer”、“GPS”、“Humidity Sensor”、“Magnetometer”、“Compass”、“Bluetooth”、“Wi-Fi”、“Mobile”、“Smartphone”、“Smart Phone”、“Mobilephone”、“Mobile Phone”
ウェアラブル技術関連論文	タイトル又は要旨で以下のいずれかの語句を使用した Web of Science Core Collection に収載されている原著論文 “Wearable Device”、“Wearable Technology”、“Wearable Sensor”、“Wearable Health”
バイオバンク関連論文	タイトル又は要旨で“Biobank”という語句を使用した Web of Science Core Collection に収載されている原著論文の中から、実際にバイオバンクに由来する分子プロファイル情報（試料の利用や質問票の情報のみを利用した研究は除く）を利用し、それらの研究結果を報告していることが要旨の記載から判断可能であった論文
大規模ゲノムデータ関連論文	タイトル又は要旨で以下のいずれかの語句を使用した Web of Science Core Collection に収載されている原著論文 “Genome-wide association study”（“GWAS”を含む）、“Mendelian Randomization”（“Mendelian Randomisation”を含む）
論文著者が所属する組織と国	Web of Scienceで構造化された所属の情報と国の情報を解析に利用した。国の単位は、各論文内で記載されている記述を利用しているため、台湾等の地域は論文内の表記の通り集計した。また、イングランド、ウェールズ、スコットランド、北アイルランドは英国として包含して集計を行った。

1 章 製薬産業が関与する学術研究の整理とリアルワールドデータの利用

1.1 製薬産業におけるリアルワールドデータの現状

近年、製薬産業におけるリアルワールドデータの重要性は非常に高くなってきている。リアルワールドデータは、従来の臨床試験では得られない現実世界における疫学情報、健康状態、医薬品の使用状況、効果および安全性に関する貴重な情報を提供し、治療方針の評価において重要な役割を果たす。このようなデータの利用場面は、薬事関連の分野にまで拡大しており、医薬品の承認申請に必要な臨床試験や、市販後の安全性監視においても、リアルワールドデータの活用可能性の検討がプロセスの一環として組み込まれるようになってきている。米国の食品医薬品局（FDA）は、2018年にリアルワールドデータ利用に関するフレームワークを発出し¹、昨年末にはリアルワールドデータとリアルワールドエビデンスの利用推進を目的とした新しい組織となるCDER Center for Real-World Evidence Innovation（CCRI）の設置を発表した²。欧州においても同様に、欧州医薬品庁（EMA）がリアルワールドデータの薬事利用に関するフレームワークを発出しており³、リアルワールドデータの利用をサポートする Data Analysis and Real World Interrogation Network（DARWIN EU）を設置している⁴。国内においても、製造販売後の医薬品の安全性監視を中心に関連する通知が発出されており⁵、限られた事例ではあるものの、医薬品承認申請時におけるリアルワールドデータの利用事例も報告されている⁶。ここで示した取り組みは全体の一部に過ぎず、各地域の規制当局は、リアルワールドデータから得られる知見を医薬品に関する意思決定に反映させるための仕組みを整備しつつある。

リアルワールドデータの薬事利用については、各地域の規制当局を中心に検討が進められている段階であり、実際に薬事評価で利用された事例はまだ限られている。一方、製薬企業内部では、臨床試験の戦略策定、市販後の製品エビデンス創出、患者のセグメンテーションなど多岐にわたる用途において、リアルワールドデータが活用されてきている。リアルワールドデータを利用した調査には、収集される情報の信頼性や網羅性などの観点で様々な課題が存在する一方で、臨床試験のような制約された条件下で得られる限られた患者集団からの情報のみでは捉えきれない、様々な利用場面における安全性や有効性に関する情報が入手できるという利点がある。製薬産業では、基礎的な創薬研究だけでなく、臨床試験や市販後の臨床研究を含む、自社が関与した多くの研究成果を学術論文として積極的に公表している。リアルワールド

1 U.S. Food and Drug Administration, Framework for FDA's Real-World Evidence Program, 2018, <https://www.fda.gov/media/120060/download>

2 U.S. Food and Drug Administration, CDER establishes new Center for Real-World Evidence Innovation, 2024, <https://www.fda.gov/drugs/drug-safety-and-availability/cder-establishes-new-center-real-world-evidence-innovation>

3 European Medicines Agency, Real-world evidence framework to support EU regulatory decision-making, 2024, https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-2nd-report-experience-gained-regulator-led-studies-february-2023-february-2024_en.pdf

4 European Medicines Agency, Data Analysis and Real World Interrogation Network (DARWIN EU), <https://www.ema.europa.eu/en/about-us/how-we-work/big-data/real-world-evidence/data-analysis-real-world-interrogation-network-darwin-eu>

5 厚生労働省、製造販売後の医薬品安全性監視における医療情報データベースの利用に関する基本的考え方について、2017, https://www.mhlw.go.jp/web/t_doc?dataId=00tc2712&dataType=1

6 Asano, J., Sugano, H., Murakami, H., Noguchi, A., Ando, Y., & Uyama, Y. 2025. PMDA Perspective on Use of Real-World Data and Real-World Evidence as an External Control: Recent Examples and Considerations. *Clinical Pharmacology & Therapeutics*.

データを利用した研究も例外ではなく、既に多くの研究成果が学術論文として公表され、医薬品の効果や安全性に関する新たな知見が提供されている。製薬企業が研究成果を学術論文で公表する背景には、研究成果を広く共有することによるイノベーションの促進や、研究開発における連携の強化、査読付きの学術誌で公開することによる科学的信頼性の獲得、さらに研究の透明性の向上といった様々な目的がある⁷。製薬産業における研究成果のアウトプットを示す指標としては、医薬品の上市、特許出願、医薬品の安全性情報の周知などが挙げられるが、これらの指標の中でも学術論文による研究報告は、アウトプットの頻度が高く、各企業の研究開発力を示す重要な指標の一つとして位置付けられる。製薬産業におけるリアルワールドデータの利用は、主に大規模な製薬企業を中心に利用が開始されて数年が経過した段階であり、薬事申請や特許出願に直接繋がる事例はまだ限られているため、現時点では大局的な分析が困難な段階にある。そこで本稿では、製薬産業におけるリアルワールドデータ利用の実態を調査するために、最も適切な指標であると考えられる学術論文の情報に基づき、多角的な分析を行った。

1.2 調査に用いたデータベースと調査方法

今回の調査は、Clarivate 社の Web of Science Core Collection に収載されている学術論文を基に実施した。Web of Science Core Collection には、品質基準を満たした22,000誌を超えるジャーナルに由来する情報が含まれており、科学、社会科学、芸術、人文科学にわたる250以上の分野をカバーする網羅性の高い情報源として知られている。このデータベースは、文部科学省の科学技術・学術政策研究所による科学研究のベンチマーキングにも利用されている⁸。学術論文は、主に大学等の研究機関に所属する研究者によって執筆されることが多く、薬学に関する研究分野においては、産業界に所属する著者が関与する論文は全体の2割程度となっている。今回の調査は、医薬品事業売上高が上位10社に入るグローバル製薬企業又は国内製薬企業が関与した論文を、著者の所属を基に抽出し、その中から各章で扱うトピックに関連する論文を、タイトル又は要旨に含まれる語句と、必要に応じて研究分野の情報をを用いて特定した⁹。各トピックでの抽出条件の詳細は、各章に記載している。Web of Science Core Collectionでは、各学術誌に対して複数の関連分野が割り当てられており、研究分野に関する調査においては、Clarivate社によって割り当てられた研究分野を利用した。製薬企業の学術論文に関する分析を行った後に、各トピックに関連する研究の国際的な関係性や競争力をより広範に把握するため、製薬企業の関与の有無に限らず該当するトピックを扱う全ての論文に関する調査も実施した。ただし、トピック全体に関する調査の結果に含まれる論文の大半は、大学等の研究機関で行われた研究の実績に基づくものであり、その点に留意して解釈を行う必要がある。調査には、2014～2023年の10年間に出版された原著論文のデータを利用しており、それらのデータは、2024年11～12月にかけてデータベースにアクセスした際の情報に基づく。

1.3 製薬産業が関与する学術論文の全体像

はじめに、直近の製薬企業が関与する学術研究全体の特徴を俯瞰するために、医薬品事業売上高上位10社のグローバル製薬企業に所属する研究者が著者として含まれる2023年に出版された全ての原著論文を調

7 McKelvey, M., & Rake, B. 2020. Exploring scientific publications by firms: what are the roles of academic and corporate partners for publications in high reputation or high impact journals?. *Scientometrics*, 122(3), 1323-1360.

8 文部科学省 科学技術・学術政策研究所、科学研究のベンチマーキング2023, <https://www.nistep.go.jp/research/science-and-technology-indicators-and-scientometrics/benchmark>

9 日本製薬工業協会、DATA BOOK 2024, https://www.jpma.or.jp/news_room/issue/databook/ja/index.html

査した。タイトルと要旨がデータベース内に収載されていた7,691報を調査の対象とした。掲載された学術誌とそれらの研究分野の頻度集計の上位をそれぞれ表1-1、表1-2に示す。製薬企業が関与する論文が多く投稿されている学術誌としては、Nature Communications や Scientific Reports、PLOS ONE など、領域横断的なオープンアクセスの学術誌が多く見られた。その他には、包括的な薬学研究、プロセス化学や分子構造等を対象とする学術誌など、疾患領域に依存しない学術誌が上位の大半を占めた。さらに、研究分野の頻度集計では、薬学や化学といった分野の論文が最も多く見られ、それらに続いて疾患領域として、悪性腫瘍や免疫学、神経科学といった製薬産業の近年の重点領域を確認することができた。

続いて、論文のキーワードに用いられている語句について集計を行った。多くの学術誌では、研究者が関連論文を検索するためのツールとして、各論文にキーワードを記載することが求められている。調査対象となった論文において、高頻度で使用されていたキーワードを表1-3に示す。キーワードの集計結果では、新型コロナウイルス感染症や医薬品の安全性、薬物動態に関するトピックを扱う論文が最も多かった。その他には、製薬産業が従来から重点的に研究分野として扱ってきたトピックが上位に並ぶ中で、Quality of life や、Real world evidence、Machine learning といった、近年業界内で特に注目されているトピックも上位に挙がっており、これらの分野も現在の製薬産業において重要な研究課題となっていることが分かる。一部の学術誌では、キーワードが採用されていないため、キーワードが記載されていない論文が約30%存在することや、表記の揺れにより類似表現の名寄せが考慮されていない等の制約があり、これらの要

表1-1 グローバル上位製薬企業が関与した論文の投稿先（2023年）

	投稿先	論文数
1	Nature Communications	127 (1.7%)
2	Scientific Reports	93 (1.2%)
3	Organic Process Research & Development	81 (1.1%)
4	Journal of Pharmaceutical Sciences	80 (1.0%)
5	PLOS ONE	78 (1.0%)
6	Advances in Therapy	77 (1.0%)
7	Journal of Medicinal Chemistry	68 (0.9%)
8	Molecular Pharmaceutics	64 (0.8%)
9	CPT : Pharmacometrics & Systems Pharmacology	62 (0.8%)
10	Journal of Clinical Oncology	56 (0.7%)
10	Vaccine	56 (0.7%)

出所：Web of Science®クラリベイトをもとに作成

表1-2 グローバル上位製薬企業が関与した論文の研究分野（2023年）

	研究分野	論文数
1	Pharmacology & Pharmacy	1,444 (18.8%)
2	Chemistry	947 (12.3%)
3	Oncology	817 (10.6%)
4	Research & Experimental Medicine	612 (8.0%)
5	Immunology	560 (7.3%)
6	Biochemistry & Molecular Biology	537 (7.0%)
7	Science & Technology - Other Topics	500 (6.5%)
8	Neurosciences & Neurology	350 (4.6%)
9	General & Internal Medicine	340 (4.4%)
10	Infectious Diseases	291 (3.8%)

出所：Web of Science®クラリベイトをもとに作成

表1-3 グローバル上位製薬企業が関与した論文のキーワード（2023年）

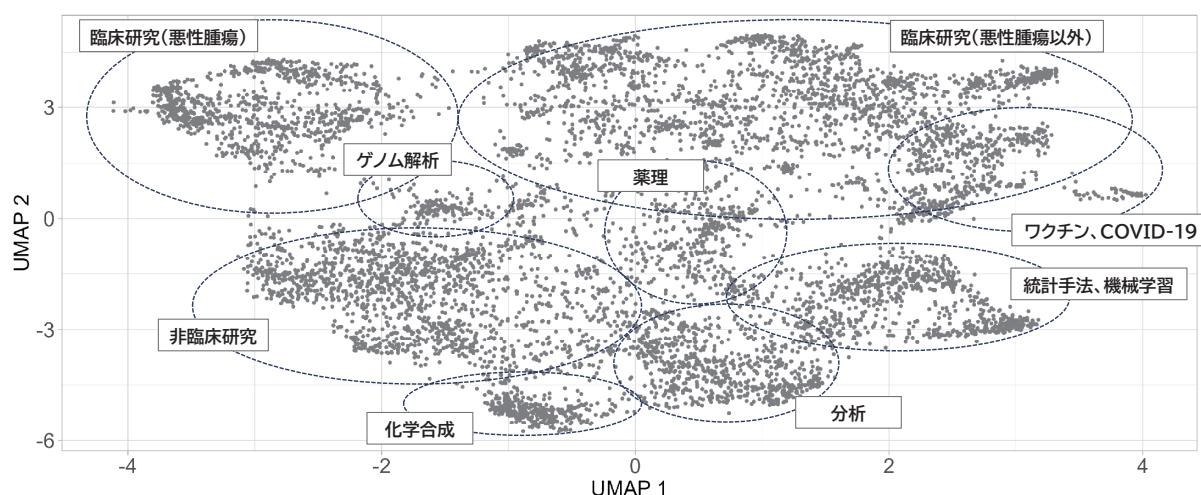
	キーワード	論文数
1	COVID-19	127 (1.7%)
2	Safety	113 (1.5%)
3	Pharmacokinetics	111 (1.4%)
4	Immunotherapy	93 (1.2%)
5	Epidemiology	88 (1.1%)
5	Quality of life	88 (1.1%)
7	Real world evidence	86 (1.1%)
8	Heart failure	83 (1.1%)
9	Machine learning	76 (1.0%)
9	SARS-CoV-2	76 (1.0%)

出所：Web of Science®クラリベイトをもとに作成

因が集計結果に影響を与えている可能性があることには留意が必要である。

製薬企業が関与する論文の傾向をより詳しく把握するために、論文のタイトルおよび要旨の記載をベクトル化し、それらを二次元に圧縮して可視化を行った結果を図1-1に示す。各文章は、PV-DBOW（Paragraph vectors with a distributed bag of words model）を用いてベクトル化し¹⁰、得られた分散表現をUMAP（Uniform Manifold Approximation and Projection）で圧縮して二次元に投影した¹¹。この可視化方法により、論文内で使用されている語句に基づいて、異なる論文間の類似性を距離として定量的に示すことが可能となり、論文の全体像について視覚的理解を深めることができる。図内の点線で囲まれた区分は、論文内で多く使用されている表現（単語のBigram）を基にグループ化した。各グループに含まれる論文のタイトルと要旨において高頻度で使用されている表現を表1-4に示す。論文内で使用された表現を基にグループの特徴を確認すると、化学合成、創薬段階の分析、非臨床研究、薬理、臨床研究（市販後の研究

図1-1 グローバル上位製薬企業が関与した学術論文の全体像（2023年）



出所：Web of Science®クラリベイトをもとに作成

各プロットは論文を示しており、論文のタイトル及び要旨の記述を基に2次元に投影している。

10 Le, Q., & Mikolov, T. 2014. Distributed representations of sentences and documents. *International conference on machine learning* (pp. 1188-1196). PMLR.

11 McInnes, L., Healy, J., & Melville, J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint:1802.03426*.

表1-4 各グループ中に高頻度で使用された表現（単語の Bigram）

臨床研究（悪性腫瘍）		臨床研究（悪性腫瘍以外）		ワクチン、COVID-19	
頻出表現		頻出表現		頻出表現	
1	Free Survival	1	Adverse Events	1	COVID 19
2	Adverse Events	2	Real World	2	Immune Responses
3	Progression Free	3	Clinical Trials	3	Adverse Events
4	Response Rate	4	Confidence Interval	4	SARS CoV
5	Hazard Ratio	5	Cohort Study	5	Clinical Trials
ゲノム解析		薬理		統計手法、機械学習	
頻出表現		頻出表現		頻出表現	
1	Genome Wide	1	Adverse Events	1	Clinical Trials
2	Wide Association	2	Plasma Concentration	2	Drug Development
3	Association Studies	3	Concentration Time	3	Real World
4	Meta Analysis	4	Physiologically Based	4	Treatment Effect
5	UK Biobank	5	Drug Interaction	5	Simulation Study
非臨床研究		化学合成		分析	
頻出表現		頻出表現		頻出表現	
1	Mouse Model	1	Building Blocks	1	Mass Spectrometry
2	Cell Lines	2	Cross Coupling	2	Monoclonal Antibody
3	Single Cell	3	Coupling Reactions	3	Drug Product
4	Gene Expression	4	Late Stage	4	Liquid Chromatography
5	Stem Cell	5	Reaction Conditions	5	Machine Learning

出所：Web of Science[®]クラリベイトをもとに作成

グループ内の論文で使用されている表現（Bigram）を基に作成した。ストップワードが含まれる表現は集計対象外とした。語形変化を問わない形で集計を行ったが、表中では最も多用された語形を示した。

を含む）といった研究開発のプロセスに沿って大きく分類される結果となった。医薬品の市販後の研究について報告する論文で使用された語句は、臨床研究に関する論文の中で使用される語句との間に類似性が見られたため、これらが二次元空間上で明確に区別されることはなかった。一方で、悪性腫瘍を対象とした臨床研究においては、エンドポイントの評価手法として、Time to event analysisが広く利用される傾向にあることから、他の疾患に関する臨床研究で使用される語句とは異なる表現が多く見られ、疾患としてのグループが形成された。また、表1-3で示したように COVID-19関連の論文が数多く存在したため、疾患として独立したグループが形成された。その他の研究分野としては、ゲノム解析や、統計手法、機械学習に関連する研究もグループが形成された。各グループの論文で使用された頻出表現からも、この分類が各論文の特徴を概ね反映したものであることが類推される。

1.4 製薬産業におけるリアルワールドデータの位置づけ

グローバル上位製薬企業が関与した2023年に出版された論文全体におけるキーワードの集計においても、表1-3に示すように“Real world evidence”という表現の使用頻度が全体の7番目に位置しており、リアルワールドデータから得られる情報の重要性が製薬産業において一層高まっていることが分かる。リアルワールドデータのトピックを本稿で扱うにあたり、まず「リアルワールドデータ」と「リアルワールドエビデンス」という語句が指し示す内容を明確に整理することが重要である。はじめに、FDA等の組織から公表されている定義を表1-5に示す。薬事分野におけるリアルワールドデータは、複数の組織で若干の差異が見られるものの、「多様な情報源から得られる日常診療において収集される患者の健康状態や治療

表1-5 Real World Data と Real World Evidence の定義

Real World Data (RWD)	
FDA ¹²	Real-world data are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. Examples of RWD include data derived from electronic health records, medical claims data, data from product or disease registries, and data gathered from other sources (such as digital health technologies) that can inform on health status.
EMA ¹³	Real-World Data (RWD) are data that describe patient characteristics (including treatment utilisation and outcomes) in routine clinical practice. In broader terms, RWD represent data captured in routine care which are not collected in a clinical trial and are relevant to the subject (e.g., age, sex, ethnicity etc.), the disease, the treatment, interactions with the healthcare system, as well as social and environmental factors influencing health status.
厚生労働省 ¹⁴	リアルワールドデータとは、臨床試験とは異なり、実臨床の環境において収集された各種データを指す。
NICE ¹⁵	We define real-world data as data relating to patient health or experience or care delivery collected outside the context of a highly controlled clinical trial. Real-world data can be routinely collected during the delivery of health or social care. (中略) Real-world data can be quantitative or qualitative. Common data types include patient demographics, health behaviours, medical history, clinical outcomes (including patient-reported outcomes), patient or user experiences, resource use, costs, omics, laboratory measurements, imaging, free text, test results and patient-generated data.
Real World Evidence (RWE)	
FDA ¹²	Real-world evidence is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.
EMA ¹³	Through the analysis of RWD, Real-World Evidence (RWE) is generated to answer research questions.
NICE ¹⁵	We define real-world evidence as evidence generated from the analysis of real-world data.

出所：各機関のホームページ等の記載をもとに作成

実態に関するデータ」と定義されることが多い。また、リアルワールドエビデンスは、「リアルワールドデータの分析によりもたらされるエビデンス」と定義されており、こちらの表現は多くの組織の定義で共通している。

次に、リアルワールドデータが、学術論文の中で実際にどのような分野で利用されているかについて調査を行った。図1-1で示した2023年に出版された7,691報の論文において、“Real World” という語句がタイトル又は要旨に使用されている論文を濃いプロット、その他の論文を薄いプロットで表示した結果を図1-2に示す。表1-5で、リアルワールドデータは、主に日常診療での情報に由来することが示されているが、論文での表記においても、“Real World” という語句を使用した論文の大部分は臨床研究のグループに属しており、ゲノム解析や非臨床研究に関連する論文では、“Real World” という表現がほとんど使用さ

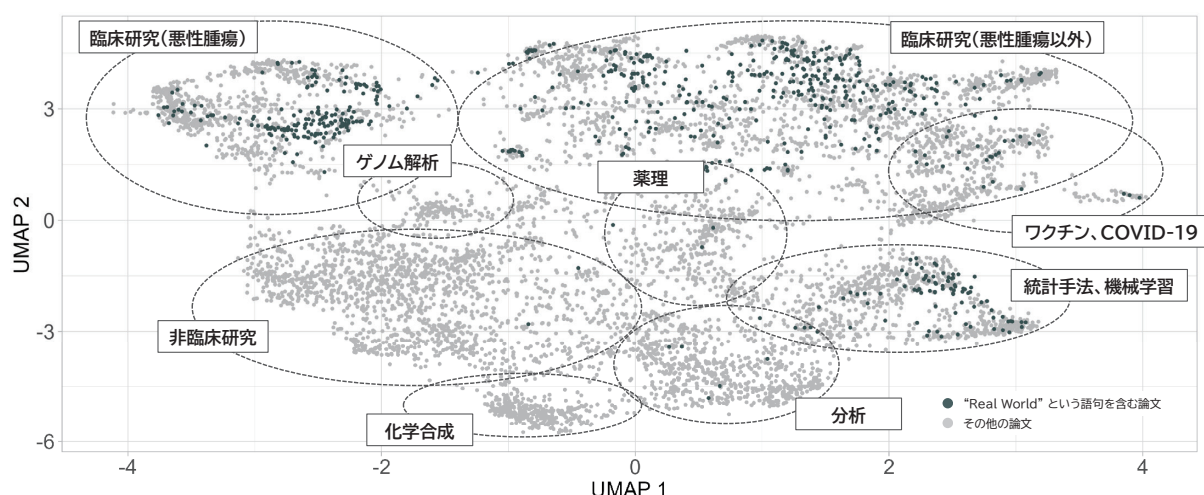
12 U.S. Food and Drug Administration, Real-World Evidence, <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>

13 European Medicines Agency, Data Quality Framework for EU medicines regulation: application to Real-World Data - Draft, 2024, https://www.ema.europa.eu/en/documents/other/draft-data-quality-framework-eu-medicines-regulation-application-real-world-data_en.pdf

14 厚生労働省、第6回 創薬力の強化・安定供給の確保等のための薬事規制のあり方に関する検討会 資料1、2023、https://www.mhlw.go.jp/stf/newpage_36390.html

15 National Institute for Health and Care Excellence, NICE real-world evidence framework, 2022, <https://www.nice.org.uk/corporate/ecd9/chapter/introduction-to-real-world-evidence-in-nice-decision-making>

図1-2 製薬企業が関与した論文全体におけるリアルワールドデータの位置づけ（2023年）



出所：Web of Science®クラリベイトをもとに作成

各プロットは論文を示しており、論文のタイトル及び要旨の記述を基に2次元に投影している。

れていないことが確認された。臨床研究以外の分野では、統計手法や機械学習に分類される論文においても、リアルワールドデータの解析方法に関する論文が存在していることが分かる。

現時点で、学術研究において利用されているリアルワールドデータは、主に医療機関や薬局に由来する情報が中心となっている。一方で、より広義でのリアルワールドデータが指し示す情報源は、FDAやNICEの定義で例示されているように多岐にわたり、患者情報、臨床検査値、処置、投薬、臨床転帰、医療費、オミックス、画像情報、さらにはデジタル技術に基づくデータなど、幅広い情報が含まれている。2018年に公表された論文では、リアルワールドデータとして利用される代表的な情報の種類が整理されており、それらの内容を表1-6に日本語で示す¹⁶。

前項の分析で用いたタイトル又は要旨に“Real World”という語句が使用されている論文からサンプリングを行い、論文の内容を精査した結果、その大多数が表1-6で示す情報のうち、医療機関、投薬、請求データという3区分に由来するデータに関する論文であることが判明した。この傾向は、図1-2からも一部推察することが可能であり、オミックス関連のデータがリアルワールドデータという呼称で利用されていないことが分かる。現状では、情報の測定技術やデータベース化の進展、個人情報保護に関する規制の影響により、上記で示した3区分以外の情報を研究で利用する際は、主に臨床試験の枠組みで情報が入手されることが多くっており、リアルワールドデータと表現して利用される場面は限定的になっていると類推される。

本稿では、これらの背景を踏まえ、次に示すトピックについて詳細に調査する。2章では、現時点でリアルワールドデータという表現で利用されることが多くなっている医療機関、投薬、請求データの情報を中心とした狭義のリアルワールドデータに関する調査を紹介する。3章では、日常生活において継続的に取得される情報として注目されているモバイルヘルスに関する調査を紹介する。4章では、主に研究分野において利用される分子プロファイルに関する調査を紹介する。製薬産業の最前線における情報利用の実態を把握するため、本稿では主に近年出版されたグローバル上位の製薬企業が関与した論文を基に実態の把握を行った。

16 Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. 2018. Innovation at the intersection of clinical trials and real-world data science to advance patient care. *Clinical and translational science*, 11 (5), 450-460.

表1-6 リアルワールドデータの主な情報源

情報	内容
医療機関	患者情報、電子カルテ、臨床検査値、診断、処置、病理組織情報、放射線画像、微生物学的検査結果、診療記録、入退院記録、全身状態
投薬	処方情報、投与情報（投与量、投与経路、医薬品コード）、併用薬、販売時情報（処方薬、OTC薬）、リフィル処方箋、アレルギー
請求データ	医科レセプト、調剤レセプト、その他の医薬品や処置の情報
分子プロファイル	ゲノムと遺伝情報（一塩基多型、パネル検査）、マルチオミックス情報（プロテオミクス、トランスクリプトーム、メタボロミクス、リポドミクス）、その他のバイオマーカー
家系	患者とその家族に関連する健康状態およびアレルギーに関する過去のデータ、喫煙、アルコール
モバイルヘルス	フィットネストラッカー、ウェアラブルデバイス、その他の活動量や身体機能を計測する健康アプリケーション
環境	気候、汚染物質、感染症、生活習慣（食事、ストレス）、その他の環境的・職業的要因
患者報告	患者報告アウトカム、調査票、日記（食生活、習慣）、Personal Health Record (PHR)、有害事象報告、Quality of Life (QOL) 指標
ソーシャルメディア	患者交流、Twitter（現在はX）、Facebook、ブログ
文献	疾病負担、臨床的特徴、有病率・罹患率、治療実態、資源の利用と費用、疾病管理、QOL 指標

出所：Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. 2018. Innovation at the intersection of clinical trials and real-world data science to advance patient care. Clinical and translational science, 11(5), 450-460.

2章 リアルワールドデータの利用の実態

2.1 製薬産業で利用されるリアルワールドデータの特徴

製薬企業が関与する学術論文を基に、リアルワールドデータの利用実態に関して詳細な調査を行った。はじめに、医薬品事業売上高が上位10社に入るグローバル製薬企業に所属する研究者が、著者として含まれる2014～2023年の10年間に出版された原著論文を対象とし、タイトル又は要旨に“Real World”という語句が使用された論文を抽出した。抽出された論文のタイトル又は要旨において、“Real World”と共に使用された表現（単語のBigram）を集計し、高頻度で使用されていた表現を表2-1に示す。この結果からも、治療パターンや臨床診療などの表現が多く使用されていることが分かり、製薬企業が関与するリアルワールドデータに関する研究において、医療機関で取得される情報や請求データが主要な情報源として利用されていることが推察される。

論文内で使用されたリアルワールドデータの情報源は、大別すると表2-2で示す3つの情報源に集約される。患者報告アウトカム等、これらの情報以外を利用した研究も一部存在したが、本章では、最も中心的な情報源として利用されているこれらの日常診療に基づく情報を、現時点で研究に利用可能な狭義のリアルワールドデータとして定義して調査を行った。調査の過程で、これらの情報源を指す表現として高頻度で使用されていた用語を特定し（抽出表現：“Real World”、“Health Record”、“Medical Record”、“Claim Database”、“Claim Data”、“Medical Data”、“Data Mart”、“Research Database”、“Patient Registry”、“Registry Data”）、今回の調査において、これらの表現をタイトル又は要旨に含む論文を、リアルワールドデータ関連論文と定義し、分析の対象とした。

表2-1 グローバル上位製薬企業が関与した論文中で Real world と共起する表現

	頻出表現	論文数
1	World Data	667 (20.5%)
2	Clinical Trial	583 (17.9%)
3	Clinical Practice	565 (17.3%)
4	Patient Receive	538 (16.5%)
5	Treatment Pattern	524 (16.1%)
6	Observational Study	470 (14.4%)
7	Cohort Study	458 (14.0%)
8	Adverse Event	445 (13.7%)
9	World Set	437 (13.4%)
10	Confidence Interval	399 (12.2%)

出所：Web of Science®クラリベイトをもとに作成
グループ内の論文で使用されている表現（単語のBigram）を基に作成した。ストップワードが含まれる表現は集計対象外とし、語形変化を問わない形で集計を行った。

表2-2 製薬企業が関与する論文において高頻度で利用された情報

情報源	内容
電子カルテ	患者の診療経過の記録。患者の傷病名や治療行為等に加えて、臨床検査値やバイタルサインの情報、患者の主訴などが含まれる。
請求データ	患者が保険診療を受けた際に、医療機関が保険者に請求する診療報酬の明細書。患者の傷病名や治療行為、医薬品の処方情報等が含まれる。
レジストリ（疾患登録等）	疾患の情報収集を目的とし、特定の疾患へ罹患した患者等を登録したデータベース。治療歴や疾患の経過など疾患特異的な情報が含まれる。

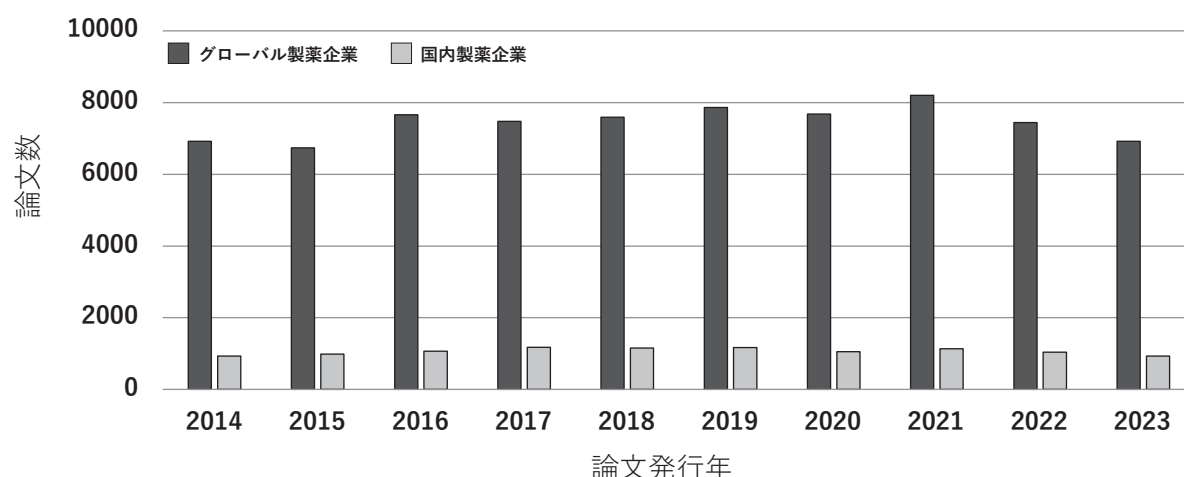
2.2 製薬産業によるリアルワールドデータ利用の傾向

製薬産業におけるリアルワールドデータの利用状況を俯瞰的に把握するため、医薬品事業売上高が上位10社に入るグローバル製薬企業をグローバル上位製薬企業（以降、グローバル製薬企業）、日本国内に本社機能を置く医薬品事業売上高が上位10社の製薬企業を国内上位製薬企業（以降、国内製薬企業）とそれぞれ定義して、これらの企業の社員が著者として関与したリアルワールドデータ関連論文を2014～2023年の10年間にわたり調査した。

はじめに、製薬産業が関与する学術論文の総数を把握するために、グローバル製薬企業と国内製薬企業が関与した学術論文全体の年次推移を図2-1に示す。今回の調査対象となっている製薬企業は、2022年度の売上を基に選定されているが、調査期間内に急激に売上を伸ばして上位に入った製薬企業はほとんど存在しないため、調査対象企業の選択方法が年次変化に与える影響は限定的であると考えられる。全体の論文数の傾向としては、グローバル製薬企業で年間7,000報程度（1社平均700報程度）、国内製薬企業で年間1,000報程度（1社平均100報程度）が出版されており、各年における論文数は概ね一定となっていることが分かる。グローバル製薬企業と国内製薬企業の論文数の差異は、研究開発への投資規模の違いに起因していると考えられる。これらの論文の中における、リアルワールドデータ関連の論文数の年次推移を図2-2に示す。グローバル製薬企業と国内製薬企業が関与するリアルワールドデータ関連論文は、2014年以降増加しており、2023年に出版された論文数は、グローバル製薬企業に所属する著者が関与する論文が年間700報程度（1社平均70報程度）、国内製薬企業に所属する著者が関与する論文が年間100報程度（1社平均10報程度）となっている。これらの結果を踏まえて、各年の全体の論文数におけるリアルワールドデータ関連論文が占める割合の推移を図2-3に示す。グローバル製薬企業と国内製薬企業ではリアルワールドデータ関連論文の論文数には差異が見られるものの、全体の論文数に占める割合は概ね同程度であり、2014年にはリアルワールドデータ関連論文の全体に占める割合が2%程度であったのに対し、2023年にはその割合が10%以上に達していることが分かる。この結果からも国内外の製薬企業における研究活動の中でリアルワールドデータの存在感が大きくなっていることが示唆される。

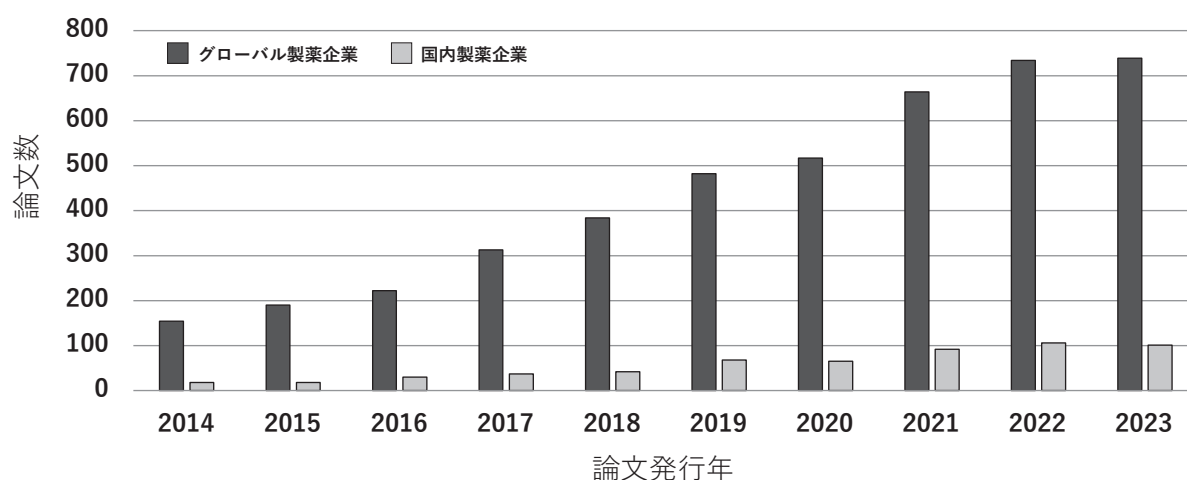
リアルワールドデータに関連する研究の実施について、製薬企業間での分布を確認するため、グローバル製薬企業と国内製薬企業のそれぞれにおいて、10年間に出版されたリアルワールドデータ関連論文全体を分母として、各企業が関与する論文の割合を降順で図2-4に示す。複数の製薬企業の著者による共著論

図2-1 製薬企業が関与する学術論文数の推移



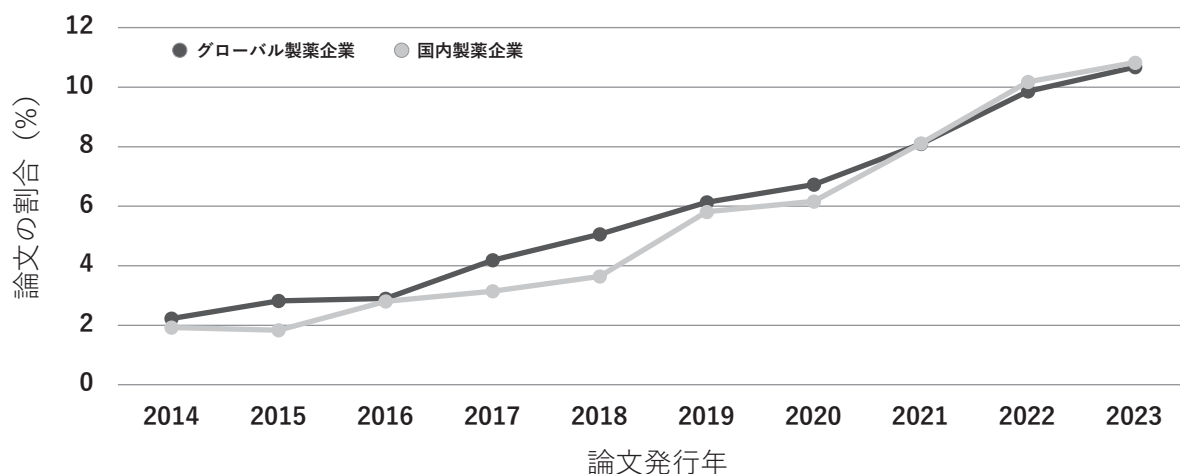
出所：Web of Science®クラリベイトをもとに作成

図2-2 製薬企業が関与するリアルワールドデータ関連論文数の推移



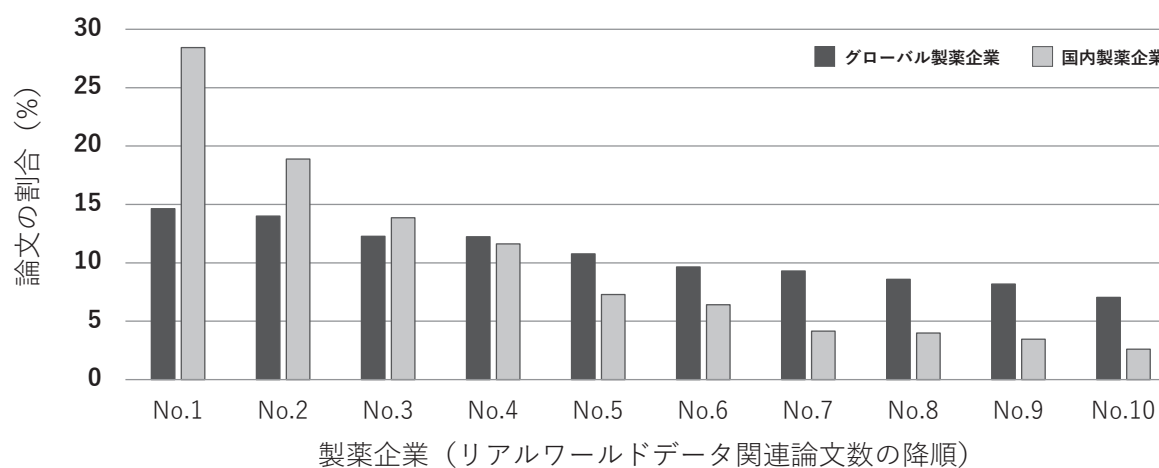
出所：Web of Science®クラリベイトをもとに作成

図2-3 製薬企業が関与するリアルワールドデータ関連論文割合の推移



出所：Web of Science®クラリベイトをもとに作成

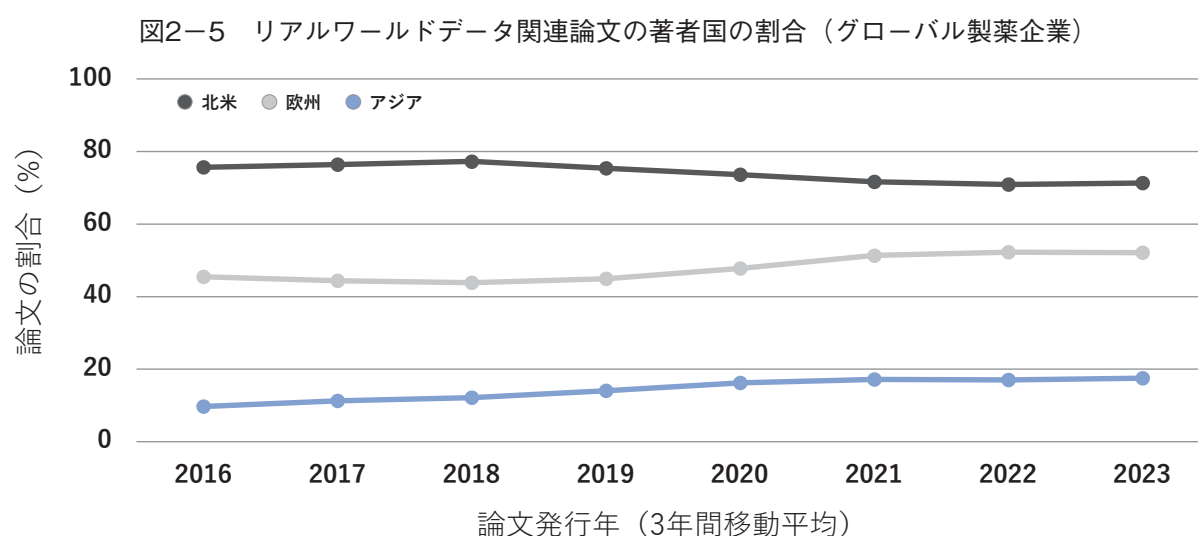
図2-4 各企業が関与するリアルワールドデータ関連論文の割合



出所：Web of Science®クラリベイトをもとに作成

文も存在するため、それらの論文は双方の企業で1つの論文としてカウントした。グローバル製薬企業では、各企業が関与する論文の割合は7～15%の範囲に収まり、企業間の差は小さいことが確認された。このことから、各企業が扱う製品や開発パイプラインの疾患領域、活動拠点となる国に依存せず、全ての企業で一定数のリアルワールドデータに関連する研究が行われていることがわかる。一方で、国内製薬企業では、各企業が関与する割合は3～28%の範囲であり、グローバル製薬企業と比較して大きな差が確認された。特に、一部の企業に研究が集中していることが明らかとなり、リアルワールドデータに関する研究への注力が企業間で偏っていることを示唆している。今回の調査対象となっている製薬企業では、全てのグローバル製薬企業の研究開発費は、国内製薬企業で最も高額である企業の研究開発費を上回っていることから⁹、現時点では、リアルワールドデータに関する研究を率先して進めることができているのは、一定規模以上の製薬企業が中心となっていると考えられる。

リアルワールドデータを利用した研究が積極的に実施されている地域を把握するため、論文著者の所在地の傾向を調査した。今回の調査では、論文に記載された地域情報を基に、製薬企業に所属する著者だけでなく、製薬企業に所属する著者と共著者となっている研究者の所在地も含めて調査を行った。調査対象となっているグローバル製薬企業は、本社機能を北米に持つ企業が5社、欧州に持つ企業が5社であり、調査結果はこれらの本社の所在地の影響を受ける可能性があることに留意する必要がある。グローバル製薬企業のリアルワールドデータ関連論文において、著者が所在する国を調査した結果、グローバル製薬企業の10社全てで米国の著者が関与している論文が最も多いことが分かった。グローバル製薬企業のリアルワールドデータ関連論文における、北米地域、欧州地域、アジア地域の著者の関与割合の3年移動平均の推移を図2-5に示す。複数の地域の著者による共著論文も存在するため、それらの論文は、双方の地域で1つの論文としてカウントした。グローバル製薬企業が関与するリアルワールドデータ関連論文の著者所在地の分布は10年間にわたり大きな変化は見られず、約70%の論文に北米の著者が関与しており、北米が研究の中心地となっていることが分かる。欧州とアジアの著者が関与する割合はそれぞれ僅かに増加しているものの、近年ではそれぞれ約40%と20%に留まっている。一方で、国内製薬企業が関与するリアルワールドデータ関連論文の同様の集計結果を図2-6に示す。年ごとに変動は見られるものの、北米の著者が関与する論文の割合は長期間にわたって高くなっており、直近の時期においても50%以上を占めている。アジア地域の著者は50%に達しておらず、これは国内製薬企業が関与するリアルワールドデータ関連論文の半数以上において、日本の著者が関与していないことを示しており、これらの研究は主に米国内の子会

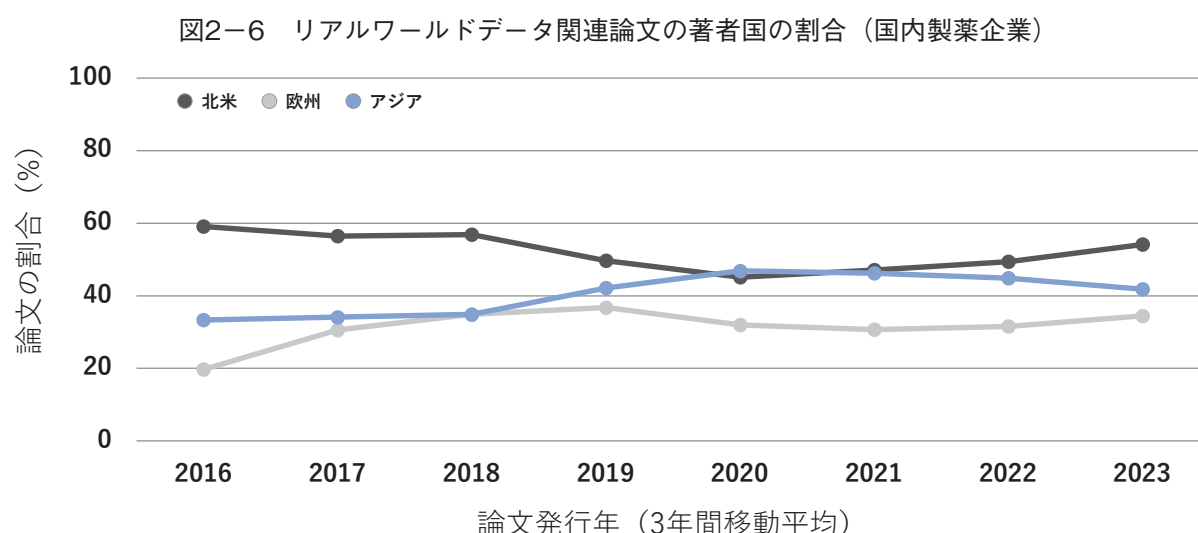


出所：Web of Science[®]クラリペイトをもとに作成

社や部門で実施されていると推察される。国内製薬企業のリアルワールドデータ関連論文数は少なく、単年での集計は結果に大きな変動が見られるため、図2-5と図2-6では表示した年から3年遡った移動平均を示している。

リアルワールドデータ関連論文の内容を確認するために、グローバル製薬企業と国内製薬企業の論文を対象に、それらが掲載された学術誌と、研究分野の頻度集計を行った。上位をそれぞれ表2-3、表2-4に示す。全体の論文の傾向と同様に、領域横断的な学術誌に掲載された論文も複数確認されたが、それに加えて、悪性腫瘍、呼吸器疾患、リウマチ、糖尿病、神経疾患、循環器疾患など、製薬産業が関与するリアルワールドデータ関連論文が研究対象としている疾患領域も確認することができた。

調査対象のリアルワールドデータ関連論文の著者構成については、各企業又はその子会社に所属する著者のみで構成された論文は、グローバル製薬企業の論文と国内製薬企業の論文でそれぞれ12%と11%であり、大半の論文が製薬企業の著者と外部の組織に所属する著者との共著で執筆された論文となっている。調査対象となっているグローバル製薬企業と共に多くの論文に関与した組織を表2-5に示す。グローバル製薬企業における共著者の所属組織を見ると、主に欧米の大学や医療機関との共同研究が多く実施されていることが確認された。加えて、データ解析やデータ集積を担当する企業との共著も確認された。国内製薬企業は論文数が少ないため本集計の結果は割愛する。



出所：Web of Science®クラリベイトをもとに作成

表2-3 製薬企業が関与したリアルワールドデータ関連論文の投稿先

	投稿先	論文数
1	Current Medical Research and Opinion	181
2	Advances in Therapy	167
3	PLOS ONE	124
4	Future Oncology	112
4	Journal of Managed Care & Specialty Pharmacy	112
6	Journal of Medical Economics	102
7	International Journal of Chronic Obstructive Pulmonary Disease	71
8	BMJ Open	66
9	Rheumatology and Therapy	62
10	Diabetes, Obesity and Metabolism	54

出所：Web of Science®クラリベイトをもとに作成

表2-4 製薬企業が関与したリアルワールドデータ関連論文の研究分野

	研究分野	論文数
1	Oncology	721
2	Pharmacology & Pharmacy	720
3	General & Internal Medicine	571
4	Research & Experimental Medicine	485
5	Health Care Sciences & Services	481
6	Neurosciences & Neurology	299
7	Respiratory System	297
8	Cardiovascular System & Cardiology	283
9	Immunology	277
10	Rheumatology	255

出所：Web of Science®クラリベイトをもとに作成

表2-5 グローバル製薬企業と共に論文著者となった組織

	著者が所属する組織	論文数
1	Harvard University	241
2	University of California System	225
3	Analysis Group Inc.	214
4	University of London	193
5	Harvard Medical School	148
6	Assistance Publique Hopitaux Paris (APHP)	141
7	Duke University	139
8	IQVIA	137
9	University of Texas System	132
10	Karolinska Institutet	127

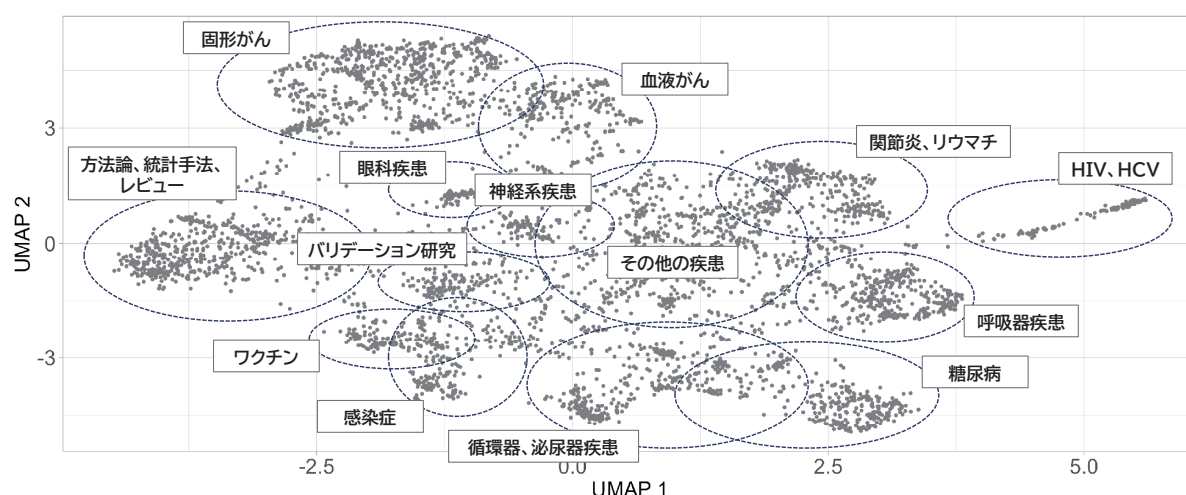
出所：Web of Science®クラリベイトをもとに作成

2.3 リアルワールドデータ関連論文の特徴

本項からは、詳細な内容の精査が可能となる論文数が出版されているグローバル製薬企業のリアルワールドデータ関連論文を基に、実際に研究に利用されたリアルワールドデータの傾向を調査した。グローバル製薬企業の10年間のリアルワールドデータ関連論文のタイトルおよび要旨の記載をベクトル化し、それらを二次元に投影した結果を図2-7に示す。可視化には、図1-1で行った解析と同様の手法を用いた。リアルワールドデータ関連論文を、論文内で使用されている語句で分類すると、主に疾患領域単位のグループが確認された。疾患領域以外では、リアルワールドデータを利用する研究の方法論や、統計手法、バリデーション研究などのグループが確認された。グローバル製薬企業が関与した論文において、多くの研究がなされていた疾患領域として、研究分野の集計においても上位となっていた悪性腫瘍（固形がん、血液がん）、呼吸器疾患、リウマチ、糖尿病、神経疾患、循環器疾患に加えて、眼科疾患や、感染症、ワクチンに関する記載が確認され、製薬産業が関与するほぼ全ての疾患領域においてリアルワールドデータが積極的に活用されていることが推察される。この分析で検出された疾患領域は、各企業の製品や開発パイプラインと連動するため、製薬企業の外部を含めたリアルワールドデータが利用された研究全体の傾向を示すものではないことに留意する必要がある。

ここまで、リアルワールドデータ関連論文の特徴について大局的に確認してきたが、具体的な研究内容や利用されている情報の特徴をより深く理解するために、図2-7でグループ化した疾患領域の中で、特に

図2-7 グローバル製薬企業が関与したリアルワールドデータ関連論文の全体像



出所：Web of Science®クラリベイトをもとに作成
各プロットは論文を示しており、論文のタイトル及び要旨の記述を基に2次元に投影している。

表2-6 悪性腫瘍（固形がん）領域での高被引用論文

	論文	被引用数	データ名称	データ取得国
1	STK11 and KEAP1 mutations as prognostic biomarkers in an observational real-world lung adenocarcinoma cohort	156	- Flatiron Health - Foundation Medicine	米国
2	Evaluating eligibility criteria of oncology trials using real-world data and AI	132	- Flatiron Health	米国
3	Circulating tumor DNA-guided treatment with pertuzumab plus trastuzumab for HER2-amplified metastatic colorectal cancer: a phase 2 trial	125	- SCRUM-Japan Registry	日本

出所：Web of Science®クラリベイトで抽出した論文に医薬産業政策研究所で調査した内容を追加して作成

多くの研究が実施されている固形がん、血液がん、呼吸器疾患の3つの疾患領域と、その他の疾患領域において、各疾患領域で被引用数の多い論文を3論文ずつ抽出し、それらの影響力が大きい研究の内容を精査することで、製薬産業にとって重要度が高いデータベースの特徴の把握を行った。被引用数の多さは発行時期が古い論文の方が高い結果となるため、最新の研究成果が十分に反映されない可能性があることには留意が必要であるが、一定の傾向は確認が可能であると考え。各グループに含まれる論文と、疾患領域に関連する単語を含む論文から、リアルワールドデータを実際に利用し、その研究結果を報告した論文（プロトコル論文、レビュー論文、前向き安全性監視等の報告を除く）を抽出した。本項で示す被引用数は、2024年11～12月にかけて集計されたものであり、現在の被引用数とは異なる。

2.3.1 悪性腫瘍（固形がん）

対象となる論文の中で、最も被引用数が多い3つの論文を表2-6に示す。表に示す被引用数が多い論文は、いずれも悪性腫瘍の疾患領域に特化したデータベースを利用している点で共通している。各論文では、無増悪生存期間、全生存期間、奏効率などの疾患領域特異的な有効性評価のエンドポイントが利用されており、これらの情報の利用可否が固形がんの研究において重要であることが示唆される。また、全ての論文で解析対象集団の定義に遺伝子変異の情報が利用されていることも特徴的である。さらに、各論文では喫煙状況や ECOG performance status、疾患のステージなどの進行度に関する情報も解析時の調整に用い

られていた。利用されたデータの規模の観点では、各論文の解析対象集団は10人から5,000人程度まで幅があるものの、いずれの論文も解析対象集団の人数は1万人を下回る規模であり、評価項目の一つである全生存期間の中央値も1～2年程度と、比較的小規模かつ短期間のデータに基づいて実施されていることが確認された。固形がんの領域では、奏功や増悪といった医師の判断を伴う情報や、患者セグメント、疾患の進行度などに関連する疾患特異的な情報が特に重要であり、個人の詳細な情報が取得されている、対象疾患に特化して構築されたデータベースから得られる研究成果の価値が高いことが推察される。

論文1と論文2で利用されたFlatiron Health社のデータベースは、米国の電子カルテデータに由来している。同社は、悪性腫瘍に特化した電子カルテシステムを医療機関に提供しており、このシステムを通じて、疾患特異的な情報を構造化された形で収集することが可能となっている。これにより、他の情報源と比較して、疾患に関連した詳細なデータを取得できる点で優位性を持っている。また、Flatiron Health社は、コンパニオン診断を行うFoundation Medicine社との提携により、臨床情報とゲノム情報の双方が利用可能となっていることから、より広範な研究に対応可能となっている¹⁷。2024年10月には、同社が保有する患者データが400万人を超え、さらに、このデータを利用した論文が1,000報を超えたことを発表しており、多くのエビデンス創出に貢献している¹⁸。論文3で利用された日本のSCRUM-Japanは、国立がん研究センターが主導する産学連携がんゲノムスクリーニングプロジェクトである。2024年時点で、国内の約215の医療機関に加え、製薬企業や検査会社など17社以上が参加していることが公表されている¹⁹。このプロジェクトは、臨床情報に加えてがん遺伝子パネル検査の情報を有しており、疾患特異的な情報を前向きに収集している点において他の情報源と比較して優位性を持っている。

2.3.2 悪性腫瘍（血液がん）

対象となる論文の中で、最も被引用数が多い3つの論文を表2-7に示す。論文1と論文3では、固形がんと同様に悪性腫瘍の疾患領域に特化したデータベースが利用された。一方、論文2では、請求データに由来するデータベースが利用されており、疾患特異的な詳細な情報を含まないデータベースを利用した研究も、多くの文献で引用される成果に繋がっていることが分かる。いずれの論文でも無増悪生存期間およ

表2-7 悪性腫瘍（血液がん）領域での高被引用論文

	論文	被引用数	データ名称	データ取得国
1	Toxicities and outcomes of 616 ibrutinib-treated patients in the United States: a real-world analysis	359	- Connect CLL Registry	米国
2	Trends in overall survival and costs of multiple myeloma, 2000-2014	254	- MarketScan - Medicare Supplemental	米国
3	Big data analysis of treatment patterns and outcomes among elderly acute myeloid leukemia patients in the United States	227	- SEER - Medicare	米国

出所：Web of Science[®]クラリベイトで抽出した論文に医薬産業政策研究所で調査した内容を追加して作成

17 Foundation Medicine, Press Release: Foundation Medicine and Flatiron Health Collaborate to Develop First-in-Class Data Platform to Accelerate Precision Medicine for Cancer, <https://www.foundationmedicine.com/press-releases/foundation-medicine-and-flatiron-health-collaborate-to-develop-first-in-class-data-platform-to-accelerate-precision-medicine-for-cancer>

18 Flatiron Health, Press Release: Flatiron Health surpasses 1,000 research publications using high-quality real-world data from its de-identified patient database, <https://resources.flatiron.com/press/flatiron-health-surpasses-1000-research-publications>

19 国立がん研究センター 東病院、SCRUM-Japanとは、https://www.scrum-japan.ncc.go.jp/scrum_japan/about/index.html

び全生存期間を有効性評価のエンドポイントとしており、この傾向は固形がんにおける傾向と一致している。生存期間を算出する根拠となる治療開始日と死亡日は、客観性の高いエンドポイントであり、保険の情報やがん登録等の情報からも取得可能である。そのため、全生存期間を疫学的な指標として調査する場合は、疾患領域に特化したデータベース以外を用いても推定が可能となっている。論文2と論文3は、米国における治療パターンや医療費に関する統計を主目的としていたため、個人単位の詳細な情報よりも、長期間にわたって追跡された患者が多く含まれるデータベースの利用が優先されたと推察される。請求データを基にした論文内では、細胞遺伝学的所見や細胞機能異常に関する情報が含まれないことがリミテーションとして述べられていた。利用されたデータの規模の観点では、論文1は、約600人を対象として、細胞遺伝学的な所見等の詳細な情報を用いた小規模な研究であった。それに対して、請求データを基本とした論文2と論文3では、それぞれの解析対象患者数は約18,000人および約8,000人と比較的大規模な研究であった。論文1と論文3の観察期間又は全生存期間の中央値は2年以下と比較的短期間あったが、論文2は5年程度の生存期間の中央値を観察するために、より長期のデータが利用されていた。血液がんの領域では、固形がんと比較すると、大規模かつ長期間にわたる情報を用いた研究が、多くの文献で引用される成果に繋がっている傾向がある。

論文1で利用された Connect CLL Registry は、米国の199医療機関が参加する慢性リンパ性白血病の前向きレジストリであり、全体では1,494人の患者が登録され、最長5年間追跡が行われている。追跡データは研究期間中、約3か月ごとに収集された²⁰。データには、患者背景や病歴等の情報だけでなく、白血病の診断とモニタリングに重要となる臨床検査値やフローサイトメトリーの結果、健康関連 QOL の結果も含まれている。論文2は、Merative 社が管理する MarketScan および Medicare Supplemental のデータベースを利用した研究であり、主に民間保険の請求データに由来する。MarketScanは、米国全土で毎年約4,000万人のデータが利用可能となる長期間のデータベースであり、対象人数や観察期間の観点では、米国内で最大規模のデータベースとなっている。論文3は、米国の国立がん研究所が管理する SEER Database および CMS (Center for Medicare and Medicaid Services) が管理する Medicare の登録・請求データベースを利用した研究であり、主に公的ながん登録および公的保険の請求データに由来する。SEER Databaseは、米国人口の約48%をカバーする大規模ながん登録であり、毎年47万人以上の新規症例が登録されている²¹。

2.3.3 呼吸器疾患

対象となる論文の中で、最も被引用数が多い3つの論文を表2-8に示す。表に示す被引用数が多い論文は、いずれも欧州の電子カルテに由来するデータベースを利用している点で共通しており、特定の治療法に注目し、安全性または有効性に関する予後の評価を行った研究であった。特に論文2と論文3は、有効性評価指標として、アレルギー喘息の治療効果を評価する Global Evaluation of Treatment Effectiveness (GETE) や、COPDの評価尺度である Modified Medical Research Council Dyspnea Scale (mMRC) といった重症度分類を用いており、疾患特異的なエンドポイントが利用されていた。日常診療において電子カルテに記録される疾患状態の指標や、日常診療で記録される情報を基に遡及的に導出可能な指標をエンドポイントとして利用することが可能であれば、特定の疾患に依存しない電子カルテのデータベースを用いて

20 Mato, A., Nabhan, C., Kay, N. E., Weiss, M. A., Lamanna, N., Kipps, T. J., ... & Sharman, J. P. 2016. Real-world clinical experience in the Connect[®] chronic lymphocytic leukaemia registry: a prospective cohort study of 1494 patients across 199 US centres. *British journal of haematology*, 175(5), 892-903.

21 Daly, M. C., & Paquette, I. M. 2019. Surveillance, epidemiology, and end results (SEER) and SEER-medicare databases: use in clinical research for improving colorectal cancer outcomes. *Clinics in colon and rectal surgery*, 32(01), 061-068.

表2-8 呼吸器疾患領域での高被引用論文

	論文	被引用数	データ名称	データ取得国
1	Adverse outcomes from initiation of systemic corticosteroids for asthma: long-term observational study	277	- OPCR - CPRD	英国
2	Omalizumab effectiveness in patients with severe allergic asthma according to blood eosinophil count: the STELLAIR study	171	- EHR (医療機関)	フランス
3	The inevitable drift to triple therapy in COPD: an analysis of prescribing pathways in the UK	135	- OPCR	米国

出所：Web of Science®クラリペイトで抽出した論文に医薬産業政策研究所で調査した内容を追加して作成

有効性の評価が行える疾患領域が存在することが確認された。一般的に、疾患特異的な電子カルテや疾患レジストリを情報源とするデータベースと比較して、対象疾患を特定せず、多数の医療機関が連携して情報を収集している電子カルテデータベースは、患者数確保の側面で優位性がある。日常診療において記録される情報から導出できるエンドポイントが研究に利用できる場合や、研究対象となった疾患領域に特化したデータベースが存在しない場合には、これらの広範なデータベースの利用が選択肢に挙がる。利用されたデータの規模の観点では、論文2は、約900人を対象とした比較的小規模な研究であったが、論文1と論文3は、それぞれ約5万人、約1万人の解析対象患者数の確保が可能となる大規模なデータベースを利用した研究であった。また、論文2は1年間の短い研究であったが、論文1と論文3の観察期間は10年前後となっており、悪性腫瘍の領域の研究と比較すると、長期間の情報を利用している傾向が確認された。呼吸器疾患の領域では、電子カルテ等の一般診療で取得されるデータを大規模かつ長期的に収集したデータベースから得られる研究成果の価値がより高いことが推察される。

論文1と論文3では、Optimum Patient Care社が管理する電子カルテデータベースである Optimum Patient Care Research Database (OPCRD) が利用されていた。このデータは英国の1,100を超えるプライマリケアの医療機関に由来する2,400万人を超える患者の情報が含まれる。OPCRDは、電子カルテの情報を基本としているが、患者への質問票も含まれており、100を超える出版物の研究に利用されている²²。さらに、論文1では、追加して、国立の研究所が支援して情報が収集されている Clinical Practice Research Datalink (CPRD) も利用された。このデータも2,000を超えるプライマリケアの医療機関の電子カルテに由来する6,000万人を超える患者の情報が含まれており、その中の25%の患者は20年以上の観察期間となっていることが特徴となっている²³。

2.3.4 その他の疾患

上記で言及していない疾患に関する論文の中で、最も被引用数が多い3つの論文を表2-9に示す。被引用数が多い論文は、COVID-19のワクチンの有効性に関して報告した2報の論文と、2型糖尿病の治療法の有効性に関して複数の国のデータベースを利用して大規模に実施された研究を報告する論文であった。COVID-19のワクチンの有効性に関する2報の論文は、いずれも2021年に発表されており、2020年末から開始されたワクチン接種の有効性評価を目的に、大規模なデータを利用して迅速に研究が行われたことが、多くの文献で引用される要因となったと考えられる。このような感染症流行時に迅速に研究に利用可能な大規模データベースは限られており、研究に利用できる医療情報基盤の整備が整っていたイスラエルの国

22 Optimum Patient Care, Optimum Patient Care Research Database, <https://opcrd.optimumpatientcare.org>

23 National Institute for Health and Care Research, Clinical Practice Research Datalink, <https://www.cprd.com>

表2-9 その他の疾患領域での高被引用論文

	論文	被引用数	データ名称	データ取得国
1	Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data	1,169	- Nationwide Surveillance Data	イスラエル
2	Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study	832	- EHR (Kaiser Permanente Southern California)	米国
3	Lower Risk of Heart Failure and Death in Patients Initiated on Sodium-Glucose Cotransporter-2 Inhibitors Versus Other Glucose-Lowering Drugs The CVD-REAL Study (Comparative Effectiveness of Cardiovascular Outcomes in New Users of Sodium-Glucose Cotransporter-2 Inhibitors)	657	- MarketScan - Diabetes Prospective Follow-Up initiative - National full-population registries - CPRD	複数 (6 か国)

出所：Web of Science® クラリベイトで抽出した論文に医薬産業政策研究所で調査した内容を追加して作成

家的なデータベースや、米国のHealth maintenance organization (HMO) であるKaiser Permanenteのデータベースは、公衆衛生の観点において極めて重要な役割を果たしたことが分かる。一方、2型糖尿病の治療法に関する論文は、SGLT2阻害薬の有効性を評価した研究であり、複数のデータベースを利用して二重盲検試験の結果を検証する形で実施された。この研究では、異なる国で整備された複数の大規模なデータが利用されており、リアルワールドデータの有用性を示す代表的な事例となっている。利用されたデータの規模の観点では、各論文の解析対象集団は約30万から約650万人の規模で、いずれも大規模な研究であったが、平均で1年以内という短期間のデータに基づいて行われた。

論文1と論文2では、各国のHMOで収集された情報が利用されており、個人の識別番号で情報の追跡が可能である点が特徴である。HMOが保有するデータベースの中でも、Kaiser Permanente（米国）やMaccabi（イスラエル）のデータベースは、これらの論文以外にも複数の研究で利用されていることが確認されており、リアルワールドデータを利用した研究において、重要な情報源として認識されている。また、論文3では、前述した米国のMarketScanに加え、英国、ドイツ、スウェーデン、ノルウェー、デンマークといった国々の患者登録の情報が利用されており、診断コードや死亡データを基にした大規模な国際的な研究が実施された。

2.4 被引用数上位の論文から見るデータに求められる要件

多くの文献で引用され、影響力の大きい成果に繋がった研究で利用されたデータベースの特徴は多岐にわたっており、研究のリサーチクエスションに応じて求められるデータベースの要件は異なっている。特に、利用頻度の高いデータベースは大きく二つに分類できる。一つは、電子カルテやレジストリに基づき、詳細な疾患情報を比較的小規模で収集するデータベースであり、もう一つは、請求データのように情報の構造化や標準化が進んでおり、大規模に情報を収集することが可能となっているデータベースである。データベースの選択は、主に調査の目的（エンドポイントの捕捉可否、コホートの定義可否など）とデータの規模（対象患者数、追跡期間など）に基づき決定される。これらの要件は疾患領域にも関連しており、例

例えば悪性腫瘍の研究では、予後に影響を与える因子として遺伝子変異や細胞機能異常などの分子プロフィール情報や、疾患のステージ情報の有無が重視される傾向がある。一方、慢性疾患が多く含まれる呼吸器疾患の研究では、長期的な観察期間の確保が重要な要素とされる傾向がある。米国の MarketScan のように、大規模で長期間にわたる情報が含まれるデータベースが存在する中においても、他のデータベースを用いた研究の被引用数が多いことから、より精緻な研究結果を導くためには、現在の構造化された請求データに含まれる情報の拡充が望まれていることが分かる。また、データの内容だけでなく、社会的価値の高い研究結果を導出するためには、データ利用の即時性や国際的なデータ連携の強化も不可欠である。迅速かつ国際的なデータ提供の基盤となるデータの標準化やガバナンス体制の構築は一朝一夕に整備できるものではないため、持続的な検討と改良が必要である。

2.5 国際的なリアルワールドデータ関連論文の実態

製薬企業が関与するリアルワールドデータ関連論文は、約90%が外部組織との共著によるものであり、表2-5で示したように大学や医療機関と共著で執筆されるものが多い。リアルワールドデータ関連論文全体においても、製薬企業が関与する研究は限定的であり、これらの分野の研究の進展には大学をはじめとする外部の研究機関の研究力の向上が不可欠である。そこで、リアルワールドデータ関連の研究の全体の潮流を把握するために、本項では、製薬企業が関与していない研究も含めて、特に国際的な研究の実施に焦点を当てて調査を行った。

本項までの調査では、リアルワールドデータ関連論文で多用される特徴的な表現と、製薬企業の関与を基に、医療や製薬分野でのリアルワールドデータの利用に関連する論文を特定した。しかしながら、リアルワールドという表現は、医療や製薬分野以外でも利用されるため、製薬企業の関与を抽出条件から除外すると、医療や製薬分野に関連しない論文が結果に含まれることとなる。それらの論文を可能な限り除外するために、製薬企業が関与したリアルワールドデータ関連論文の上位20の研究分野に該当する論文のみを調査対象とし、その中からリアルワールドデータ関連論文で特徴的な表現（抽出表現：“Real World”、“Health Record”、“Medical Record”、“Claim Database”、“Claim Data”、“Medical Data”、“Data Mart”、“Research Database”、“Patient Registry”、“Registry Data”）をタイトル又は要旨に含む、2014～2023年の間に出版された原著論文を基に調査を実施することとした。医療機関において執筆された論文も含まれることから、本項の調査では、医療情報の二次利用に関する研究に加えて、医療情報の一次利用に関する研究も含まれることとなる。医療情報の二次利用を推進していくためには、一次利用のための情報の取得

表2-10 リアルワールドデータ関連論文の著者の所在地

	著者の所在地	論文数
1	米国	34,616 (38.0%)
2	中国	10,372 (11.4%)
3	英国	8,173 (9.0%)
4	台湾	5,516 (6.1%)
5	韓国	4,997 (5.5%)
6	日本	4,935 (5.4%)
7	ドイツ	4,860 (5.3%)
8	カナダ	4,439 (4.9%)
9	イタリア	4,213 (4.6%)
10	オーストラリア	3,818 (4.2%)

出所：Web of Science® クラリベイトをもとに作成

表2-11 リアルワールドデータ関連論文の著者所属組織

	著者が所属する組織	論文数
1	Harvard University	4,316 (4.7%)
2	University of California System	3,447 (3.8%)
3	Harvard Medical School	2,710 (3.0%)
4	University of London	2,444 (2.7%)
5	US Department of Veterans Affairs	1,905 (2.1%)
6	University of Texas System	1,861 (2.0%)
7	Veterans Health Administration	1,801 (2.0%)
8	Mayo Clinic	1,737 (1.9%)
9	University System of Ohio	1,708 (1.9%)
10	Brigham & Women's Hospital	1,670 (1.8%)
158	University of Tokyo	445 (0.5%)

出所：Web of Science® クラリベイトをもとに作成

および利用に関する環境整備が重要となるため、一次利用と二次利用の双方に関する研究の成果を総合的に評価する形で傾向を確認することとした。

これらの論文の著者構成について、著者の所在地と著者が所属する組織を表2-10と表2-11に示す。複数の国や所属の著者で共著となっている論文は、それぞれの国、組織で集計に加えた。集計の結果、米国の組織に所属する著者が関与している研究が、他国と比較して顕著に多いことが分かった。一方で、中国、台湾、韓国、日本といったアジア地域の著者も多く論文に関与していた。アジア地域からの著者が少ない結果となっていた前項までの国内外の製薬企業のリアルワールドデータ関連論文の著者に関する調査の結果とは異なる傾向が確認された。特に、台湾では、National Health Insurance を中心とした医療情報のデジタル化が進んでいることが知られており、服薬履歴の共有サービスや、患者自身で医療機関の利用情報等が管理できるアプリケーションなど、10年以上前から様々なシステムが整備され始めている。このような環境整備が、リアルワールドデータに関する研究を促進しており、台湾の論文数が多い要因の一つとなっていると考えられる²⁴。また、著者が所属している組織は、ハーバード大学（米国）、カリフォルニア大学システム（米国）、テキサス大学システム（米国）をはじめとした米国の研究機関が上位を占めており、米国以外の組織で上位10組織に含まれたのは、ロンドン大学（英国）のみであった。このように、大規模な研究組織が米国に集中していることが、米国からの論文発表が顕著に多い理由の一因と考えられる。日本においては、国別の比較では上位に位置するものの、組織別の集計では特に上位に位置する研究組織は存在していなかった。最も多くの論文を発出した機関は、全体の158番目となる東京大学であり、京都大学、国立がん研究センター、慶応義塾大学等で同程度の論文数が発出されている状況であった。

各国の研究における特徴を把握するために、最も論文数が多い米国、アジアで論文数が多い中国と台湾に加え、日本における各国の著者が関与する論文でタイトル又は要旨に使用されている表現（Bigram）の集計を行った。各国のリアルワールドデータ関連論文において高頻度で使用された表現を表2-12に示す。米国、中国、日本の研究においては、米国では、Electronic Health RecordsやElectronic Medical Recordsといった表現が他国と比較して高頻度で使用されていたことなど、国ごとに若干の特徴は見られるものの、共通して Medical Records という語句の使用が多く確認された。一方、台湾においては、National Health Insurance Research Database（NHIRD）という健康保険のデータベースを利用した研究が多く見

24 Lu, J. F. R., & Liang, L. L. 2024. The role of digital health under Taiwan's national health insurance system: progress and challenges. *Health Systems & Reform*, 10(2), 2375433.

表2-12 各国のリアルワールドデータ関連論文で使用された表現

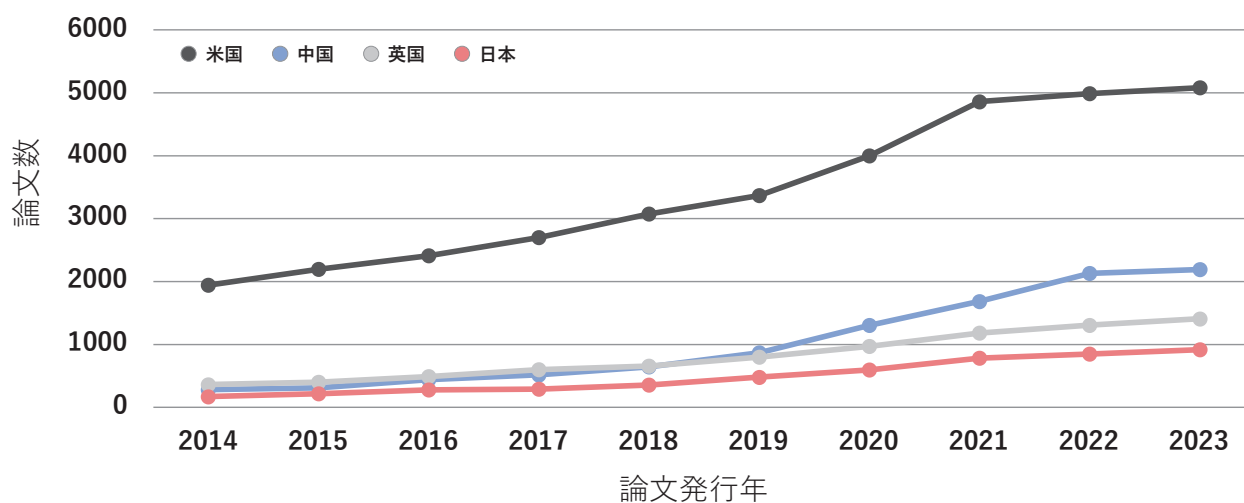
米国			中国		
	頻出表現	論文数		頻出表現	論文数
1	Medical Record	13,401	1	Medical Record	5,023
2	Real World	11,472	2	Real World	4,381
3	Health Record	8,361	3	Risk Factor	1,713
4	Electronic Health	7,902	4	Study Aim	1,668
5	Electronic Medical	5,229	5	Logistic Regression	1,437
6	Cohort Study	4,383	6	Retrospective Study	1,289
7	Confidence Interval	4,179	7	Confidence Interval	1,207
8	Logistic Regression	3,274	8	Regression Analysis	1,195
9	Health Care	3,163	9	Cohort Study	1,120
10	Risk Factor	2,896	10	Significant Difference	1,061

台湾			日本		
	頻出表現	論文数		頻出表現	論文数
1	Research Database	3,523	1	Medical Record	2,308
2	Health Insurance	3,503	2	Real World	2,115
3	National Health	3,476	3	Confidence Interval	899
4	Insurance Research	3,307	4	Risk Factor	642
5	Cohort Study	2,225	5	Study Aim	632
6	Hazard Ratio	2,019	6	Adverse Event	616
7	Population Base	2,008	7	Multivariate Analysis	577
8	Confidence Interval	1,911	8	Retrospectively Review	569
9	Taiwan National	1,161	9	Cohort Study	555
10	Medical Record	1,082	10	Regression Analysis	543

出所：Web of Science®クラリベイトをもとに作成

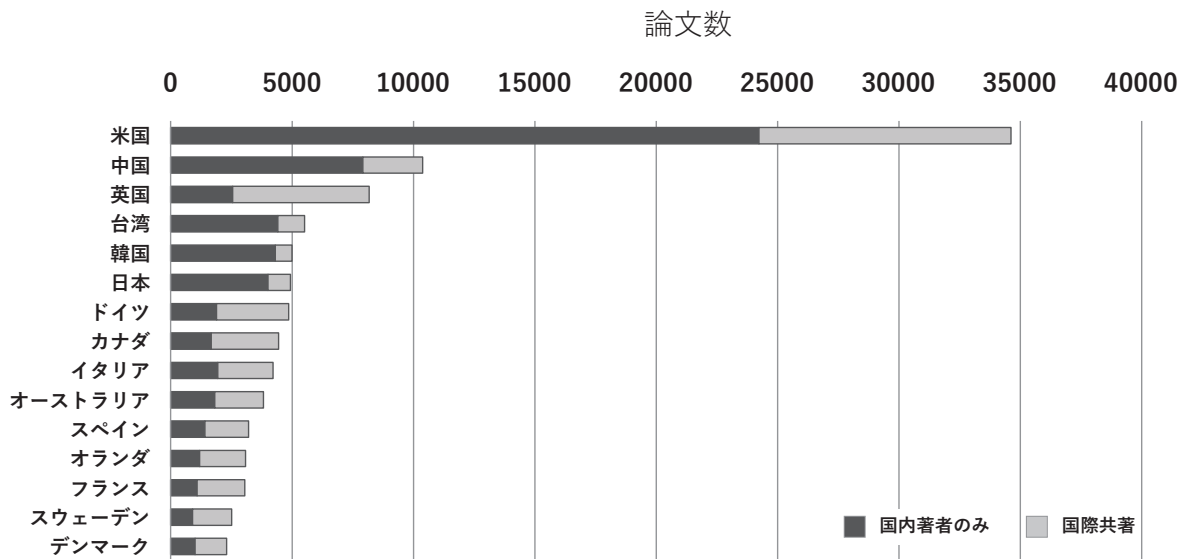
論文で使用されている表現（Bigram）を基に作成した。ストップワードが含まれる表現は集計対象外とし、語形変化を問わない形で集計を行った。

図2-8 リアルワールドデータ関連論文数の各国の推移



出所：Web of Science®クラリベイトをもとに作成

図2-9 リアルワールドデータ関連論文数と各国の国際共著



出所：Web of Science®クラリベイトをもとに作成

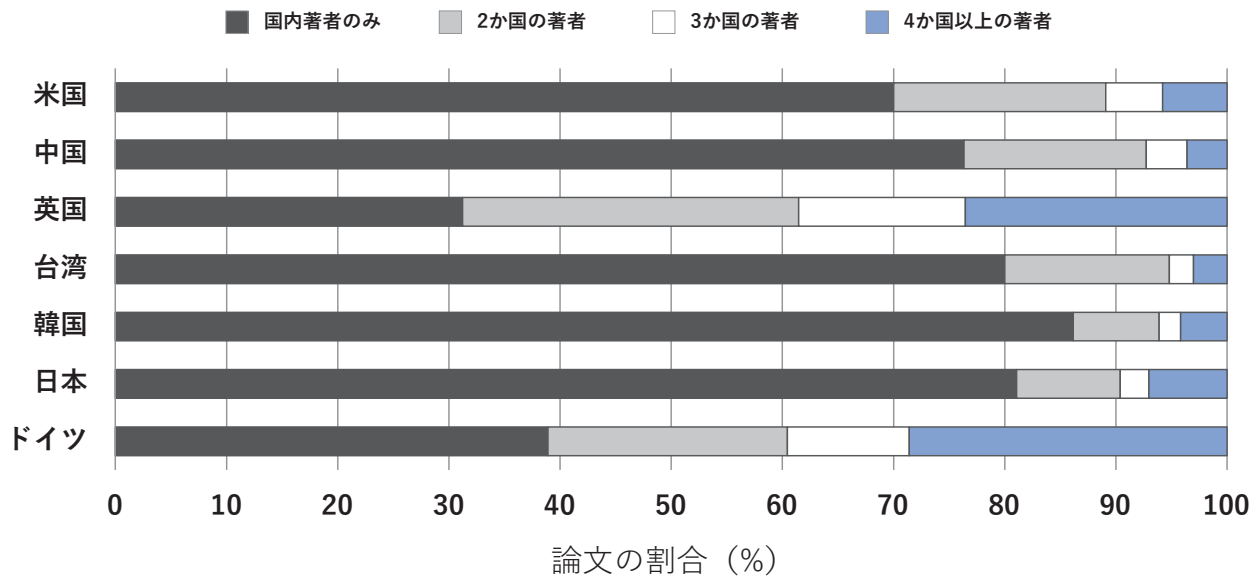
られ、他国と異なる傾向が確認できる。NHIRDは、請求データを中心とした公的なデータベースであり、がん登録、がん検診、希少疾患データベース等の様々なデータベースと統合して解析が行えることが利点となっている。特に、2014年以降、NHIRDを利用した出版物が毎年300件以上公表されていることも報告されており、台湾におけるリアルワールドデータ研究の発展において重要な役割を果たしていることが分かる²⁵。

論文数が上位となった国の出版数の年次推移を図2-8に示す。台湾と韓国と日本の論文数の推移は概ね同様の推移となっているため、代表して日本の推移を示す。図2-1で示した製薬企業が関与するリアルワールドデータ関連論文の論文数の推移と同様に、2014年以降、これらの国々が関与する論文の数は継続的に増加している。上位国における論文数の増加傾向には大きな差異は見られない。この結果から、製薬企業の外部においても、リアルワールドデータに関連する研究は各国で活発に行われるようになっていることが分かる。

リアルワールドデータ関連論文における国境を跨いだ共同研究について調査をするために、2014～2023年までの10年間に出版された論文の著者構成に関して、同一国の著者のみで構成されている論文と、複数国の著者が含まれる論文を分類して国ごとに集計を行った結果を図2-9に示す。また、当該国の研究機関に所属する研究者が一人でも含まれている論文の数を分母として、各論文の著者が所属する組織の国数の分布を図2-10に示す。論文数が上位である米国と中国では、国際共著論文（2か国以上の著者で構成される論文）の割合がそれぞれ2～3割程度であった。それらに続く、台湾、韓国、日本といった東アジアの国は国際共著論文が2割以下である一方で、英国、ドイツ、イタリア等の欧州の国では、国際共著論文が半数以上となっており、論文数が上位の国でも研究体制に違いが確認された。米国と東アジアの国では、国際共著論文の割合が低く、欧州の国際共著論文の割合が低い傾向があることは、科学技術・学術政策研究所による科学研究のベンチマーキングで報告されている結果と同様の傾向であり、国際共著論文の割合は、2019～2021年の臨床医学全体の論文の傾向と類似していることが確認された⁸。

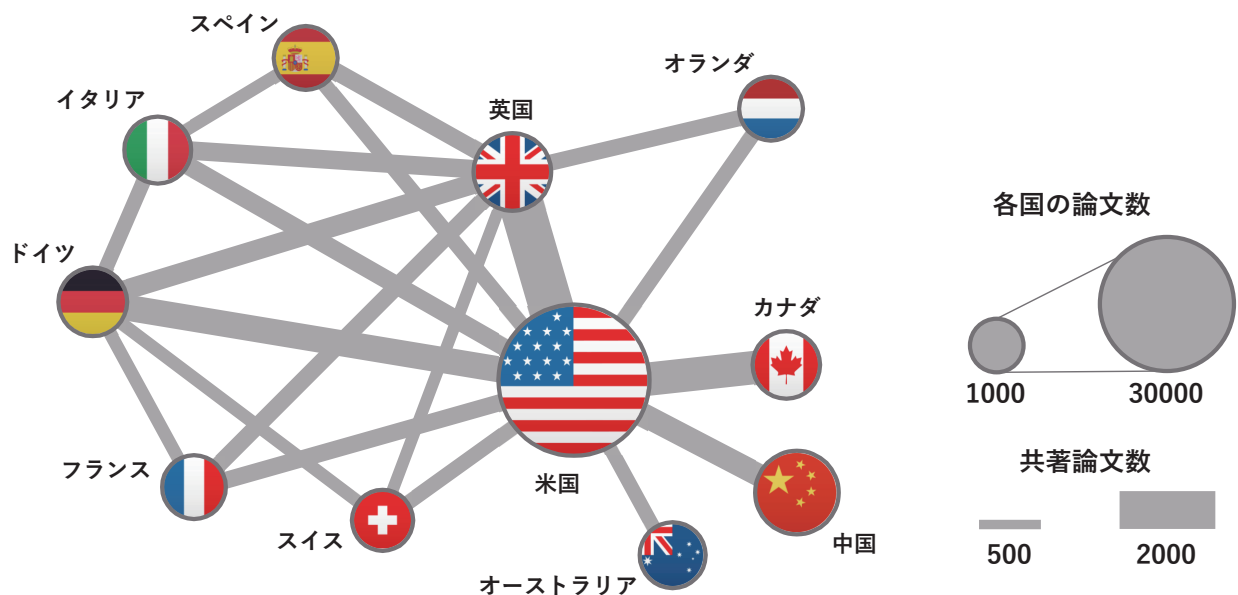
25 Lin, L. Y., Warren-Gash, C., Smeeth, L., & Chen, P. C. 2018. Data resource profile: the national health insurance research database (NHIRD). *Epidemiology and health*, 40.

図2-10 リアルワールドデータ関連論文の各国の共著関係の分布



出所：Web of Science®クラリベイトをもとに作成

図2-11 リアルワールドデータ関連論文の国際共著関係



出所：Web of Science®クラリベイトをもとに作成

一対一の共著論文数の上位20位までを抽出。各国の円の大きさに用いた論文数には単一国のみの著者による論文も含む。

最後に、リアルワールドデータ関連論文における国際連携の現状を確認する。複数国の機関の著者が含まれる論文の中で、同一論文内の著者が所属する国の組み合わせを調査した。共著論文数の上位20番目までの組み合わせを用いて、連携の状態を図示した結果を図2-11に示す。円の大きさは、国内の著者のみで構成された論文も含むすべての論文数を表し、線の太さは共著論文数を表している。リアルワールドデータ関連の研究は、米国、英国、中国が主導しているものの、国際連携の中心となっているのは米国と英国であることが分かる。中国は、上位国の中では他国との連携が比較的少なく、国内のみで研究を実施している傾向にあった。台湾、韓国、日本といった東アジアの国は論文数こそ多いものの、大半が国内著者のみの論文であるため、上位20番目（共著637報）までの連携に含まれる国は存在しなかった。日本と共に論

表2-13 日本の著者が関与するリアルワールドデータ関連論文の共著者の所在地

	著者の所在地	論文数
1	米国	546
2	英国	264
3	ドイツ	177
4	中国	159
5	イタリア	154
6	カナダ	139
7	フランス	127
8	スペイン	123
9	オーストラリア	118
10	オランダ	114

出所：Web of Science®クラリベイトをもとに作成

文を執筆している国の集計を表2-13に示す。最も多い国際共著の相手国は、546報の米国であり、2 国間の共著としては28番目に共著論文数が多い組み合わせであった。中国も上位に含まれているものの、上位国としては、英国、ドイツ、イタリア等の欧米の国が中心となっていた。

2.6 小括

2014年以降、医療機関、投薬、レセプなどの狭義のリアルワールドデータを利用した研究に関する論文は大幅に増加している。この増加傾向は特に製薬企業が関与して作成された論文に顕著であり、近年では製薬企業が関与する論文の10%以上がリアルワールドデータに言及しており、製薬産業における重要な情報源となってきた。製薬企業は、自社のパイプラインに関連するほぼすべての疾患領域でリアルワールドデータを利用した研究を実施しているが、その大半は米国を中心に実施された研究であった。この傾向はグローバル製薬企業のみならず、国内製薬企業においても見られており、米国のデータベースを用いて、米国の医療実態に則した研究成果が多く創出されている。この現象は、米国での事業が売上の大半を占める国内製薬企業のビジネスモデルに照らしても、妥当な結果であると考えられる。今後も、米国や英国などの拠点を中心に、利用価値の高いデータベースを活用することが製薬企業にとって重要であり、欧米でのエビデンスの創出や関連する解析技術の向上に注力することが、海外の製薬企業との競争力を維持するためにも不可欠であると考えられる。

一方で、海外のデータベースのみに基づいて医薬品の安全性や用法に関する意思決定がなされることは、日本国内の患者に対する情報提供の観点からは望ましくない可能性がある。日本国内の患者に適した正確な情報を提供するためには、少なくとも海外で実施されている研究と同様の研究が、国内でも実施可能となるようなデータベースの構築が不可欠である。国内の研究機関から発出されている論文数は国際的に見ても一定の存在感を示しているものの、欧米諸国と比較すると海外との共同研究の割合は少なく、各国独自で研究が実施されている現状がある。この現状の評価については一概に結論を出すことは難しいが、医薬品開発がグローバルに進んでいく中で、グローバル製薬企業が主導するリアルワールドデータに基づく意思決定においても、日本人患者に関連するデータが十分に反映される環境を整備することは、日本人患者に適した情報の提供にも重要であると考えられる。

3章 モバイルデータの利用の実態

3.1 製薬産業で利用されるモバイルデータの特徴

ウェアラブルデバイスや、生体センサーを搭載したスマートフォンが、社会に広く普及してきたことに伴い、製薬産業においてもこれらのデバイスから得られる情報の利用を検討する場面が増加してきている。これらのモバイルデータの最も大きな利点は、バイタルサインや身体活動、睡眠のパターン等の従来は局所的にしか取得できなかった患者の状態変化に関する情報を、断続的かつ客観的にリアルタイムで精緻に収集することができる点にある。製薬産業におけるモバイルデータの利用事例としては、バイタルサイン等の異常を検知することによる疾患の早期発見や早期の治療方法の提案、医薬品や医療技術評価におけるエンドポイントとしての利用、臨床試験における遠隔でのデータ収集などの用途に加えて、患者自身が疾患管理を行うことで行動変容を促し、治療効果を向上させるデバイスの開発も進められている。センサーやアプリケーションの開発は急速に進展しており、その進展に伴い、応用領域や対象疾患は変遷している。現時点では、一般的に普及し始めている加速度センサーやジャイロスコプ、圧力センサー等を搭載するデバイスや、スマートフォンで取得される患者の自身の健康状態の自己評価やアプリケーションを通じた疾患状態の追跡の情報を利用した研究が増加している。今回の調査では、製薬企業が関与する研究において、現時点で使用されているデバイスの活用実態について、学術論文を基に調査した。

3.2 製薬産業によるモバイルデータ利用の傾向

モバイルデータの利用を行った研究は限られるため、傾向の把握が可能なグローバル製薬企業が関与する研究に焦点を当てて調査を実施した。具体的には、デバイスを実際に使用した研究を特定するために、これまでに報告された複数のレビュー論文を参考にし、モバイルデータに関連する語句を整理した(“Wearable”、“Device”、“Smartwatch”、“Wristband”、“Fitness-bound”、“Fitness Band”、“Wireless Watch”、“Flexible Band”、“Smart Insole”、“Bracelet”、“Accelerometer”、“Gyroscope”、“Pedometer”、“Barometer”、“GPS”、“Humidity Sensor”、“Magnetometer”、“Compass”、“Bluetooth”、“Wi-Fi”、“Mobile”、“Smartphone”、“Smart Phone”、“Mobilephone”、“Mobile Phone”)^{26, 27, 28}。2014～2023年の10年間に出版された、グローバル製薬企業の社員が著者として関与した論文の中から、上記のモバイルデータに関連する語句がタイトル又は要旨で使用された論文を抽出した。抽出された論文から、要旨の記載を基に、実際に研究の中でモバイルデータを取得し、研究の結果を報告している149論文を特定した。これらの論文を今回の調査の中でモバイルデータ関連論文と定義した。

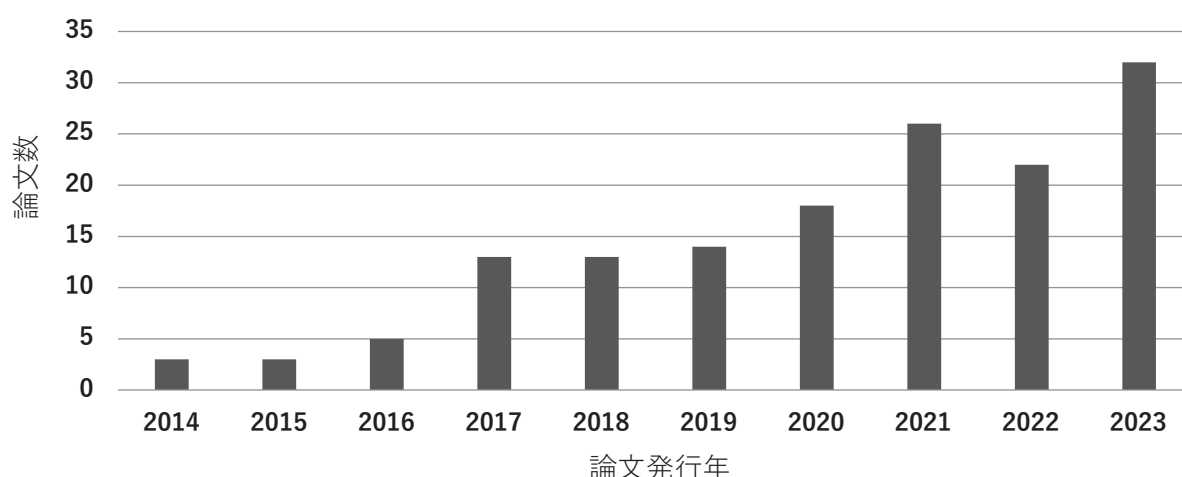
はじめに、モバイルデータ関連論文の出版数の年次推移を図3-1に示す。グローバル製薬企業におけるモバイルデータ関連論文は、2014年時点ではほとんど公表されておらず、モバイルデータを利用した研究は、近年徐々に拡大していることが確認できる。論文数自体は、2023年時点においても年間30報程度（1

26 Masoumian Hosseini, M., Masoumian Hosseini, S. T., Qayumi, K., Hosseinzadeh, S., & Sajadi Tabar, S. S. 2023. Smartwatches in healthcare medicine: assistance and monitoring; a scoping review. *BMC Medical Informatics and Decision Making*, 23(1), 248.

27 Straczekiewicz, M., James, P., & Onnela, J. P. 2021. A systematic review of smartphone-based human activity recognition methods for health research. *NPJ Digital Medicine*, 4(1), 148.

28 Rodriguez-León, C., Villalonga, C., Munoz-Torres, M., Ruiz, J. R., & Banos, O. 2021. Mobile and wearable technology for the monitoring of diabetes-related parameters: Systematic review. *JMIR mHealth and uHealth*, 9(6), e25138.

図3-1 グローバル製薬企業によるモバイルデータ関連論文数の推移



出所：Web of Science®クラリベイトをもとに作成

表3-1 グローバル製薬企業が関与したモバイルデータ関連論文の投稿先

	投稿先	論文数
1	JMIR mHealth and uHealth	11
2	Sensors	9
3	npj Digital Medicine	7
4	Journal of Medical Internet Research	6
5	American Heart Journal	4
5	PLOS ONE	4
5	Scientific Reports	4

出所：Web of Science®クラリベイトをもとに作成

表3-2 グローバル製薬企業が関与したモバイルデータ関連論文の研究分野

	研究分野	論文数
1	Health Care Sciences & Services	26
1	Medical Informatics	26
1	Neurosciences & Neurology	26
4	Cardiovascular System & Cardiology	15
5	Engineering	14

出所：Web of Science®クラリベイトをもとに作成

社平均3報程度)にとどまっており、グローバル製薬企業が関与する論文全体の中に占める割合は小さい。

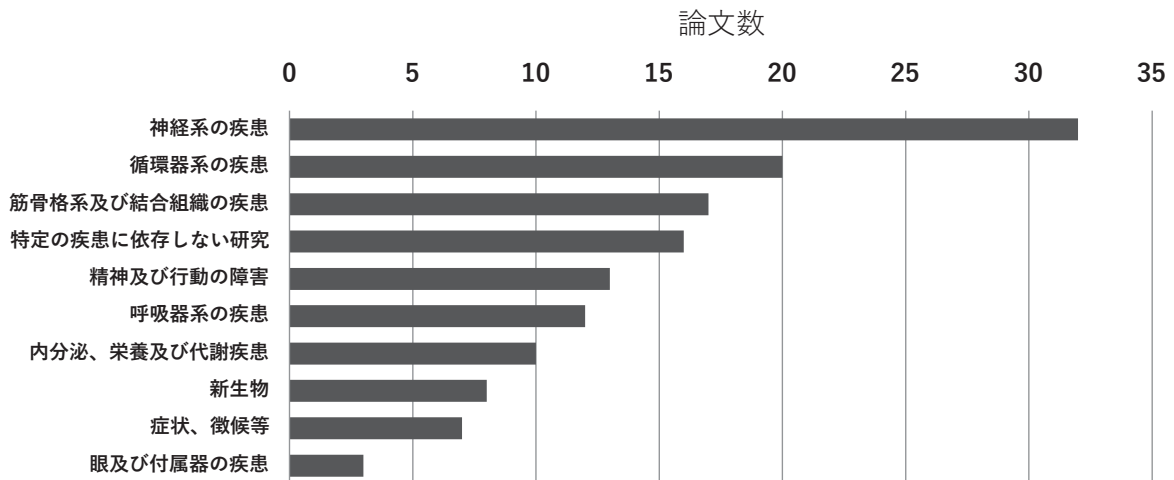
グローバル製薬企業のモバイルデータ関連論文の内容を確認するために、掲載された学術誌およびそれらの研究分野の頻度集計の上位をそれぞれ表3-1、表3-2に示す。リアルワールドデータ関連論文の投稿先で見られた傾向とは異なり、疾患領域に関連する学術誌で上位に位置するものは循環器疾患に関する学術誌のみとなり、医療情報やデジタル機器に関する学術誌が多く確認された。このことから、医療機関で取得される情報や請求データに由来するデータベースを利用する研究とは異なり、モバイルデータの臨床応用に関する報告は、デジタル関連に特化した学術誌で報告される基礎的な研究段階にあることが示唆される。研究分野の集計では、神経疾患および循環器疾患が上位に位置しており、これらの疾患から研究が着手されていることが分かる。

表3-3 グローバル製薬企業が関与したモバイルデータ関連論文のキーワード

	キーワード	論文数
1	Physical activity	14
1	Smartphone	14
3	Multiple sclerosis	10
4	Accelerometer	8
4	mHealth	8
4	Mobile phone	8

出所：Web of Science®クラリペイトをもとに作成

図3-2 グローバル製薬企業によるモバイルデータ関連論文数の対象疾患



出所：Web of Science®クラリペイトで抽出した情報を利用し、医薬産業政策研究所で情報を作成

続いて、論文のキーワードに用いられている語句について集計を行った。モバイルデータ関連論文において、高頻度で利用されていたキーワードを表3-3に示す。この結果から、身体活動が主要な測定対象となっていること、測定機器としてスマートフォンや加速度計が頻繁に使用されていること、疾患領域として多発性硬化症の研究が多くの論文で取り扱われていることなどが類推される。

3.3 モバイルデータ関連論文の特徴

グローバル製薬企業のモバイルデータ関連論文に関して、各論文の方法の項に記載されている情報を基に、研究に利用されたモバイルデータの傾向を調査した。論文内の記載から研究対象とされた疾患を特定し、ICD-10の章分類に従って分類を行った結果を図3-2に示す。最も多くの論文で研究対象とされた疾患領域は神経系の疾患であった。神経系の疾患は、表3-2で集計した研究分野の集計においても、上位に位置しており、表3-3で集計したキーワードに多く挙げられた多発性硬化症もこの分類に含まれる。次に多いのは循環器系の疾患であり、こちらも研究分野の集計において上位に位置した疾患であった。その後に、特定の疾患に依存しない研究や、筋骨格系の疾患、精神疾患が続いた。リアルワールドデータ関連論文と同様に、この分析で検出された疾患領域は各企業の製品や開発パイプラインと連動しているため、製薬企業の外部を含めたモバイルデータが利用された研究全体の傾向を示しているわけではないことに留意する必要がある。それぞれの疾患群を対象とした論文から、被引用数の多い論文を抽出し、使用されるデバイスの特徴把握を行った。

3.3.1 神経系の疾患

神経系の疾患に関する論文では、多発性硬化症とパーキンソン病を対象とした論文がそれぞれ11報と10報であり、これらの疾患を対象とする研究が大半を占めていた。その他には、アルツハイマー病、片頭痛、てんかんが続いた。多発性硬化症とパーキンソン病は共に身体機能の異常が症状に現れる疾患である。現時点で一般的に用いられているデバイスであるスマートフォンやウェアラブルデバイスの多くは、加速度センサーやジャイロセンサーをはじめとした慣性センサーを搭載しており、身体機能関連の情報（活動量、動作安定性、振戦等）の取得に適している。さらに、スマートフォンやタブレットでは、疾患情報を取得するための専用アプリケーションが開発されており、手の運動機能の情報や、音声認識による発話特徴の情報を取得する研究も行われている。これらのデバイスを介して測定される複数の指標は、多発性硬化症やパーキンソン病の症状の進行や変化をモニタリングするための手段として適しているため、他の疾患と比較して多くの研究が進められていると推察される。神経系の疾患で、最も多く使用されたデバイスは、スマートフォンであり、次いでリストバンド型の慣性センサー、腰装着型の慣性センサー、インソール型の圧力および慣性センサーが続いた。

対象となる論文の中で、最も被引用数が多い3つの論文を表3-4に示す。これらはすべて多発性硬化症又はパーキンソン病を対象とした、50人前後の少数規模の臨床研究であった。引用数が多い論文には、スマートフォン又はウェアラブルデバイスに搭載された慣性センサーを用いて運動徴候の情報を取得する論文が多く含まれており、客観的な評価および主観的な評価を継続的かつ遠隔で実施するための研究が実施されていた。

3.3.2 循環器系の疾患

循環器系の疾患に関する論文では、心不全と心房細動を対象とした論文がそれぞれ7報ずつであり、これらの疾患を対象とする研究が大半を占めていた。その他には、急性冠症候群、脳卒中などに関する研究が行われていた。心不全の疾患状態の評価には、定期的な身体活動と睡眠習慣の安定が QOL 指標として

表3-4 神経系の疾患領域での高被引用論文

	論文	被引用数	デバイスと主な測定項目
1	Evaluation of Smartphone-Based Testing to Generate Exploratory Outcome Measures in a Phase 1 Parkinson's Disease Clinical Trial	175	スマートフォン - 慣性センサー（加速度センサー、ジャイロセンサー、磁気センサー）による運動徴候 - マイクと録音による持続発声 - タッチスクリーンによる運動徴候
2	Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device	82	リストバンド型センサー - 慣性センサー（加速度センサー、ジャイロセンサー、磁気センサー）による運動徴候
3	Adherence and Satisfaction of Smartphone- and Smartwatch-Based Remote Active Testing and Passive Monitoring in People With Multiple Sclerosis: Nonrandomized Interventional Feasibility Study	81	スマートフォン - 質問票の回答による QOL 調査 - アプリケーションによる認知テスト - 慣性センサー（加速度センサー、ジャイロセンサー、磁気センサー）による運動徴候 - タッチスクリーンによる運動徴候 スマートウォッチ - 慣性センサー（加速度センサー、ジャイロセンサー、磁気センサー）による運動徴候

出所：Web of Science®クラリベイトで抽出した論文に医薬産業政策研究所で調査した内容を追加して作成

表3-5 循環器系の疾患領域での高被引用論文

	論文	被引用数	デバイスと主な測定項目
1	Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation The mSToPS Randomized Clinical Trial	327	パッチ型心電図モニター －ECG 記録による心房細動、頻脈等の電気信号異常
2	Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: A randomized study	124	スマートフォン －アプリケーションによる服薬管理と患者指導（運動、体重、喫煙）による生活習慣の改善
3	OUTSTEP-HF: randomised controlled trial comparing short-term effects of sacubitril/valsartan versus enalapril on daily physical activity in patients with chronic heart failure with reduced ejection fraction	60	リストバンド型センサー －慣性センサー（加速度センサー、ジャイロセンサー、磁気センサー）による身体活動

出所：Web of Science®クラリベイトで抽出した論文に医薬産業政策研究所で調査した内容を追加して作成

用いられており、これらの指標の改善が、医薬品評価のエンドポイントとしても利用されている。加速度センサーやジャイロセンサーをはじめとした慣性センサーは、これらの活動量の情報の取得に適している。また、心房細動は心電図で検知されるため、パッチ型心電図モニターによる症状検知に関する研究が多く行われていた。循環器系の疾患で使用されたデバイスは、疾患ごとに特徴が見られ、心不全を対象とした研究では、手首に装着する慣性センサーが主に使用されていた一方で、心房細動を対象とした研究では、胸部に装着するパッチ型心電図モニターが主に使用されていた。

対象となる論文の中で、最も被引用数が多い3つの論文を表3-5に示す。それぞれ、心房細動、心筋梗塞、心不全を対象とした臨床研究であった。全て100人以上の被験者数が参加する中規模以上の研究であり、パッチ型心電図モニターや慣性センサーの使用は、研究段階から実用化を検証する段階に移行しつつあることが分かる。また、疾患の異常検知や状態把握を目的としたデバイスの使用だけでなく、スマートフォンを活用した行動変容を促す生活習慣指導のアプリケーションの研究開発にも応用されており、これらの研究にも製薬企業の関与が確認された。

3.3.3 筋骨格系及び結合組織の疾患

筋骨格系及び結合組織の疾患に関する論文では、関節リウマチと膝関節症を対象とした論文がそれぞれ、7報と5報であり、これらの疾患を対象とする研究が大半を占めていた。その他には、体軸性脊椎関節炎、骨粗鬆症、痛風などに関する研究が行われていた。関節リウマチは、症状の経過が多様であり、改善と悪化の周期的な変動を伴うことも多い疾患であるため、継続した疾患状態の追跡が疾患管理に重要となる。関節リウマチは、臨床試験においても患者報告のアウトカムが主要な評価項目として使用されており、スマートフォンを用いてこれらのアウトカムを継続的かつ遠隔で収集する研究が多数存在していた。また、こわばりや疼痛が発生した際にリアルタイムで記録を行うことで、想起バイアスを軽減する効果も期待されている。膝関節症に関しても、関節リウマチと同様に、スマートフォンを用いて疼痛の報告を行った研究や、足や腰に慣性センサーを装着して、歩行能力や起立動作など歩行障害の進行度を評価する研究が行われていた。

対象となる論文の中で、最も被引用数が多い3つの論文を表3-6に示す。それぞれ、変形性膝関節症、関節リウマチ、痛風を対象とした臨床研究であった。全ての研究で、スマートフォンアプリケーションを用いて患者の自己報告によるアウトカムを収集しており、定期的に疼痛が生じる疾患が多い筋骨格系及び

表3-6 筋骨格系及び結合組織の疾患領域での高被引用論文

	論文	被引用数	デバイスと主な測定項目
1	Assessing the Impact of a Novel Smartphone Application Compared With Standard Follow-Up on Mobility of Patients With Knee Osteoarthritis Following Treatment With Hylan G-F 20: A Randomized Controlled Trial	65	リストバンド型センサー －歩数計 －慣性センサーによる睡眠の質、時間 スマートフォン －質問票の回答による QOL 調査 －リストバンド型センサーとアプリケーションの連携による自己情報の閲覧
2	Using a ResearchKit Smartphone App to Collect Rheumatoid Arthritis Symptoms From Real-World Participants: Feasibility Study	53	スマートフォン －質問票の回答による QOL 調査 －慣性センサー（加速度センサー、ジャイロセンサー）による手首関節の屈伸評価
3	What Outcomes are Important for Gout Patients? In-Depth Qualitative Research into the Gout Patient Experience to Determine Optimal Endpoints for Evaluating Therapeutic Interventions	40	スマートフォン －発作時のリアルタイム症状報告

出所：Web of Science®クラリベイトで抽出した論文に医薬産業政策研究所で調査した内容を追加して作成

結合組織の疾患領域においては、リアルタイムかつ遠隔で症状を記録する手法に関する研究が重点的に行われている。

3.4 被引用数上位の論文から見る利用の特徴

製薬企業によるモバイルデータを用いた研究事例からも、ウェアラブルデバイスや、生体センサーを搭載したスマートフォンが、従来は測定が不可能であった生体情報の捕捉を可能にすることと、既存の指標取得の簡易化や追跡性の向上を推進することの両面で大きく貢献していることが分かる。特に、前者の従来の技術では測定が不可能であった生体情報を捕捉する研究においては、慣性センサーに由来する情報が中心的な役割を果たしており、取得された情報を処理するアルゴリズムに関する研究も並行して実施されている。短時間の使用事例としては、歩行リズム、立ち上がり動作、起居動作、転倒検知などの身体活動を直接評価することが可能となりつつあり、長時間の使用事例としては、日中の活動量、活動強度、運動活動時間、睡眠時間、睡眠パターンなど、QOLと密接に関連する指標を定量化する研究が進んでいる。特に、後者の長時間の活動パターンの把握は、医薬品開発における治療効果の評価に加え、QOL評価の重要性が高まる中で、患者の日常生活活動の改善に関連する指標として重要な役割を果たすことが期待される。製薬産業における具体的な利用場面の事例として、慢性疾患の症状改善や加齢に伴う運動障害、心肺機能の低下、術後回復の定量的評価などが挙げられ、これらの用途において今後利用頻度が増加することが予測される。また、既存の指標取得の簡易化や追跡性の向上に関する研究では、スマートフォンやスマートウォッチで利用可能な疾患独自のアプリケーション開発が複数進められていた。これらのアプリケーションの使用により、患者が医療機関に来院することなく定期的に自己報告や疾患状態に関するデータを収集できるため、より正確な疾患状態の把握や医薬品評価の精度向上に寄与することが期待される。さらに、分散型臨床試験やオンラインでの臨床試験の導入が検討される中で、遠隔での疾患状態の把握は今後より必要性が増加することが想定され、関連する技術の開発が加速する分野であると考えられる。

3.5 国際的なウェアラブル技術関連論文の実態

製薬企業にとってセンサーやアプリケーションの開発は、従来の専門領域と異なるため、他産業や学術機関との協業が強く求められる。これまでに、大手の製薬企業とヘルステック企業の提携は多くなされており、国内の製薬企業においても、大塚製薬と Verily Life Sciences 社との提携、中外製薬と Biofourmis 社の提携、マルホと Sibel Health 社の提携など製薬企業におけるモバイルデータの臨床応用には、外部組織の研究開発の進展が極めて重要となっている。本項では、それらのモバイルデータ関連の基礎研究および応用研究の全体的な潮流を把握することを目的として、製薬企業が関与していない研究も含めて、特に国際的な研究の実施に焦点を当てて調査を行った。モバイルデータ関連論文の抽出に使用された語句はヘルスケア分野のみではなく、広域な分野の論文で使用されるため、今回は製薬産業において最も頻繁に使用されるウェアラブル技術に限定して、技術開発が行われている国について調査を行った。今回の調査では、タイトル又は要旨にウェアラブル技術の論文で多く利用される表現（“Wearable Device”、“Wearable Technology”、“Wearable Sensor”、“Wearable Health”）を含む2014～2023年の10年間に出版された原著論文をウェアラブル技術関連論文として調査した。

これらの論文の著者構成について、著者の所在地と著者が所属する組織を表3-7と表3-8に示す。複数の国や所属の著者で共著となっている論文は、それぞれの国、組織で集計を行った。多くの分野の論文数で上位となる米国、中国、英国はウェアラブル技術の分野においても上位に位置していたが、韓国やイタリア、インドなどもこの分野で多くの論文を発出しており、研究を主導していることが分かる。また、著者が所属する組織に関しても、カリフォルニア大学システム（米国）、ハーバード大学（米国）、ロンドン大学（英国）といった他の分野の集計でも上位に位置する組織に加えて、中国科学院（中国）やソウル大

表3-7 ウェアラブル技術関連論文の著者の所在地

	著者の所在地	論文数
1	中国	1,363 (26.7%)
2	米国	1,319 (25.8%)
3	韓国	481 (9.4%)
4	英国	443 (8.7%)
5	イタリア	290 (5.7%)
6	インド	257 (5.0%)
7	日本	256 (5.0%)
8	オーストラリア	249 (4.9%)
9	カナダ	244 (4.8%)
10	台湾	216 (4.2%)

出所：Web of Science®クラリベイトをもとに作成

表3-8 ウェアラブル技術関連論文の著者所属組織

	著者が所属する組織	論文数
1	Chinese Academy of Sciences	166 (3.2%)
2	University of California System	143 (2.8%)
3	Harvard University	81 (1.6%)
4	University of London	80 (1.6%)
5	Seoul National University	59 (1.2%)
77	University of Tokyo	25 (0.5%)

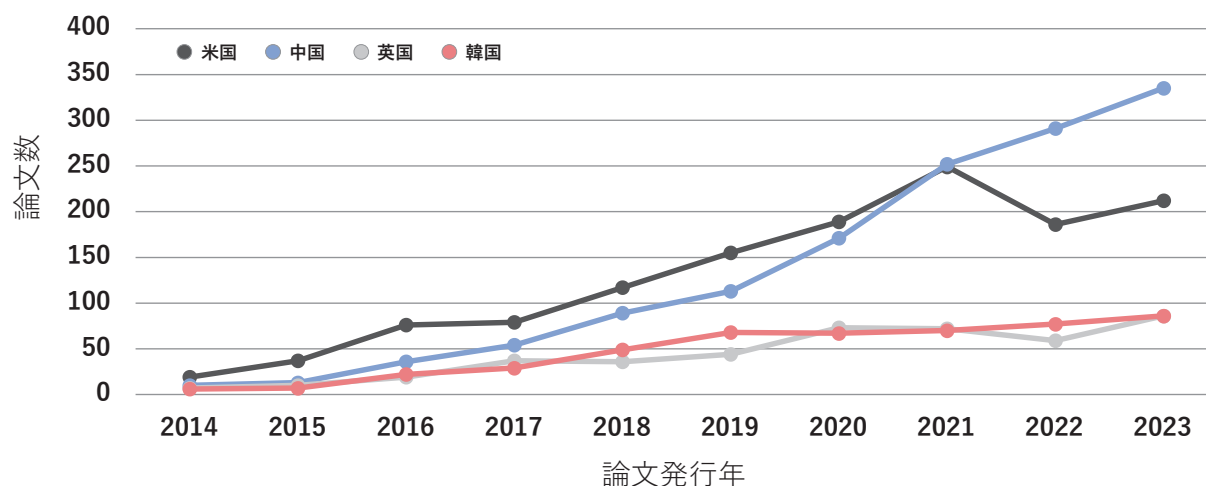
出所：Web of Science®クラリベイトをもとに作成

学校（韓国）が上位に入っており、出版された論文数が多い国では、その研究活動の中心となる組織が国内に存在していることがわかる。日本は全体で7番目に位置しており、最も多くの論文を発出した機関は東京大学であった。

論文数が上位になった国の出版数の年次推移を図3-3に示す。ウェアラブル技術関連論文数は、製薬企業が関与した論文の傾向と同様に、2014年時点ではいずれの国でもほとんど研究が実施されていなかったものの、近年は継続的に増加しており、中国と米国が牽引している。

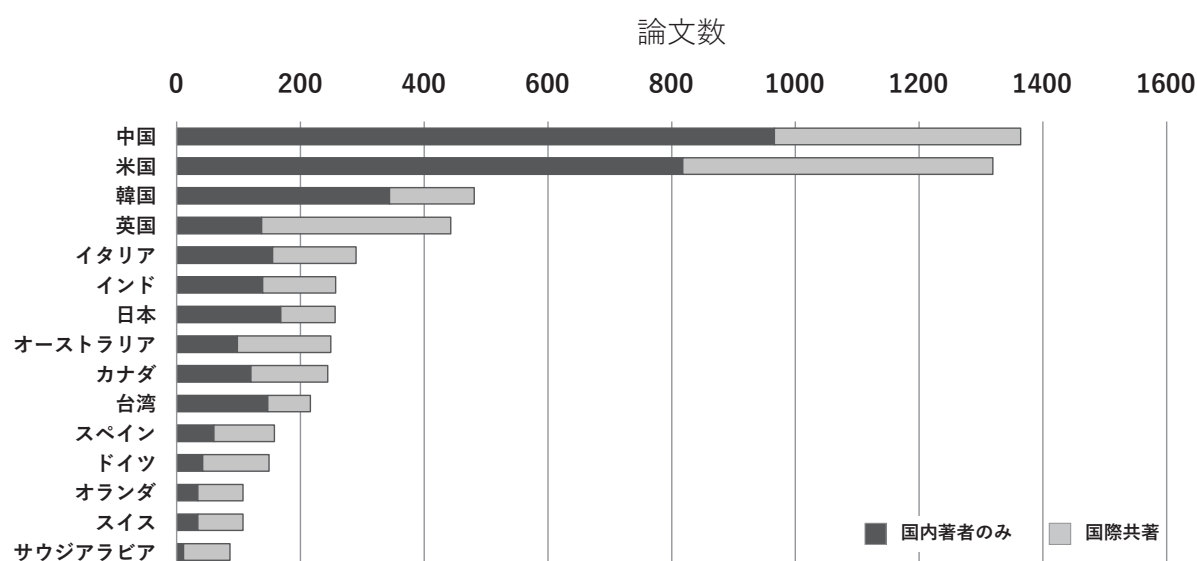
ウェアラブル技術関連論文における国境を跨いだ共同研究について調査をするために、2014～2023年までの10年間に出版された論文の著者構成に関して、同一国の著者のみで構成されている論文と、複数国の著者が含まれる論文を分類して国ごとに集計を行った結果を図3-4に示す。また、当該国の研究機関に所属する研究者が一人でも含まれている論文の数を分母として、各論文の著者が所属する組織の国数の分布を図3-5に示す。論文数が上位である中国と米国では、国際共著論文（2か国以上の著者で構成される論

図3-3 ウェアラブル技術関連論文数の各国の推移



出所：Web of Science®クラリベイトをもとに作成

図3-4 ウェアラブル技術関連論文数と各国の国際共著

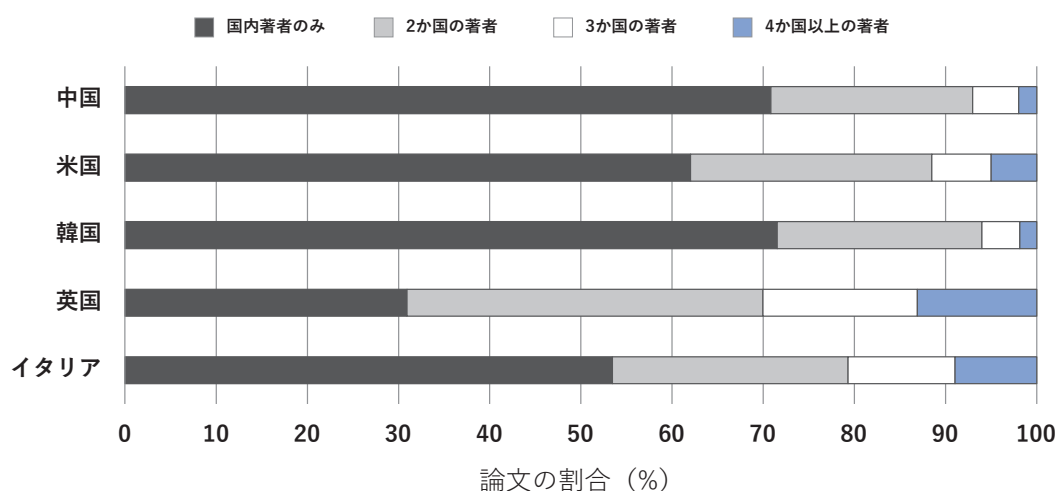


出所：Web of Science®クラリベイトをもとに作成

文)の割合がそれぞれ3～4割程度であった。それらに続く、韓国は3割程度であり、リアルワールドデータ関連論文と比較すると、国際共著論文の割合は高い傾向にあった。

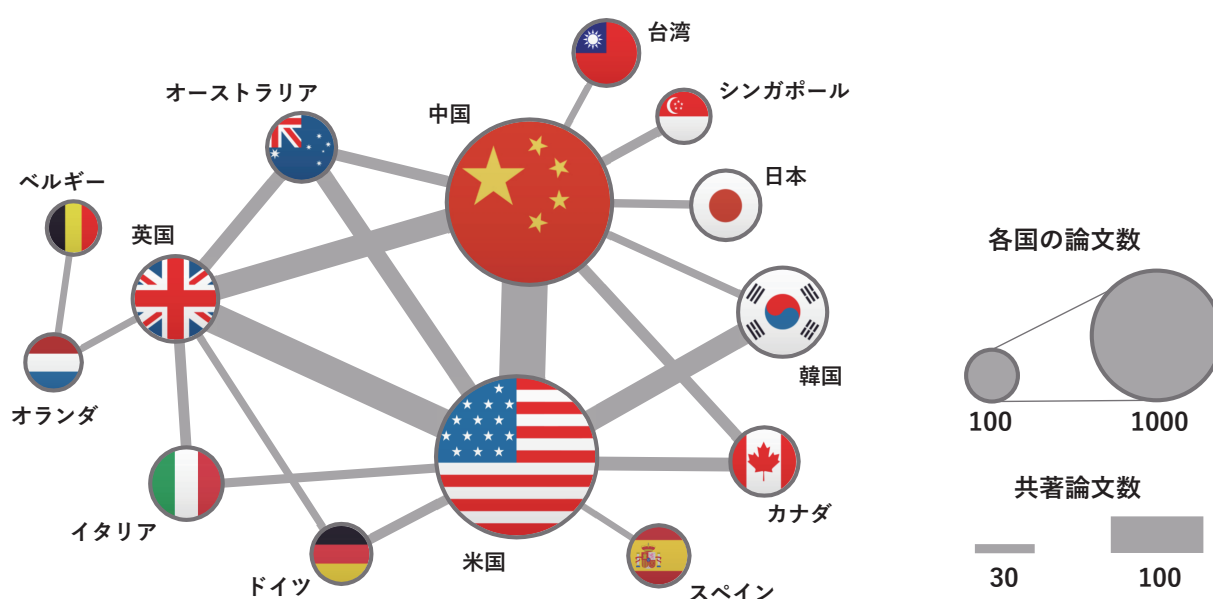
最後に、ウェアラブル技術関連論文における国際連携の現状を確認する。複数国の機関の著者が含まれる論文の中で、同一論文内の著者が所属する国の組み合わせを調査した。共著論文数の上位20番目までの組み合わせを用いて、連携の状態を図示した結果を図3-6に示す。ウェアラブル技術に関連する研究も、リアルワールドデータ関連論文と同様に中国、米国、英国が主導しており、これらの国が国際連携の中心も担っていることが分かる。リアルワールドデータ関連論文では、中国は相対的に国際共著論文が少ない傾向にあったが、ウェアラブル技術関連論文においては、米国を除く他国と比較して圧倒的に多くの論文を発行していることもあり、国際連携においても存在感が大きいことが分かる。さらに、欧州地域の国々は英国を中心に連携が行われており、アジア地域の国々は中国を中心に連携が進んでいる。このように、

図3-5 ウェアラブル技術関連論文の各国の共著関係の分布



出所：Web of Science®クラリペイトをもとに作成

図3-6 ウェアラブル技術関連論文の国際共著関係



出所：Web of Science®クラリペイトをもとに作成
 一対一の共著論文数の上位20位までを抽出。各国の円の大きさに用いた論文数には単一国のみの著者による論文も含む。

地理的に近接する地域での国際共著論文の割合が高くなっていることもこの分野の特徴として挙げられる。ウェアラブル技術関連論文の国際共著論文数は、多い国間でも10年間で120報程度となっており、限られた論文数の中での集計となっていることには留意が必要である。

3.6 小括

製薬企業におけるモバイルデータに関連する研究は、現時点での報告数は限られているものの、2010年代後半から増加傾向にある。特に、スマートフォンやウェアラブルデバイス、スマートウォッチに搭載された慣性センサーは、疾患状態の把握において広く使用されている。センサーやアプリケーションの開発は急速に進展しており、その進展に伴い、応用領域や対象疾患は変遷している。特定の疾患領域に対応したデバイスも登場しており、例えば、パッチ型の心電図モニターや、持続血糖モニター、服薬管理機能が付いた吸入器などが研究に使用されている。医薬品の評価においては、精度や再現性が高い評価手法が求められるため、定量的かつ客観的なデータを取得可能にする技術は、特に高い親和性を持つと考えられる。

一方で、現時点ではモバイルデータの利用には解決すべき課題がいくつか存在している。まず、最も重要な課題の一つとして、データの精度と信頼性の向上が求められる。現状使用されているセンサーは自身で着脱可能なものが多く、装着方法や着脱が測定結果に影響を及ぼす可能性があり、精緻な評価が困難となることが懸念されている。また、それに伴う課題として、得られたデータの正確な解釈方法に関する研究も進められている最中であり、具体的には、実際の活動をより精密に表現するアルゴリズムの開発や、センサーフュージョン技術の活用など様々な検討がなされている。測定機器自体に関しても、測定機器のサイズや防水性、バッテリーの駆動時間などの技術的課題も挙げられ、これらの改善は製薬業界におけるモバイルデータの実用化には不可欠となる。

製薬産業においてモバイルデータを活用していくためには、他産業や学術機関との協業が必須であり、既に複数の製薬企業が連携を始めている状況である。現時点では、モバイルデータに関連する研究の連携の中心は米国であり、国内の製薬企業が連携を進めている相手も米国のヘルステック企業が多い傾向にある。ウェアラブル技術関連論文の中には、日本に本社を置く製薬企業がスポンサーとなる研究も数件確認されているが、その大半が米国の拠点が中心となり実施されていることが著者の構成からも推察される^{29,30}。前項で、国内の製薬企業が関与するリアルワールドデータ関連の研究は、海外の拠点を中心に実施されている研究が多いことについて触れたが、リアルワールドデータ関連の研究で見られた傾向以上に米国で研究が実施される割合が高くなっている可能性がある。この分野では、中国や韓国などの東アジア地域の国も米国との共同研究を数多く実施しており、グローバルな視点から見てもその存在感が大きいことが確認された。一方で、日本の論文数や海外との共同研究数は、他の東アジアの上位国と比較すると少ない状況にある。国内の製薬企業は、モバイルデータを利用した疾患状態の把握や医薬品の評価を通して新たな知見を導出するために、米国を中心とした国際的な技術進展を注視しつつ、米国や他の東アジア諸国の研究機関や企業との連携強化や投資を継続的に検討していくことが、今後の国際競争力を維持していく上で重要であると考えられる。

29 Peters-Strickland, T., Pestreich, L., Hatch, A., Rohatagi, S., Baker, R. A., Docherty, J. P., ... & Walling, D. P. 2016. Usability of a novel digital medicine system in adults with schizophrenia treated with sensor-embedded tablets of aripiprazole. *Neuropsychiatric Disease and Treatment*, 2587-2594.

30 Cormack, F., McCue, M., Taptiklis, N., Skirrow, C., Glazer, E., Panagopoulos, E., ... & Barnett, J. H. 2019. Wearable technology for high-frequency cognitive and mood assessment in major depressive disorder: longitudinal observational study. *JMIR mental health*, 6(11), e12814.

4章 大規模ゲノムデータの利用の実態

4.1 製薬産業で利用される分子プロファイル情報の特徴

分子プロファイル情報の解析技術の進展とそのコスト低下に伴い、大規模に収集されたこれらのデータを創薬プロセスに適用するための研究が増加している。特に、ゲノミクス、プロテオミクス、メタボロミクスなどのオミックスデータは、創薬や最適な治療法の提案において重要な役割を果たすことが期待されている。加えて、分子プロファイル情報は、電子カルテや請求データ等の医療機関における診療の情報や、生活習慣に関する情報と組み合わせることによって、疾患メカニズムの解明、疾患や体質に関する遺伝的な関連性の特定、薬物標的の同定、さらには個別化医療の推進に貢献すると考えられている。本邦においても、これらのデータを効果的に活用するための基盤整備が進められており、健康・医療戦略本部の健康・医療データ利活用基盤協議会や、ゲノム医療協議会などの場で、大規模なゲノム情報やリアルワールドデータを活用するためのプラットフォーム構築について議論が行われている³¹。

特にゲノムデータは、全ゲノムおよび全エクソームシーケンシング技術や遺伝子パネル検査の技術的進展により、大規模なデータの収集と解析が可能となった。この技術的進展に伴い、各国でヒトゲノムに関する複数のプロジェクトが進行しており、ゲノムデータへのアクセスはこれまで以上に容易になってきている。中でも、製薬企業がアクセスしやすい大規模なゲノムデータとして、バイオバンクやゲノムデータベースが注目されており、これらのデータが新たな創薬研究の情報源として活用され始めている。表1-4で示したゲノム解析のグループにおける頻出語句の集計結果の上位に、UK Biobankの名称が挙がっているように、既にグローバル上位製薬企業において、関連する研究が多数実施されていることが分かる。これらのデータベースの中でも、英国のUK BiobankやフィンランドのFinnGenなどのバイオバンクに由来する情報や、米国のThe Cancer Genome Atlas Program (TCGA)などの疾患コホート研究は、特に製薬企業からのアクセスが多い情報源として知られている。これらの分子プロファイルのデータベースでは、ゲノムデータをはじめとする多様なオミックスデータや臨床情報が大規模に収集され、解析されており、製薬企業との連携を通じて創薬研究に利用されている。特に、健康な個人に関する情報が豊富なバイオバンクは、生殖細胞系列変異に関連する遺伝的疾患の解析に強みを持つ一方で、悪性腫瘍を対象としたTCGA等の疾患コホートは、体細胞変異やがん細胞の遺伝子発現に関連する研究において重要な役割を果たしている。今回の調査では、製薬企業によるバイオバンクへの投資が増加している現状を背景に、今後さらに利用の拡大が予想される大規模バイオバンクで管理されるゲノムデータを中心に、その利用実態について調査を行った。

4.2 製薬産業によるバイオバンクに由来する情報の利用の傾向

製薬企業によるバイオバンクに由来する情報の利用傾向を確認するために、2014～2023年にかけて出版された、グローバル製薬企業の社員が著者として関与した論文の中から、タイトル又は要旨の中に“Bio-bank”という単語を含む論文を抽出した。その中で、要旨の記載を基に、実際にバイオバンクに由来する分子プロファイル情報（試料の利用や質問票の情報のみを利用した研究は除く）を利用し、その研究結果

31 内閣官房 健康・医療戦略室、第10回 健康・医療データ利活用基盤協議会 第12回 ゲノム医療協議会 資料4-3、https://www.kantei.go.jp/jp/singi/kenkouiryou/data_rikatsuyou/dail0/gijisidai.html

を報告している論文を特定した。調査対象として抽出された論文は104論文であり、これらの論文を今回の調査の中でバイオバンク関連論文と定義した。論文の執筆時にタイトルと要旨に使用できる文字数に制限があるため、情報源として単にデータベースの名称のみが記載されている論文も存在しており（FinnGen等）、本調査で抽出された論文は、データベース名の中に“Biobank”という単語が明示的に含まれるバイオバンク（UK Biobank 等）の情報を利用した研究を報告する論文に偏っている可能性があることに留意する必要がある。

バイオバンク関連論文が掲載された学術誌およびその研究分野の頻度集計の上位をそれぞれ表4-1、表4-2に示す。現時点でバイオバンクにおいて利用可能となっているオミックス関連情報はゲノムデータが中心となっているため、特に遺伝学に特化した学術誌への投稿が多く見られた。また、これらの情報を利用した研究は基礎研究の分野とどまらず、臨床医学系の学術誌においても結果が報告されており、特に個別化医療や精密医療の分野における臨床応用に関する研究が実施されている。さらに、一部の論文はインパクトファクターの高い学術誌に掲載されており、その背景には製薬企業からの資金提供によって測定されたデータに対して、資金提供者である製薬企業が優先的なアクセス権を有していることが影響していると考えられる。一部の大規模なバイオバンクでは、上記のように製薬企業からの資金提供を受けて試料の測定が実施されており、今回の調査対象に含まれる製薬企業も各プロジェクトに複数社が参加している。英国のUK Biobankでは、8社の製薬企業（AbbVie、Alnylam Pharmaceuticals、AstraZeneca、Biogen、Bristol-Myers Squibb、Pfizer、Regeneron、武田薬品工業）が全エクソーム配列データの測定を行うプロジェクトを実施しており、13社の製薬企業（Alnylam Pharmaceuticals、Amgen、AstraZeneca、Biogen、Bristol Myers Squibb、Calico、Genentech、GlaxoSmithKline、Janssen Pharmaceutical Companies、Novo Nordisk、Pfizer、Regeneron、武田薬品工業）が血漿プロテオームを測定するプロジェクトを実施している。フィンランドのFinnGenも、現時点で14社が資金提供研究パートナーとして参加する官民連携プロジェクトとして知られている。

続いて、論文内で使用されている表現（単語のBigram）について集計を行った。バイオバンク関連論

表4-1 グローバル製薬企業が関与したバイオバンク関連論文の投稿先

	投稿先	論文数
1	Nature Genetics	11
2	Nature Communications	8
3	The American Journal of Human Genetics	7
4	Circulation: Genomic and Precision Medicine	5
4	Nature	5

出所：Web of Science®クラリベイトをもとに作成

表4-2 グローバル製薬企業が関与したバイオバンク関連論文の研究分野

	研究分野	論文数
1	Genetics & Heredity	30
2	Science & Technology - Other Topics	24
3	Biochemistry & Molecular Biology	10
4	Cardiovascular System & Cardiology	9
5	Endocrinology & Metabolism	8
5	General & Internal Medicine	8
5	Neurosciences & Neurology	8

出所：Web of Science®クラリベイトをもとに作成

表4-3 グローバル製薬企業が関与したバイオバンク関連論文中で使用された表現

	頻出表現	論文数
1	Genome Wide	53
2	Wide Association	52
3	Association Study	48
4	Odd Ratio	25
5	Genetic Variant	20
6	Meta Analysis	19
7	Risk Factor	17
8	Genetic Risk	16
9	Rare Variant	15
10	Exome Sequence	14
10	Mendelian Randomization	14

出所：Web of Science®クラリベイトをもとに作成
グループ内の論文で使用されている表現（単語の Bigram）を基に作成した。ストップワード又は Biobank という単語が含まれる表現は集計対象外とし、語形変化を問わない形で集計を行った。

文において、高頻度で利用されていた表現を表4-3に示す。この結果から、Genome-wide association study (GWAS) に関連する研究が最も多く行われていることが確認され、製薬企業が関与する研究では、遺伝的要因に関連する疾患発症のリスク因子の特定などを目的として生殖細胞系列変異に関する調査が多く行われていることが確認された。さらに、製薬企業が関与する研究において、大規模なエクソーム配列データを利用した研究も始まっていることが確認された。

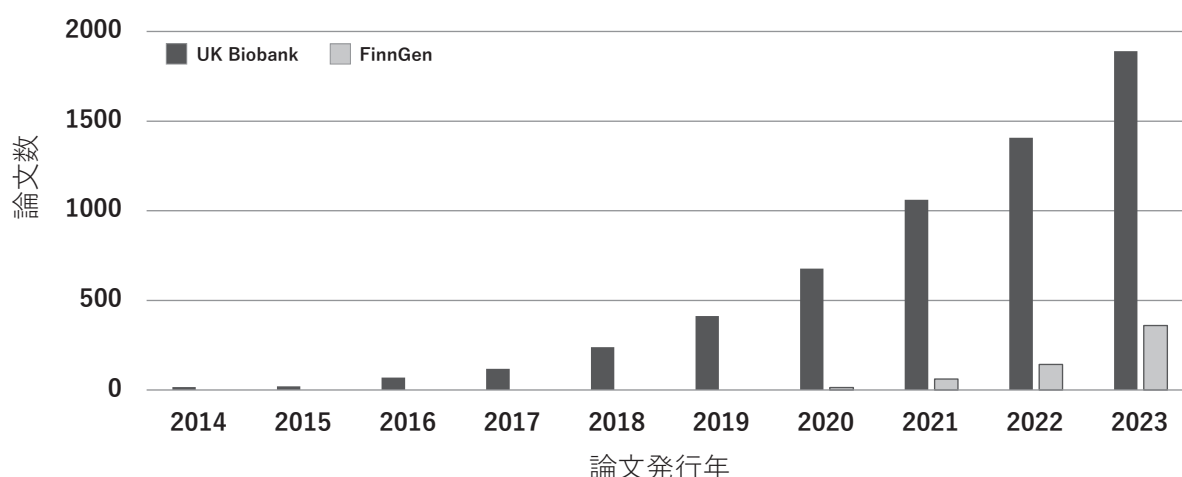
4.3 バイオバンクに関する研究の特徴

前項で触れた製薬企業からの資金提供の影響もあり、UK Biobank と FinnGen はグローバルの上位製薬企業による利用が特に多いバイオバンクとして知られている。前項で実施したバイオバンク関連論文の調査においても、大多数の研究が両者のバイオバンクに由来する情報を基に行われていたことが確認されている。製薬企業の利用が集まっているバイオバンクの特徴について調査を行うため、これら二つのバイオバンクに関連する全ての研究について、製薬企業が関与していない研究も含めて調査を行った。タイトル又は要旨の中に“UK Biobank” (“United Kingdom Biobank” を含む) 又は“FinnGen” という語句含む2014～2023年の間に出版された原著論文を抽出した。論文抽出における限界として、バイオバンクから提供される情報を利用した論文のみを抽出することが不可能であったため、本項の調査ではバイオバンクから分譲された生体試料を用いて実施された研究も含まれることに留意が必要である。

UK Biobank と FinnGen について言及した論文の出版数の年次推移をそれぞれ図4-1に示す。UK Biobankについて言及した論文は、2014年以降増加しており、現在に至るまで増加傾向が継続している。FinnGen に関しても、データのダウンロードが可能となった2020年以降、FinnGen について言及した論文の数は増加を続けている。

続いて、論文内で使用されている表現 (Bigram) の集計を行った。それぞれのバイオバンクに言及した論文において、高頻度で利用されていた表現を表4-4に示す。これらの結果から、双方のバイオバンクの情報を生体試料を用いた研究において、GWAS やメンデルのランダム化解析が多く実施されていることが確認できる。UK Biobank では、研究者は非識別化された大部分の個人レベルの情報へのアクセスが可能である一方で、FinnGen では、現時点ではプロジェクトに参画している組織の研究者のみが非識別化された個人レベルの情報にアクセス可能となっている。そのため、広く利用可能な情報は、GWASの結果などの要約さ

図4-1 各バイオバンクに言及した論文数の推移



出所：Web of Science®クラリベイトをもとに作成

表4-4 各バイオバンクについて言及した論文中使用された表現

UK Biobank			FinnGen		
	頻出表現	論文数		頻出表現	論文数
1	UK Biobank	5,817	1	Mendelian Randomization	464
2	Genome Wide	1,874	2	Genome Wide	411
3	Wide Association	1,743	3	Wide Association	384
4	Association Study	1,633	4	Association Study	362
5	Mendelian Randomization	1,248	5	Randomization Study	311
6	Confidence Interval	1,246	6	Inverse Variance	298
7	Risk Factor	1,171	7	Variance Weight	298
8	Odd Ratio	966	8	UK Biobank	268
9	Hazard Ratio	950	9	Odd Ratio	258
10	Cohort Study	933	10	Sample Mendelian	217

出所：Web of Science®クラリベイトをもとに作成

論文で使用されている表現 (Bigram) を基に作成した。ストップワードが含まれる表現は集計対象外とし、語形変化を問わない形で集計を行った。

れた情報に限定されており、今回の調査結果に含まれる論文には、これらの要約情報を利用した情報を利用した研究が多く含まれている。このようなアクセス制限も影響し、FinnGenについて言及した論文では、UK Biobank について言及した論文と比較しても、要約情報を利用したメンデルのランダム化解析やメタアナリシスといった表現が上位に見られる結果となっている。

UK Biobank と FinnGen について言及した論文の著者の所在地を表4-5に示す。バイオバンクに関連した研究は、米国、英国、中国、オーストラリア、スウェーデン、オランダ、ドイツ等で数多く実施されており、この傾向は UK Biobank と FinnGen の双方で大きな差異は見られない。FinnGen に言及する論文においては、中国の著者が多いことが確認されており、これは FinnGen のデータが近年利用可能となったことと関連していると考えられる。この点については、後述する大規模ゲノムデータ関連論文の実態に関する調査において言及する。

バイオバンクに関連した研究の特徴を明確にするため、悪性腫瘍領域に特化したデータベースである TCGA について言及した論文に関する調査結果を参考情報として示す。2014～2023年の10年間に出版され

表4-5 各バイオバンクについて言及した論文の著者の所在地

UK Biobank			FinnGen		
	著者の所在地	論文数		著者の所在地	論文数
1	英国	2,561	1	中国	407
2	米国	2,290	2	米国	117
3	中国	1,713	3	英国	80
4	オーストラリア	657	4	スウェーデン	79
5	スウェーデン	545	5	フィンランド	69
6	オランダ	513	6	ドイツ	21
7	ドイツ	477	7	オーストラリア	17
8	カナダ	369	8	オランダ	17
9	フランス	274	9	エストニア	16
10	デンマーク	258	10	デンマーク	13

出所：Web of Science®クラリベイトをもとに作成

表4-6 グローバル製薬企業が関与した TCGA に関する論文中で使用された表現

	頻出表現	論文数
1	Cancer Genome	20,804
2	Gene Expression	7,675
3	Enrichment Analysis	4,828
4	Differentially Expressed	4,413
5	Cox Regression	4,338
6	Expression Omnibus	3,813
7	Immune Cell	3,756
8	Poor Prognosis	3,614
9	Related Genes	3,606
10	Regression Analysis	3,582

出所：Web of Science®クラリベイトをもとに作成

論文で使用されている表現 (Bigram) を基に作成した。ストップワードが含まれる表現は集計対象外とし、語形変化を問わない形で集計を行った。

表4-7 TCGA に関する論文の著者の所在地

	著者の所在地	論文数
1	中国	14,877
2	米国	3,969
3	ドイツ	626
4	韓国	596
5	日本	561
6	英国	430
7	カナダ	401
8	台湾	381
9	イタリア	363
10	オーストラリア	263

出所：Web of Science®クラリベイトをもとに作成

た、タイトルまたは要旨に“The Cancer Genome Atlas”を含む原著論文を調査した。TCGAについて言及した20,804論文において、高頻度で利用されていた表現を表4-6に示す。TCGAでは腫瘍組織から取得されたトランスクリプトームに関連するデータが主に利用されており、エンリッチメント解析や発現変動に関連する研究が多数行われていることが確認された。腫瘍の分子サブタイプに関する解析が中心に行われており、バイオバンクで多く扱われている生殖細胞系列変異に関する研究とは目的やアプローチが異なる研究が多く実施されていることが分かる。さらに、TCGAに言及した論文の著者の所在地については、表4-7に示す通り、東アジアの国々が上位に多く見られており、悪性腫瘍領域に特化した研究に関しては、国内においても外部データの利用が進んでいることが確認された。日本が上位に位置する結果は、日本が悪性腫瘍領域の研究において国際的な競争力を有していることと一致している³²。

4.4 国際的な大規模ゲノムデータ関連論文の実態

分子プロファイル情報は、広義のリアルワールドデータに含まれる情報の中でも特に大学等の研究機関で先駆的に研究が進められている重要な分野であり、創薬への応用を目指す上で、これら基礎研究の発展は不可欠である。その中でも特に、大規模なゲノムデータは分子プロファイル情報の中核を成しており、製薬産業においても利用が開始されている段階にある。本項では、表4-4の集計結果を踏まえ、現時点で大規模なゲノムデータを利用した生殖細胞系列変異に関する研究において中心的な役割を果たしているGWASとメンデルのランダム化解析に関する研究に焦点を当て、製薬企業が関与していない研究も含めて、特に国際的な研究の実施に関して調査を行った。GWASは、特定の疾患や形質に関連する遺伝的多型を特定するために、ゲノム全域に渡って網羅的に検索する手法であり、メンデルのランダム化解析は、バリエントがランダムに分配されるという法則を利用し、バリエントを操作変数として利用して、観察研究において生じる交絡を軽減する手法である。これらの手法は、医薬品に関連する研究においても、疾患に関連する遺伝子多型の特定や、新たな医薬品ターゲットの発見、薬物の代謝や反応に影響を与える遺伝的要因の同定、医薬品の反応や副作用を予測するためのバイオマーカー開発等での利用が期待されている。これらを利用した創薬フレームワークは国内の研究チームからも報告されている³³。論文のキーワードを利用した今回の調査方法では、GWASを実施した研究のみを抽出することが困難であったため、GWASを実施した研究とそれらの結果を利用した研究の双方を含む調査を行った。今回の調査では、タイトル又は要旨に“Genome-wide association study”（“GWAS”を含む）又は“Mendelian Randomization”（“Mendelian Randomisation”を含む）の語句を含む2014～2023年の10年間に出版された原著論文を広義の大規模ゲノムデータ関連論文とした。これらのキーワードを利用する研究として農学や植物科学、動物学等も存在するため、それらの研究分野に投稿された論文は除いて集計を行った。

大規模ゲノムデータ関連論文の著者構成について、著者の所在地と著者が所属する組織を表4-8と表4-9に示す。複数の国や所属の著者で共著となっている論文は、それぞれの国、組織で集計に加えた。表4-5で示した各バイオバンクについて言及した論文と同様の傾向となっており、大規模ゲノムデータ関連の研究はこれらの国が主導していることが分かる。また、著者が所属する組織を確認すると、ハーバード大学（米国）、カリフォルニア大学システム（米国）、ロンドン大学（英国）等の大学は、これまでも複数の分野の集計で上位に位置しており、他領域に渡ってリアルワールドデータに関連する研究を主導して

32 厚生労働省、第9回 国立高度専門医療研究センターの今後の在り方検討会 資料1、https://www.mhlw.go.jp/stf/newpage_02860.html

33 Namba, S., Konuma, T., Wu, K. H., Zhou, W., Global Biobank Meta-analysis Initiative, & Okada, Y. 2022. A practical guideline of genomics-driven drug discovery in the era of global biobank meta-analysis. *Cell genomics*, 2(10), 100190.

表4-8 大規模ゲノムデータ関連論文の著者の所在地

	著者の所在地	論文数
1	米国	12,193 (45.7%)
2	中国	8,090 (30.3%)
3	英国	5,655 (21.2%)
4	ドイツ	3,068 (11.5%)
5	オランダ	2,428 (9.1%)
6	オーストラリア	2,231 (8.4%)
7	スウェーデン	2,094 (7.9%)
8	カナダ	2,040 (7.7%)
9	フランス	1,558 (5.8%)
10	デンマーク	1,389 (5.2%)
12	日本	1,329 (5.0%)

出所：Web of Science®クラリベイトをもとに作成

表4-9 大規模ゲノムデータ関連論文の著者所属組織

	著者が所属する組織	論文数
1	Harvard University	3,062 (11.5%)
2	University of California System	2,332 (8.7%)
3	University of London	2,247 (8.4%)
4	Harvard Medical School	1,853 (6.9%)
5	National Institutes of Health - USA	1,717 (6.4%)
6	Massachusetts Institute of Technology	1,307 (4.9%)
7	Broad Institute	1,256 (4.7%)
8	University of Cambridge	1,247 (4.7%)
9	Karolinska Institutet	1,245 (4.7%)
10	Massachusetts General Hospital	1,226 (4.6%)
89	RIKEN	436 (1.6%)

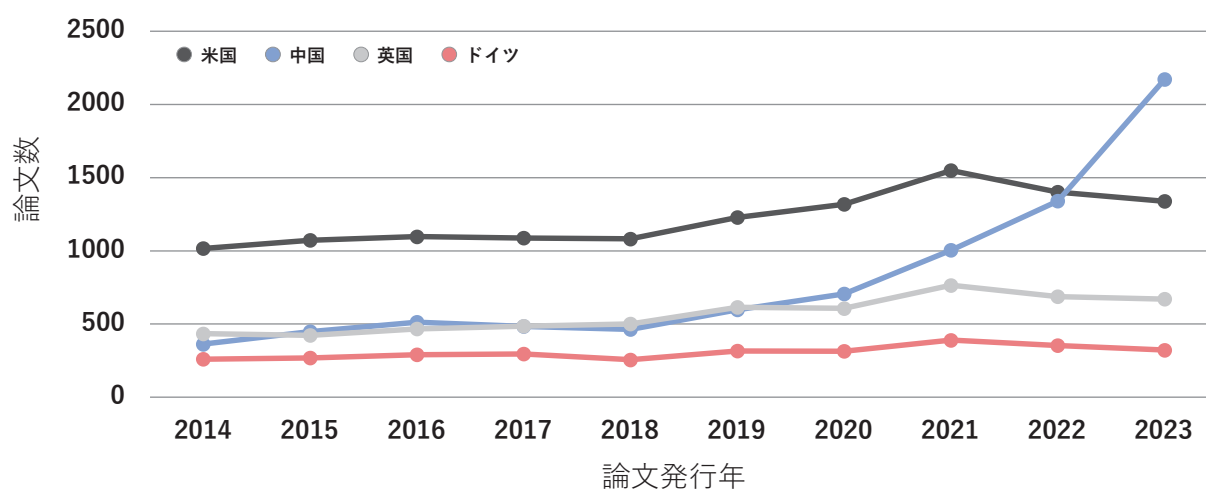
出所：Web of Science®クラリベイトをもとに作成

いることが分かる。日本は全体の12番目であり、最も多くの論文を発出した機関は理化学研究所であった。

論文数の上位4か国の出版数の年次推移を図4-2に示す。各バイオバンクに関する論文の増加傾向とは異なり、これらの基礎的な手法を利用した研究は、2014年以降、各国でわずかな増加を見せつつ、一定の論文数を維持していることが確認できる。上位の国の中では、中国の論文数が他国と比べて急増しており、2023年には米国を上回る論文数を公表していた。中国のこの顕著な論文数の増加傾向は、リアルワールドデータやウェアラブル技術に関する論文数の増加と比較しても顕著であり、特にゲノム関連の研究へ注力していることが分かる。科学技術・学術政策研究所による科学研究のベンチマーキングにおいても、基礎生命科学分野では中国が大きくシェアを伸ばしていることが報告されており、これらの動向とも一致する結果となっている⁸。この中国における直近の大規模ゲノムデータに関する論文の急増は、直近でデータが公開されたFinnGenについて言及した論文において、中国の著者が多く含まれていたことにも大きく影響していると推測される。

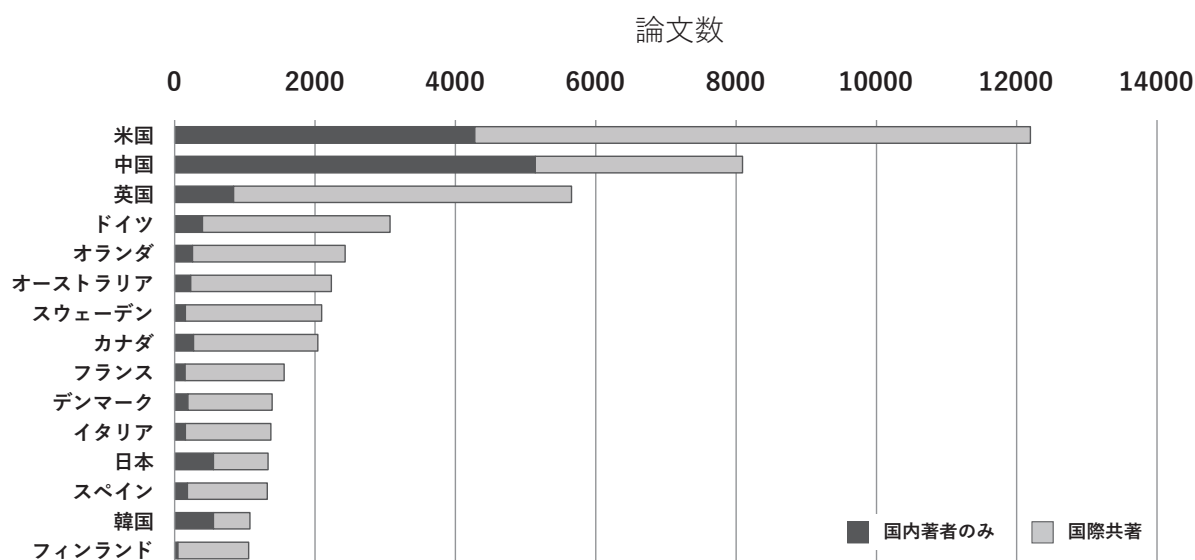
大規模ゲノムデータ関連論文における、国境を跨いだ共同研究について集計を行った結果を図4-3に示す。また、当該国の研究機関に所属する研究者が一人でも含まれている論文の数を分母として、各論文の

図4-2 大規模ゲノムデータ関連論文数の各国の推移



出所：Web of Science®クラリベイトをもとに作成

図4-3 大規模ゲノムデータ関連論文数と各国の国際共著

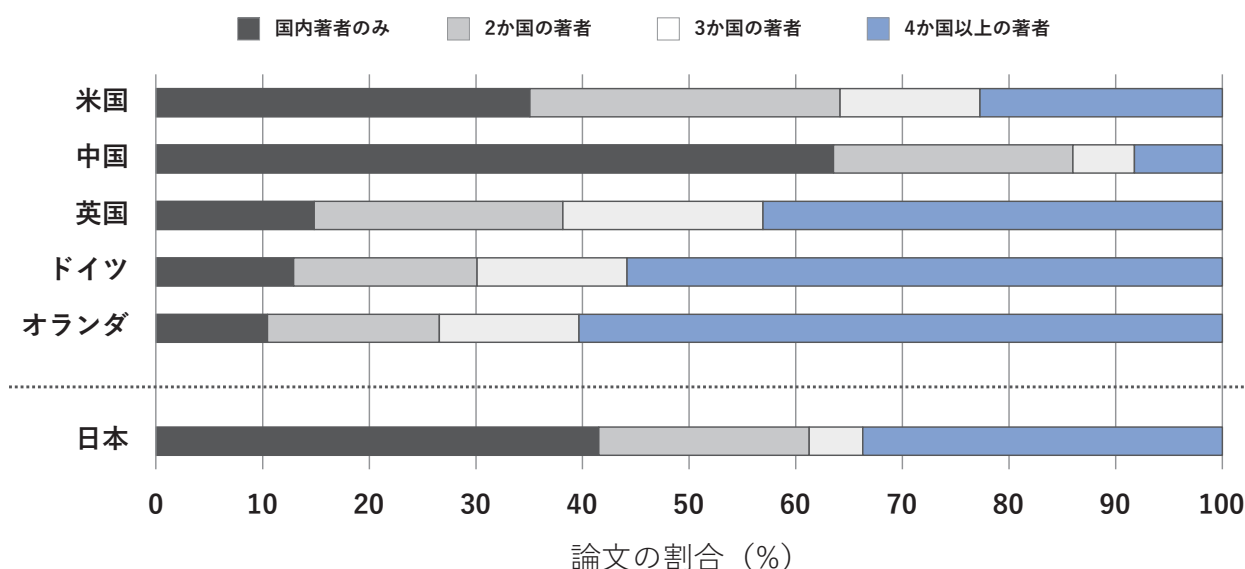


出所：Web of Science®クラリベイトをもとに作成

著者が所属する組織の国数の分布を図4-4に示す。論文数が上位に位置する米国では、国際共著論文（2か国以上の著者で構成される論文）の割合が6～7程度であったのに対し、中国では、国際共著論文の割合が3～4程度であり、論文数が上位の国でも研究体制に違いが確認された。他の分野と同様に欧州の国の国際共著論文の割合は比較的高くなっており、日本や韓国といった東アジアの国は米国と同程度となっていることが確認された。

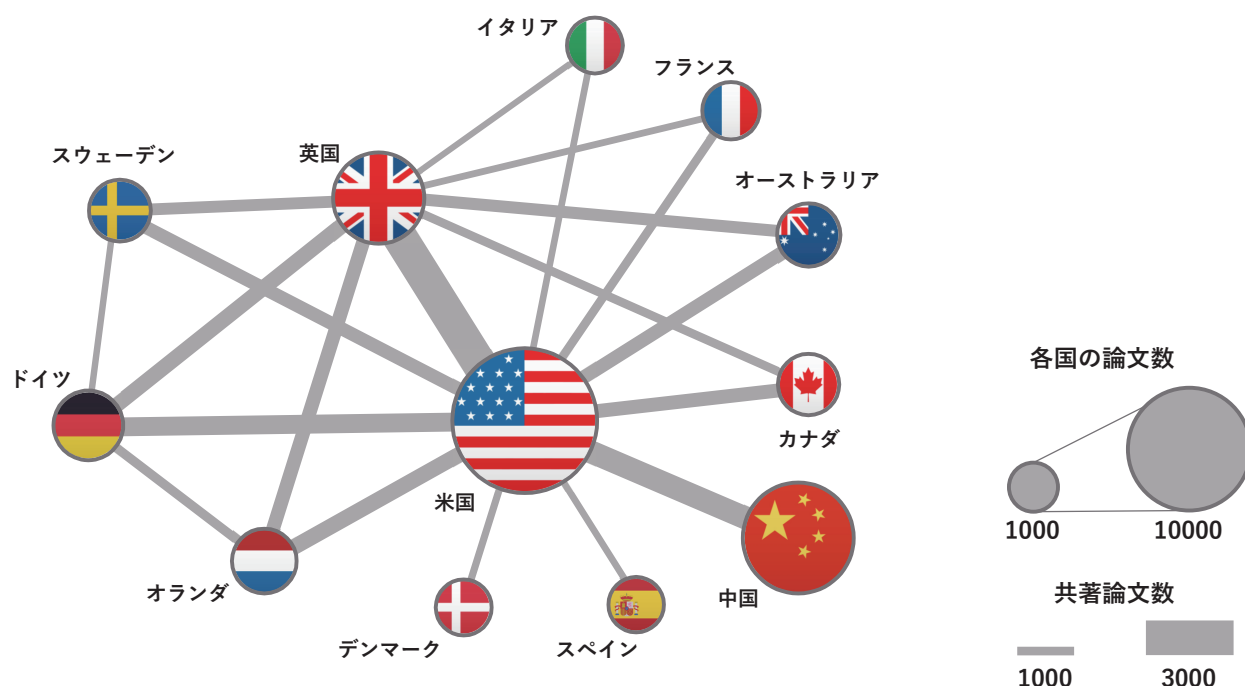
最後に、大規模ゲノムデータ関連論文における国際連携の現状を確認する。複数国の機関の著者が含まれる論文の中で、同一論文内の著者が所属する国の組み合わせを調査した。共著論文数の上位20番目までの組み合わせを用いて、連携の状態を図示した結果を図4-5に示す。大規模ゲノムデータ関連論文は、米国、中国、英国が主導しているものの、国際連携の中心となっているのは米国と英国であり、中国は他国との連携が比較的少なく、国内のみで研究を実施している傾向にあった。ドイツ、オランダ、スウェーデンは、欧州内での共同研究を多く行っていることが分かる。他の分野と異なり、スウェーデンやデンマー

図4-4 大規模ゲノムデータ関連論文の各国の共著関係の分布



出所：Web of Science®クラリベイトをもとに作成

図4-5 大規模ゲノムデータ関連論文の国際共著関係



出所：Web of Science®クラリベイトをもとに作成
 一対一の共著論文数の上位20位までを抽出。各国の円の大きさに用いた論文数には単一国のみの著者による論文も含む。

ク等の北欧諸国が多く論文に関与していることは、大規模ゲノムデータ関連論文の特徴である。日本は、国内著者のみで作成された論文が多く、上位20番目（共著807報）までの連携には含まれなかった。日本と共に論文を出版している国の集計を表4-10に示す。最も多い国際共著の相手国は、532報の米国であり、2国間の共著では45番目に共著論文数が多い組み合わせであった。中国も上位に含まれているものの、他の上位国としては、英国、ドイツ、カナダ等の欧米の国が中心となっていた。

表4-10 日本の著者が関与する大規模ゲノムデータに関する論文の共著者の所在地

	著者の所在地	論文数
1	米国	532
2	英国	316
3	ドイツ	246
4	カナダ	212
5	中国	208
6	オーストラリア	195
7	スウェーデン	188
8	オランダ	185
9	フランス	168
10	イタリア	163

出所：Web of Science[®]クラリベイトをもとに作成

4.5 小括

製薬産業における分子プロファイル情報は、特に研究開発の初期段階において極めて重要な役割を果たす。特に大規模なゲノムデータやトランスクリプトームデータは、バイオバンクや The Cancer Genome Atlas Program (TCGA) などの疾患コホートをはじめとする各種データベースが、製薬企業の研究活動に利用されるようになってきている。UK Biobank 等のバイオバンクでは、複数の製薬産業による資金提供により全ゲノム・全エクソームシーケンスが実施されており、これらのデータの活用が進む背景には、製薬産業の積極的な関与も存在する³⁴。現時点では、バイオバンクのデータを利用した、製薬企業が関与する論文数は多く無いものの、製薬企業による投資やバイオバンクの情報が利用可能となって間もないことを考えると、今後は製薬企業におけるデータ活用が広がっていくことが見込まれる。実際に、製薬企業以外の研究機関においては、バイオバンクについて言及した論文の数は近年増加しており、特に米国、英国、中国での研究活動が活発となっている。中でも、中国の大規模ゲノムデータに関する論文数の増加は顕著であり、現在では米国を上回る水準に達している。この傾向は、前述した他の情報源を利用した研究の傾向とは若干異なる状況となっている。中国の研究の中には、UK Biobank や FinnGen の GWAS の結果を利用したものも多く、他国のデータベースから得られる知見を積極的に活用していることが分かる。日本国内においても複数のバイオバンクの整備が進められている一方で、大規模ゲノムデータ関連論文の論文数を見ると、欧米の上位国や中国と比較して伸び悩んでいる現状がある。悪性腫瘍に関するデータベースに関連する論文は、他国と比較して遜色ない水準にあり、今後はこれらのデータベースを用いた研究の知見が他の疾患領域にも波及していくことが期待される。

日本国内のデータ基盤構築という側面では、遺伝子変異の発現率が人種によって異なる疾患も存在することを考慮すると、日本人に特化した医薬品の創出を行うためには、大規模な日本人のゲノムデータベースが不可欠となる³⁵。実際に、ゲノムデータに関する多くの研究はヨーロッパ系集団を対象に実施されており、その結果、集団の代表性の不均衡が生じていることが報告されている³⁶。国内の会議においても、これ

34 文部科学省、世代医療実現のための基盤形成に関する作業部会（第2回）資料1-2、https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu2/113/siryo/mext_00002.html

35 厚生労働省、第1回 がんに関する全ゲノム解析等の推進に関する部会 資料2-1、https://www.mhlw.go.jp/stf/newpage_07324.html

36 Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A. R., & Kuchenbaecker, K. 2022. A roadmap to increase diversity in genomic studies. *Nature medicine*, 28(2), 243-250.

らのゲノムデータベースの中に日本人やアジア人が少ないことが指摘されており、国際的に利用可能なデータベースを構築していくことも、日本人に対するエビデンス創出において極めて重要となる³⁷。本稿でも言及した TCGA に関しても、アジア人が占める割合は全体の 3 %にとどまっており、その結果として TCGA を用いた研究結果が少数派の人種に対して一般化できるかどうか不確実であることについても指摘されている³⁸。UK Biobank が国外の研究者にも広く情報を公開しているように、日本も海外に日本人の情報を公開し、海外で実施される研究におけるエビデンスに含めるような情報の利用方法も検討していく必要があると考える。

37 厚生労働省、第 9 回 ゲノム医療推進法に基づく基本計画の検討に係るワーキンググループ議事録、https://www.mhlw.go.jp/stf/newpage_49450.html

38 Spratt, D. E., Chan, T., Waldron, L., Speers, C., Feng, F. Y., Ogunwobi, O. O., & Osborne, J. R. 2016. Racial/ethnic disparities in genomic sequencing. *JAMA oncology*, 2(8), 1070-1074.

おわりに

リアルワールドデータの活用は、製薬産業において急速に進展しており、その影響力は今後さらに拡大していくことが予想される。従来の臨床試験では十分に捉えきれなかった、多様な患者集団における治療法の有効性や安全性を、実際の医療現場から得られるデータに基づいて評価できるようになることで、医薬品に関する情報収集の枠組みが変革され、従来のエビデンス創出のアプローチに新たな視点が加わるようになる。これらのデータの利用が与える影響は製薬産業に留まらず、医療政策や実際の診療現場においても大きな変革をもたらし、製薬産業は外部環境の変化に柔軟に対応する戦略的な意識改革を求められる可能性がある。例えば、より広範で精緻な患者情報を迅速に把握できるようになれば、医療機関や保険者はよりの確で個別化された治療戦略を策定できるようになる。このような利用は、治療の質の向上に寄与するだけでなく、医療費の効率的な削減にもつながると考えられる。患者個人に関連する多様な情報が意思決定に利用される未来を実現するためには、情報収集の基盤整備と並行して、人工知能や機械学習技術の進展が不可欠である。これらの技術が進化することにより、大規模なリアルワールドデータの解析手法はさらに高度化し、より患者ごとに最適化された治療法の提案や予後予測の精度向上に繋がることが期待される。

現在、リアルワールドデータに関する研究は、特に米国を中心に実施されており、日本は米国との共同研究が相対的に少ないことについて本稿で触れた。リアルワールドデータは、各地域の医療実態、遺伝的特性、行動特性を定量的に表現することができる情報源であることから、医薬品の売上のシェアが最も多く、製薬産業のビジネスの中心となっている米国や、シェアを拡大している中国のリアルワールドデータに関する研究の重要性は大きくなっている。一方で、日本人に対する適切な医療や医薬品の提供を行う観点では、海外のデータだけではなく、日本人の情報を含んだデータベースを用いたエビデンスの創出が行われることが望ましい。海外で得られた研究結果の日本人における再現性の確認や、日本人により適した医療提供に関する研究を進めるためには、海外で構築が進む医療情報基盤に準ずる基盤を遅滞なく整備することが必要である。また、グローバル全体で実施される研究において、日本人を含むデータを統合して利用していくためには、海外のデータと同水準の品質管理と一貫性を確保する必要があり、国際的なデータ標準化に積極的に参画することも重要であると考えられる。海外の動向に目を向けて医療情報基盤を構築することにより、最適な治療法を迅速かつ効率的に提供する社会の実現に繋がっていく。

また、3章と4章で触れたように、今後のリアルワールドデータの展望として、ますます多様化するデータソースの統合が進むことが予想される。医療機関で取得される情報に加え、ウェアラブルデバイスから得られる生体データや遺伝子情報、環境データなどを統合的に解析することが可能となれば、より精緻で包括的な患者像を構築することに繋がる。このような統合的な解析により、疾患の早期発見や予防、治療後の経過観察が精緻化され、個別化されたケアが提供されることに繋がる。製薬企業は、これらの新たなデータや技術を活用することによって、患者の個別ニーズに応じた医薬品や関連する情報の提供が可能となる。しかし、これらの技術発展や基盤構築には多大な投資とリソースが必要であり、大学や製薬企業、医療機関など、多様なステークホルダーの連携強化が不可欠である。これらの基盤は、医薬品開発のみならず、医療全体の質向上にも寄与し、最終的には患者の生活の質の向上に繋がることが期待される。このような基盤構築を実現するために、海外の最新技術の動向を継続的に追跡するとともに、国内におけるデータ基盤の構築を加速させることが必要である。

日本製薬工業協会
医薬産業政策研究所

〒103-0023 東京都中央区日本橋本町 2-3-11 日本橋ライフサイエンスビルディング 7F

Tel: 03-5200-2681 Fax: 03-5200-2684

<https://www.jpma.or.jp/opir/>