

希少疾患の診断や治療の発展に向けた ビッグデータ活用の可能性と課題

川本 祐子*

要 旨

ビッグデータは、大容量のデータから予期しないパターンや隠れた知識を明らかにすることに強みがあり、医療を含む多分野において活用されてきた。医療分野においてビッグデータの活用が注目される領域の一つが、希少疾患である。世界的な公衆衛生上の課題とされる希少疾患は、世界で4億人以上の患者がいると推定されている。希少疾患は病態の解明が十分ではなく、診断基準が定まっていない場合や特異的な治療法が確立されていない場合が多いため、希少疾患をもつ患者は効果的な治療法へのアクセスが制限されている。本稿では、希少疾患の診断や治療の発展に向けたビッグデータの活用可能性と課題について紹介する。

キーワード：ビッグデータ，リアルワールドデータ，データサイエンス，希少疾患，課題

社会保障研究 2024, vol.9, no.1, pp.90-99.

I はじめに

近年、多くの人が耳にしたであろう用語の一つが、「ビッグデータ」ではないか。Google Trendsでは、世界的に「big data (ビッグデータ)」という検索語が2011年を境に上昇傾向を示した。ビッグデータという言葉自体はそれ以前から存在しているが、ビッグデータを的確に表現するコンセンサスの得られた統一的な定義は今なお存在しない。その言葉が用いられる文脈に応じてさまざまな定義や特徴づけがなされてきた、実に多面的な用語である。ビッグデータの諸定義は総じてさまざまな源から得られた大量のデータで構成されている点が共通しており [Favaretto et al. (2020)],

「Volume (データの量)」、 「Variety (データ源やデータ種の多様性)」、 「Velocity (データの速度)」、 「Variability (データの変動性)」、 「Veracity (データの正確性)」、 「Value (データの価値)」などが特徴例として挙げられる [McCue and McCoy (2017)]。

ビッグデータという言葉に馴染みが無くても、我々はその提供者となり、利用者となっている。人々の生活に密着したビッグデータの用途をいくつか例示する。インターネット上のモール型販売サイトで買い物をする際、サイトの閲覧情報や購買情報等がデータとして活用され、利用者の行動傾向によって広告やおすすめ商品の表示内容が変わる。コンビニエンスストア等においては、実店舗内での動線データや発注・在庫・購買データ、

* 東京医科歯科大学大学院保健衛生学研究科 助教

ポイントカードや決済アプリ等から得られたデータを用いて来店者の行動分析や在庫管理を行い、取り扱う商品の内容や陳列方法などが調整される。また、我々はインターネットを用いて現在地や目的地の適時かつ局所的な交通情報や気象防災情報を瞬時に把握することができる。上記のビッグデータの用途例は、広く知られたものであるが、ビッグデータは全世界的に多業界で多目的に活用されている。

その一つが医療分野での活用である。生物医学や生命科学系の文献データベースであるPubMed[®]では、2003年以降に「big data」という言葉を用いた研究が進められてきた。医療分野でのビッグデータは、電子的に記録された診療・調剤報酬等の請求データベース（レセプト情報）や疾患特異的な患者登録型データベース、健康診断情報、診療記録情報（医療機関の電子カルテ情報）、ウェアラブルデバイスから得られた計測情報、ゲノム情報などが挙げられる。さまざまな疾患や医療サービスに関するビッグデータを用いた研究が進められているが、世界的な公衆衛生上の課題とされる希少疾患〔Forman et al. (2012), United Nations (2021)〕での活用も一考の余地がある。本稿では、希少疾患の診断や治療の発展に向けたビッグデータ活用の可能性と課題について検討する。

II 希少疾患の診断や治療の発展に向けたビッグデータの活用

1 希少疾患の定義と疫学的背景

希少疾患は、一般に人口に対して患者数が小さい疾患を示すが、疾患の希少性を計る世界共通の基準は存在しない。国や地域、組織により希少疾患の定義は異なり、それらの定義で用いられる有病率の閾値の世界平均は人口10万人当たり40例という報告がある〔Richter et al. (2015)〕。希少疾患

の総数については報告数にばらつきがあり、おおむね5000-8000と推定されている〔The Lancet Neurology (2011)〕。希少疾患の多くは疫学データが乏しく患者数の正確な推定は困難であるが、全世界的に4.7億人が希少疾患に罹患している可能性がある¹⁾と推定されている〔Ferreira (2019)〕。

2 希少疾患に関する社会的課題

希少疾患は、個々の疾患の患者数の少なさから、さまざまな社会的課題を抱えてきた〔Chan et al. (2020), EURORDIS (2009), Hurvitz et al. (2021), Institute of Medicine et al. (2010), Schieppati et al. (2008)〕。米国や欧州をはじめ各国で希少疾患に対する新しい治療法の研究・開発・販売を促進する法規制²⁾がなされるまで、希少疾患用の医薬品や医療機器の開発が進まない状況があった。希少疾患に対する研究開発資金の獲得が困難であることや臨床試験の対象者数の確保が難しいこと、効果的な治療法の利用可能性や医療費の補償が限定されることは課題として挙げられている。希少疾患は病態の解明や診断基準の確立がなされていないことが多く、有病率など疫学的特徴の推定が困難になり疾患の正確な評価を妨げること、診断の遅れや誤診、治療の遅れ、誤った治療の提供なども課題となっている。希少疾患に関する情報の少なさやその治療に関する知識を持つ臨床医の特定の難しさは、医療従事者と患者・家族を悩ませてきた。それらの課題は、希少疾患患者・家族のquality of lifeや社会的機会へ影響を及ぼしている。こうした社会課題を解決する糸口として、ビッグデータの活用が試みられてきた。

3 希少疾患の診断や治療の発展に向けてビッグデータを活用した国内外の事例

一般的にビッグデータの解析では、統計解析のほか、深層学習などの機械学習が用いられることが多い。希少疾患患者の画像検査データやゲノム

¹⁾ 米国のOrphan Drug Act (<https://www.fda.gov/media/99546/download>), 欧州委員会のOrphan Medicinal Products Regulations: OMP Regulations (<http://data.europa.eu/eli/reg/2000/141/oj>), 日本の希少疾病用医薬品・希少疾病用医療機器・希少疾病用再生医療等製品の指定による支援措置 (<https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000068484.html>) など。

情報などを深層学習で解析した研究に関するスコーピングレビューによると、希少な腫瘍性疾患や遺伝性疾患、神経疾患を主な対象としており、解析の焦点は診断に関する研究が最も多く(79%)、次いで予後(13%)や治療(8.7%)、特性評価(2.4%)に関する研究であった〔Lee et al. (2022)〕。希少疾患の疾患理解や診断、治療法の開発においてビッグデータを活用した事例をいくつか紹介する。

(1) 希少疾患の病態理解や分類、特性評価

ビッグデータは、現段階では不明瞭な希少疾患の疫学的・臨床的・形態学的・医療経済的特徴をより深く理解するために用いられることがある。Jang et al. (2021) は、2006年から2017年までの韓国の国民健康保険の給付記録から、国際疾病分類第10版を基に新たにアミロイドーシスと診断された2233件のデータを抽出し、アミロイドーシスの発生率や死因、生存率を評価した。Benovoy et al. (2022) は、川崎病における冠動脈瘤の形成予防に向けた特性評価のため、川崎病患者27名から光干渉断層撮影法による冠動脈の画像データを収集し、川崎病における冠動脈の伸展性の低下(硬化)と冠動脈壁の損傷の程度との相関、および、冠動脈の伸展性の経時的変化について、深層学習を用いて評価した。Lazic et al. (2021) は、小児腫瘍神経芽腫患者から神経芽腫細胞株や骨髄吸引液、末梢血液を採取し、データマイニングや画像処理、深層学習などを経て、骨髄転移における腫瘍とその周辺細胞の特徴の多様性を調べた。Walker et al. (2016) は、Orphanetが開発したORPHAコード(希少疾患を包括的に分類し符号化したもの)と国際疾病分類第10版オーストラリア版(ICD-10-AM)の照合作業を経て、西オーストラリア州の1999年から2010年までの各希少疾患患者に関するデータをHospital Morbidity Data Collectionから収集し、報告された希少疾患の種類、患者数、利用された医療サービス、西オーストラリア州の医療システムに対する希少疾患関連の医療費を調査した。Chiu et al. (2018) は、香港の公立病院で用いられるClinical Data Analysis and Reporting System

の2005年から2016年までデータのうち、前掲のICD-10-AMの希少疾患に関するコードに基づいて、希少疾患により入院した546,673名の患者データを抽出し、香港における希少疾患の有病率と分布の推定、希少疾患患者の入院数とその割合、およびその入院関連医療費を推定した。

(2) 希少疾患の診断や予後予測の向上

希少疾患への罹患が疑われる患者が、個々の希少疾患に関する情報やその診断・治療に関する知識・技術を有する専門家・専門施設を特定することは容易ではない。また、個々の希少疾患の専門家ではない医師が、そうした患者を診ることによる診断の遅れや誤診は、治療機会の損失や誤った治療の提供を招き、予後を左右しうる。Schaaf et al. (2020) のスコーピングレビューによると、大量の疾患情報(疾患・治療法・研究に関する文献)や遺伝情報などのビッグデータに基づく、希少疾患に特異的な診断支援ソフトウェアが開発されている。プロトタイプではなく開発が完了しているシステムは14あり、症状や表現型データなどを入力して疾患関連情報を検索したり、疾患の原因である遺伝的変異を特定したり、類似患者を同定して症例を共有したりできる。Dille et al. (2020) は、ドイツの14施設からファロー四徴症患者372名の心臓磁気共鳴画像法(心臓MRI)検査データを取得した。深層学習を用いた画像データ解析によって心臓の寸法と機能の測定値を算出し、ほかの臨床指標と組み合わせることでファロー四徴症患者の予後を推定し、自動予後予測アルゴリズムの有用性を評価した。

(3) 希少疾患の治療法の開発・評価

ビッグデータは、希少疾患の診断や予後予測の向上のために用いられる場合と比較して多くはないが、治療法の開発や治療後の評価にも活用されることがある。上咽頭癌に対する治療計画の最適化を目指し、深層学習を用いたいくつかの取り組みがなされている。Ma et al. (2021) は、上咽頭癌に対する放射線治療計画においてMRIのみを使用した場合に必要となる電子密度情報を提供するた

め、早期上咽頭癌患者20名から得た1433組のT1強調磁気共鳴シミュレーション画像とコンピュータ断層撮影法（CT）シミュレーション画像を用いて、疑似CTの生成法を開発し、評価した。Xue et al. (2021) は、限られた医療資源のなかで正確な放射線量計算を行う場合に必要となる合成CTを提供するため、上咽頭癌169名から得た20926のCT画像データ（コンビームCTや線量計算のための計画CT）から高画質な合成CT画像を生成する方法を開発し、評価した。Yang et al. (2021) は、上咽頭癌における導入化学療法治療への反応予測（レスポナーとノンレスポナーの特定）によりノンレスポナーへの不必要な毒性とコストを回避するため、導入化学療法を受けた上咽頭癌患者208名のCT画像データを基に深層学習等により反応予測モデルを作成し、その性能を評価した。

希少疾患用医薬品の開発におけるビッグデータの活用例としては、ドラッグリポジショニング（承認済みの治療薬または治験中の医薬品から別の疾患に有効な薬効を見つけるための、既存薬に対する再開発の取り組み）への貢献も挙げられる。Lee et al. (2019) は、機械学習を用いて数千の臨床遺伝子発現プロファイルを統合し、各疾患に特異的な分子的特徴を同定する統一的な計算フレームワークである Unveiling RNA Sample Annotation for Human Diseases (URSAHD) を開発し、その有用性を評価した。URSAHDは、一般的で治療可能な疾患とまれで治療不可能な疾患間の特徴の類似性に基づいて、希少疾患の治療薬の潜在的な候補を予測することができる。個々の疾患遺伝子または従来の正常/疾患発現差解析を使用する計算アプローチよりも正確に近縁の疾患を区別でき、希少疾患を含むあらゆる疾患への適用が期待されている。

(4) 希少疾患情報の整備

希少疾患分野の研究や治療、知識の発展には希少疾患登録をはじめとする疾患情報の収集が重要であり、疾患情報の共有や比較のためには疾患名や疾患分類、登録項目の標準化が必須となる。希

少疾患情報の整備に向けて、ビッグデータを活用した取り組みを紹介する。Ninomiya and Okura (2022) は、日本における希少疾患用医薬品の開発に向けた科学的情動的基盤強化のため、レセプト情報・特定健診等情報データベース（NDB）を用いた全国希少疾病疫学調査を実施した。約4500の希少疾患の日本における患者数について性別・年齢別に10年分のデータを取得した後、病名を翻訳してNDBの希少疾患とOrphanetに記載のある希少疾患の対応関係を確立し、それらの有病率と自然史を比較した。自然史は、個々の疾患経過における患者の状態変化を特徴づけるものであり、疾患集団の定義や臨床的エンドポイントの選択など、新しい治療法を確立するために不可欠とされている〔US Food and Drug Administration (2019)〕。Santoro et al. (2015) は、希少疾患登録データベース間の将来的な相互接続等に貢献するため、EPIRARE プロジェクト (Building Consensus and Synergies for the EU Rare Disease Patient Registration) による希少疾患登録の活動とニーズに関する調査から得られたデータを機械学習によりデータマイニングし、希少疾患登録データベースの分類と特性評価を実施した。中国では、2016年に希少疾患専門家委員会が設置されて以降、政府主導で希少疾患への取り組みが開始された。2016～2020年の希少疾患コホート研究や2018年の希少疾患カタログの発行のほか、中国全土の標準化された希少疾患データを収集するため、National Rare Diseases Registry Systemが開始され、2020年までの最初の5年間に登録された症例の分析も実施された〔Guo et al. (2021)〕。

III 希少疾患の診断や治療の発展に向けたビッグデータ活用の可能性と課題

1 希少疾患の診断や治療の開発の現状

希少疾患の多くは、その病態が不明瞭であり、疾患を特徴づける情報は乏しく、診断基準が定められておらず、特異的な治療法が確立されていない。また、継続的なモニタリング指標の欠如も疾病管理の障壁となっている〔Hurvitz et al.

(2021)〕。そのため、希少疾患患者およびその家族は、有用な情報や適切な診断、効果的な治療法へのアクセスが制限され、Quality of lifeの著しい低下を招いてきた。画像データやゲノム情報などのビッグデータを解析する技術の向上は、病態解明や診断精度の向上、治療シーズの探索に資する研究の促進をもたらした。前述の希少疾患に対する治療法の研究・開発・販売を促進する法規制がなされたことにより、製薬会社にとって希少疾患分野での新薬開発は魅力あるものとなり、医薬品の開発も伸展している。希少疾患治療薬の情報流通や研究開発を進めるための組織²⁾も設立され、希少疾患に対する治療法の開発と市場投入を強化している〔Kaspar (2005)〕。しかしながら、希少疾患に対する法規制は比較的裕福な国や地域(一人当たりの国民総所得が高い地域)において設立されている傾向が依然あり、規制的・経済的インセンティブが働かず医薬品の開発や承認が進まない国や地域とのギャップも存在している〔Chan et al. (2020)〕。希少疾患に関する研究が飛躍的に進むなか、それらの知識が治療へ転換される率や速度は向上しておらず、医薬品の承認数の飛躍的な伸びは得られていないという指摘もある〔Tambuyzer et al. (2020)〕。日本においてはドラッグロス(日本以外の国で承認されている薬が日本では開発が予定されず、日本では有効な治療薬にアクセスできないこと)の課題もあり、2019年5月までに米国で承認された希少疾患用医薬品の約43%が日本で承認されておらず、そのうち75%は、日本での臨床試験が実施されていないことも報告されている〔武田薬品工業株式会社(2020)〕。希少疾患を対象とした臨床試験や観察研究においては、さまざまなデザイン³⁾が提案されているが、万能なデザインはない。研究者は、各デザインの利点や欠点を理解し、個々の研究か

ら生じる分析上の課題に注意を払いながら、幅広い可能性の中でデザインを設計することが求められる〔Bogaerts et al. (2015), Mitani and Haneuse (2020)〕。希少疾患での臨床試験においてデータ不足が課題として挙げられることや希少疾患用医薬品が高額であることなどから、希少疾患用医薬品のファーマコビジランス(医薬品の開発段階から市販後も一貫して医薬品の安全性を管理すること)や承認後の費用対効果評価は重要な課題である。また、病態解明や診断・治療法の開発と並行して、希少疾患情報の整備に向けた希少疾患の分類や疾患登録も徐々に進められているが、情報の有効活用(国際間あるいは疾患間の共有など)において必須となる国際標準化には更なる努力を要する。

2 希少疾患におけるビッグデータ活用への期待

ビッグデータは、大容量のデータから予期しないパターンや隠れた知識を明らかにすることに強みがある。その特徴を生かして、社会課題(例えば、保健・医療・福祉の需要増大に対する効率化やリソースの適正配分、医療者や介護者の負担軽減、疾病予防・管理技術の向上、医療格差の是正、患者・家族サービスや医療安全の向上、ほか)の解決のために活用されることが期待されている〔中島(2015)〕。希少疾患においては、非希少疾患と比較して必ずしもデータが大容量であるとは限らず、データサイズが小さいことにより臨床的に意義のある評価へと繋がらない場合があることも指摘されている〔Lee et al. (2022)〕。しかしながら、希少疾患の発生パターンや分子的特徴の特定など疾患特性の評価、診断技術の最適化、進行・予後の予測、治療に対する患者の反応予測、希少疾患用医薬品の潜在的な候補の予測(ドラッグリ

²⁾ Rare Diseases International (RDI) は世界150カ国以上の希少疾患患者団体を代表する世界的な連合体、OrphaNet は希少疾患に関する情報を提供、National Organization of Rare Diseases (NORD) は希少疾患患者とその家族を代表する非営利団体、International Rare Diseases Research Consortium (IRDiRC) は国際連携により希少疾患の研究開発の促進を目指す幅広い代表(資金提供者、企業、患者団体、専門家他)で構成された組織、など。

³⁾ 非盲検試験、歴史的対照試験、クロスオーバー試験、N-of-1試験、アンブレラ試験、バスケット試験、適応的デザイン、ベイズ流デザイン、自己対照研究デザイン、症例対照デザインなど。

ポジショニング) などにおいては、希少疾患においてもビッグデータの強みが活かされている。

3 希少疾患におけるビッグデータ活用の課題

希少疾患におけるビッグデータ活用の最大の課題は、「データの不足」である。希少疾患患者から得たデータの解析に深層学習を用いた論文のスコopingレビュー [Lee et al. (2022)] においても、レビュー対象となった論文のうち43%がデータの不十分さを課題として挙げている。データ不足へ対処する主な取り組みとしては、「データの追加」、「データの質の向上」、「データ解析方法の改善」などが挙げられる。

異なる情報源との連結により追加データを取得することでエビデンスを補完または補強することに関心が寄せられている [Alemayehu (2016)]。追加する情報は、実験研究、観察研究（患者や医療者が報告する定性的データやウェアラブル/スマート/ヘルスケアデバイスによる測定データを含む）、レセプト情報、健診データ、患者登録情報、電子カルテ、文献、自然史研究、ソーシャルメディアで公開されている情報、その他の非標準的および非従来型の情報源など、多様である。希少疾患の研究に限定されることではないが、それらの情報源を連結させる上での構造的、技術的、制度的、倫理的課題もある。例えば、電子カルテを例に挙げると、電子カルテは医療提供を円滑に行うために開発されており、研究を想定した項目が網羅的かつ構造的に欠損なく含まれるわけではない [Wu et al. (2020)]。臨床検査データなどの定量的データとともに、今後の活用が注目されるフリーテキストなどの定性的な非構造化データは、自然言語処理技術の向上が不可欠となる。フリーテキストは標準化された用語が用いられない場合が多く、解析時には同義の用語を丁寧に定義づけ・分類する作業が必要となる。多様なデータの連結・統合・解析においては、それらを行うに足るデータ管理システムも必要となる。ほかのデータベースとの連結が、制度として禁止されている場合もある。異なる情報源から得たデータを連結する際には、個人のプライバシー保護の観点

から、どのような特性の情報を誰がどの範囲で共有するのか、機密性の検討が重要である。希少疾患の場合、患者数が少ないが故に、匿名化していても個人が特定される可能性があることも留意する必要がある。

データを追加する場合も、個々のデータが低品質ではビッグデータを用いる価値は制限される。データの質を向上させるために、元データを修正したり、合成データを作成してデータを拡張させたりすることがあるが、データ収集の質を改善することは重要である。高品質データを得るためには、データ管理者がデータの品質管理体制を整え、データ利用者がデータの品質確認を徹底することが大前提となる。研究で主要あるいは副次評価項目として用いるデータや関連因子と想定されるデータがどのように産生され管理されるのかを研究計画段階から正しく理解し、研究開始後もデータ品質管理監視などを通じて、完全なデータが得られるよう一貫して努めることは意義がある。データ収集を改善するための措置が導入された結果、データの完全性と品質が向上したという報告がある [Clarke et al. (2011)]。データの品質確認においては、専門家による手作業によるデータ検証やアノテーション（さまざまな形態の個々のデータがもつ意味や範囲を特定するためのタグ付けを各データに行う工程）が必要となるが、膨大な作業となることを予め想定する必要がある。

研究対象やデータの特性を考慮しながらより適した解析方法を検討することは、データ不足が課題となる状況下でエビデンスの質を高められる可能性がある。新たなアルゴリズムの採用においては十分な批判的吟味を重ねることが重要であり、アルゴリズムの変更によって許容しがたい結果に繋がる可能性がある [Lazer et al. (2014)] ことも認識する必要がある。

IV おわりに

希少疾患の診断や治療の発展に、ビッグデータの活用が有用であることは事例で示されており、更なる活用が期待されている。しかしながら、

データ不足という課題は、今後も希少疾患でのビッグデータ活用において不可避である。異なる情報源との連結によるデータ補強や元データの合理的な修正、合成データによる拡張、データ収集・利用におけるデータ品質の管理と確認、解析技術やデータ管理環境の向上などを図ることにより、課題を克服することが求められる。その実現には、個々の研究者の努力のみならず、国際的かつ学際的な取り組みも重要である。異なる情報源の連結においては、個人のプライバシー保護とデータ利用による公益とのバランスについて、学際的な専門家を含めた十分な検討が必要であり、ガイドラインや法の整備が必要となる。希少疾患は個々の疾患の患者数が少ないからこそ、国や地域、疾患の垣根を越えた協働により、希少疾患の診断や治療に関する研究が相乗効果で発展していくことも期待したい。その一助ともなる疾患（患者）情報登録データベースの整備においては、国際的な情報共有・比較・研究への活用が進むよう、データベースの構成や用いられる疾患名、疾患分類の統一など、国際標準化が進むことも望む。最後に、プレジジョン・メディシン（日本においては精密医療とも称されるが、各個人の状態に応じた最適な治療法を最先端技術に基づいて提供すること）の推進においても、ビッグデータの活用が期待されている。希少疾患患者の臨床情報（症状や治療、医療費に関するあらゆるデータ）に加え、疾患情報やゲノム情報、生活習慣、体内外の環境情報など、多様な情報の統合と予測技術により、個々の希少疾患患者に最適な医療が特定されることは、患者の予後やQuality of lifeの向上が期待できる。一方で、国や地域における人や物、原資、時間、情報など重要資源の有限性を踏まえると、全体最適も考慮しながらプレジジョン・メディシンが進められることが重要である。

謝辞

本稿は科研費（22H03307）による研究成果の一部である。

参考文献

- Alemayehu, D., & Berger, M. L. (2016). Big Data: transforming drug development and health policy decision making. *Health services & outcomes research methodology*, 16, pp.92-102. <https://doi.org/10.1007/s10742-016-0144-x>
- Benovoy, M., Dionne, A., McCrindle, B. W., Manlhiot, C., Ibrahim, R., & Dahdah, N. (2022). Deep Learning-Based Approach to Automatically Assess Coronary Distensibility Following Kawasaki Disease. *Pediatric cardiology*, 43 (4), pp.807-815. <https://doi.org/10.1007/s00246-021-02790-z>
- Bogaerts, J., Sydes, M. R., Keat, N., McConnell, A., Benson, A., Ho, A., Roth, A., Fortpied, C., Eng, C., Peckitt, C., Coens, C., Pettaway, C., Arnold, D., Hall, E., Marshall, E., Sciafani, F., Hatcher, H., Earl, H., Ray-Coquard, I., Paul, J., ... Seymour, M. (2015). Clinical trial designs for rare diseases: studies developed and discussed by the International Rare Cancers Initiative. *European journal of cancer (Oxford, England: 1990)*, 51 (3), pp.271-281. <https://doi.org/10.1016/j.ejca.2014.10.027>
- Chan, A. Y. L., Chan, V. K. Y., Olsson, S., Fan, M., Jit, M., Gong, M., Zhang, S., Ge, M., Pathadka, S., Chung, C. C. Y., Chung, B. H. Y., Chui, C. S. L., Chan, E. W., Wong, G. H. Y., Lum, T. Y., Wong, I. C. K., Ip, P., & Li, X. (2020). Access and Unmet Needs of Orphan Drugs in 194 Countries and 6 Areas: A Global Policy Review With Content Analysis. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 23 (12), pp.1580-1591. <https://doi.org/10.1016/j.jval.2020.06.020>
- Chiu, A. T. G., Chung, C. C. Y., Wong, W. H. S., Lee, S. L., & Chung, B. H. Y. (2018). Healthcare burden of rare diseases in Hong Kong - adopting ORPHAcodes in ICD-10 based healthcare administrative datasets. *Orphanet journal of rare diseases*, 13 (1), p.147. <https://doi.org/10.1186/s13023-018-0892-5>
- Clarke, J. T., Giugliani, R., Sunder-Plassmann, G., Elliott, P. M., Pintos-Morell, G., Hernberg-Ståhl, E., Malmenäs, M., Beck, M., & FOS Investigators (2011). Impact of measures to enhance the value of observational surveys in rare diseases: the Fabry Outcome Survey (FOS). *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 14 (6), pp.862-866. <https://doi.org/10.1016/j.jval.2011.05.002>
- Diller, G. P., Orwat, S., Vahle, J., Bauer, U. M. M., Urban, A., Sarikouch, S., Berger, F., Beerbaum, P., Baumgartner, H., & German Competence Network for Congenital Heart Defects Investigators (2020). Prediction of prognosis in patients with tetralogy of

- Fallot based on deep learning imaging analysis. *Heart* (British Cardiac Society), 106 (13), pp.1007-1014. <https://doi.org/10.1136/heartjnl-2019-315962>
- EURORDIS (2009) The Voice of 12,000 Patients, <https://www.eurordis.org/publications/the-voice-of-12000-patients/> (2024年1月29日最終確認) .
- Favaretto, M., Clercq, E. D., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLOS ONE*, 15 (2), e0228987. <https://doi.org/10.1371/journal.pone.0228987>
- Ferreira C. R. (2019). The burden of rare diseases. *American journal of medical genetics. Part A*, 179 (6), pp.885-892. <https://doi.org/10.1002/ajmg.a.61124>
- Forman J, Taruscio D, Llera VA, Barrera LA, Coté TR, Edfjäll C, Gavhed D, Haffner ME, Nishimura Y, Posada M, Tambuyzer E, Groft SC, & Henter JL. (2012) International Conference for Rare Diseases and Orphan Drugs (ICORD). The need for worldwide policy and action plans for rare diseases. *Acta Paediatr*, 101 (8), pp.805-7. <https://doi.org/10.1111/j.1651-2227.2012.02705.x>
- Guo, J., Liu, P., Chen, L., Lv, H., Li, J., Yu, W., Xu, K., Zhu, Y., Wu, Z., Tian, Z., Jin, Y., Yang, R., Gu, W., Zhang, S., & Administrative Group of National Rare Diseases Registry System of China (2021). National Rare Diseases Registry System (NRDRS): China's first nation-wide rare diseases demographic analyses. *Orphanet journal of rare diseases*, 16 (1), p.515. <https://doi.org/10.1186/s13023-021-02130-7>
- Hurvitz, N., Azmanov, H., Kesler, A., & Ilan, Y. (2021). Establishing a second-generation artificial intelligence-based system for improving diagnosis, treatment, and monitoring of patients with rare diseases. *European journal of human genetics: EJHG*, 29 (10), pp.1485-1490. <https://doi.org/10.1038/s41431-021-00928-4>
- Institute of Medicine (US) Committee on Accelerating Rare Diseases Research and Orphan Product Development, Field, M. J., & Boat, T. F. (Eds.). (2010). *Rare Diseases and Orphan Products: Accelerating Research and Development*. National Academies Press (US).
- Jang, S. Y., Kim, D., Choi, J. O., & Jeon, E. S. (2021). Incidence, Cause of Death, and Survival of Amyloidosis in Korea: A Retrospective Population-Based Study. *International journal of heart failure*, 3 (3), pp.172-178. <https://doi.org/10.36628/ijhf.2021.0006>
- Kaspar R. L. (2005). Challenges in developing therapies for rare diseases including pachyonychia congenita. *The journal of investigative dermatology. Symposium proceedings*, 10 (1), pp.62-66. <https://doi.org/10.1111/j.1087-0024.2005.10208.x>
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science* (New York, N.Y.), 343 (6176), pp.1203-1205. <https://doi.org/10.1126/science.1248506>
- Lazic, D., Kromp, F., Rifatbegovic, F., Repiscak, P., Kirr, M., Mivalt, F., Halbritter, F., Bernkopf, M., Bileck, A., Ussowicz, M., Ambros, I. M., Ambros, P. F., Gerner, C., Ladenstein, R., Ostalecki, C., & Taschner-Mandl, S. (2021). Landscape of Bone Marrow Metastasis in Human Neuroblastoma Unraveled by Transcriptomics and Deep Multiplex Imaging. *Cancers*, 13 (17), p.4311. <https://doi.org/10.3390/cancers13174311>
- Lee, J., Liu, C., Kim, J., Chen, Z., Sun, Y., Rogers, J. R., Chung, W. K., & Weng, C. (2022). Deep learning for rare disease: A scoping review. *Journal of biomedical informatics*, 135, 104227. <https://doi.org/10.1016/j.jbi.2022.104227>
- Lee, Y. S., Krishnan, A., Oughtred, R., Rust, J., Chang, C. S., Ryu, J., Kristensen, V. N., Dolinski, K., Theesfeld, C. L., & Troyanskaya, O. G. (2019). A Computational Framework for Genome-wide Characterization of the Human Disease Landscape. *Cell systems*, 8 (2), pp.152-162.e6. <https://doi.org/10.1016/j.cels.2018.12.010>
- Ma, X., Chen, X., Li, J., Wang, Y., Men, K., & Dai, J. (2021). MRI-Only Radiotherapy Planning for Nasopharyngeal Carcinoma Using Deep Learning. *Frontiers in oncology*, 11, 713617. <https://doi.org/10.3389/fonc.2021.713617>
- McCue, M. E., & McCoy, A. M. (2017). The Scope of Big Data in One Medicine: Unprecedented Opportunities and Challenges. *Frontiers in Veterinary Science*, 4, 278019. <https://doi.org/10.3389/fvets.2017.00194>
- Mitani, A. A., & Haneuse, S. (2020). Small Data Challenges of Studying Rare Diseases. *JAMA network open*, 3 (3), e201965. <https://doi.org/10.1001/jamanetworkopen.2020.1965>
- Ninomiya, K., & Okura, M. (2022). Nationwide comprehensive epidemiological study of rare diseases in Japan using a health insurance claims database. *Orphanet journal of rare diseases*, 17 (1), p.140. <https://doi.org/10.1186/s13023-022-02290-0>
- Richter, T., Nestler-Parr, S., Babela, R., Khan, Z. M., Tesoro, T., Molsen, E., Hughes, D. A., & International Society for Pharmacoeconomics and Outcomes Research Rare Disease Special Interest Group (2015). *Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group*. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 18 (6),

- pp.906-914. <https://doi.org/10.1016/j.jval.2015.05.008>
- Santoro, M., Coi, A., Lipucci Di Paola, M., Bianucci, A. M., Gainotti, S., Mollo, E., Taruscio, D., Vittozzi, L., & Bianchi, F. (2015). Rare disease registries classification and characterization: a data mining approach. *Public health genomics*, 18 (2), pp.113-122. <https://doi.org/10.1159/000369993>
- Schaaf, J., Sedlmayr, M., Schaefer, J., & Storf, H. (2020). Diagnosis of Rare Diseases: a scoping review of clinical decision support systems. *Orphanet journal of rare diseases*, 15 (1), p.263. <https://doi.org/10.1186/s13023-020-01536-z>
- Schieppati, A., Henter, J. I., Daina, E., & Aperia, A. (2008). Why rare diseases are an important medical and social issue. *Lancet* (London, England), 371 (9629), pp.2039-2041. [https://doi.org/10.1016/S0140-6736\(08\)60872-7](https://doi.org/10.1016/S0140-6736(08)60872-7)
- Tambuyzer, E., Vandendriessche, B., Austin, C. P., Brooks, P. J., Larsson, K., Miller Needleman, K. I., Valentine, J., Davies, K., Groft, S. C., Preti, R., Oprea, T. I., & Prunotto, M. (2020). Therapies for rare diseases: therapeutic modalities, progress and challenges ahead. *Nature reviews. Drug discovery*, 19 (2), pp.93-111. <https://doi.org/10.1038/s41573-019-0049-9>
- The Lancet Neurology (2011). Rare neurological diseases: a united approach is needed. *The Lancet. Neurology*, 10 (2), p.109. [https://doi.org/10.1016/S1474-4422\(11\)70001-1](https://doi.org/10.1016/S1474-4422(11)70001-1)
- United Nations (2021). Resolution adopted by the General Assembly on 16 December 2021: 76/132. Addressing the challenges of persons living with a rare disease and their families. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N21/400/70/PDF/N2140070.pdf?OpenElement> (2024年1月29日最終確認)
- U.S. Food and Drug Administration (2019). Rare diseases: natural history studies for drug development. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-natural-history-studies-drug-development> (2024年1月29日最終確認)
- Walker, C. E., Mahede, T., Davis, G., Miller, L. J., Girschik, J., Brameld, K., Sun, W., Rath, A., Aymé, S., Zubrick, S. R., Baynam, G. S., Molster, C., Dawkins, H. J. S., & Weeramanthri, T. S. (2017). The collective impact of rare diseases in Western Australia: an estimate using a population-based cohort. *Genetics in medicine: official journal of the American College of Medical Genetics*, 19 (5), pp.546-552. <https://doi.org/10.1038/gim.2016.143>
- Wu, J., Wang, C., Toh, S., Pisa, F. E., & Bauer, L. (2020). Use of real-world evidence in regulatory decisions for rare diseases in the United States-Current status and future directions. *Pharmacoepidemiology and drug safety*, 29 (10), pp.1213-1218. <https://doi.org/10.1002/pds.4962>
- Xue, X., Ding, Y., Shi, J., Hao, X., Li, X., Li, D., Wu, Y., An, H., Jiang, M., Wei, W., & Wang, X. (2021). Cone Beam CT (CBCT) Based Synthetic CT Generation Using Deep Learning Methods for Dose Calculation of Nasopharyngeal Carcinoma Radiotherapy. *Technology in cancer research & treatment*, 20, 15330338211062415. <https://doi.org/10.1177/15330338211062415>
- Yang, Y., Wang, M., Qiu, K., Wang, Y., & Ma, X. (2022). Computed tomography-based deep-learning prediction of induction chemotherapy treatment response in locally advanced nasopharyngeal carcinoma. *Strahlentherapie und Onkologie: Organ der Deutschen Röntgengesellschaft ... [et al]*, 198 (2), pp.183-193. <https://doi.org/10.1007/s00066-021-01874-2>
- 武田薬品工業株式会社 (2020)「日本における希少疾患の課題」。 https://genetics.qlife.jp/documents/RD_WhitePaper.pdf (2024年1月29日最終確認)
- 中島直樹 (2015)「ビッグデータとは何か？」 *脳神経外科ジャーナル*, 24 (10), pp.664-671. <https://doi.org/10.7887/jcns.24.664>

(かわもと・ゆうこ)

Exploring the possibilities and challenges of utilizing big data in the development of diagnostic and therapeutic methods for rare diseases

KAWAMOTO Yuko*

Abstract

Big data possess a strong capability to reveal unexpected patterns and hidden insights from large volumes of data and have found applications across various fields, including healthcare. Rare diseases stand out as a domain where big data hold promise for transformative impact in healthcare. Globally, rare diseases pose a significant public health challenge, impacting more than 400 million people. Patients with rare diseases often encounter limitations in accessing effective treatments due to an incomplete understanding of their pathophysiology, undefined diagnostic criteria, and the absence of specific treatments. This study elucidates the possibilities and challenges associated with leveraging big data in the development of the diagnosis and treatment of rare diseases.

Keywords : Big data, Real-world data, Data science, Rare diseases, Challenges

* Assistant Professor, Graduate School of Health Care Sciences, Tokyo Medical and Dental University