

令和7年度厚生労働科学研究費補助金（障害者政策総合研究事業）
発達障害への地域支援に資する継続的な情報収集・活用方法・体制整備に向けた研究
(24GC1006)

分担研究報告書

匿名医療保険等関連情報を用いたデータベース構築

研究分担者 篠山 大明（信州大学医学部精神医学教室）

研究要旨

本研究の目的は、発達障害に関する情報のデータベース構築に向け、持続可能なデータ収集体制を確立することである。発達障害に関するデータベースは、社会状況や医療体制の変化を反映しつつ経時的に更新される必要があり、そのためには継続かつ効率的にデータを収集できる仕組みの整備が不可欠である。本研究では、既存の情報資源を最大限に活用し、保険診療情報を用いた疫学調査に基づく恒常的なデータ収集・集計システムの構築を目指している。本年度（研究2年目）は、初年度に構築した基盤を踏まえ、ナショナルデータベース（NDB）を用いたデータ収集体制の実装と検証を進めた。具体的には、昨年度に申請した特別抽出データの受理、データ処理環境の整備、解析可能な形式への統合作業を行い、2009年度から2023年度までの自閉スペクトラム症および注意欠如多動症の診断データを収集・集計した。その結果、全国における診断動向の把握を行うことができたと同時に、継続的なNDBデータ取得および動向監視に向けた実務上の課題も明らかとなった。しかし、これらの課題はいずれも対応可能な範囲であり、持続的なデータ収集体制の構築は十分に実現可能であると考えられた。また、保険診療データ自体が有する固有の限界は存在するものの、他の情報源と組み合わせた多面的な情報収集の一環として本手法を確立する意義は大きいと考えられた。今後、本システムが完成すれば、発達障害に関する政策立案や臨床研究を支援する疫学データを持続的に提供する基盤として機能することが期待される。

A. 研究目的

本研究は、発達障害に関する疫学情報を恒常的に収集するための、持続可能なデータベース構築手法を確立することを目的とする。過去に実施された平成30～令和元年度厚生労働科学研究費補助金（障害者政策総合研究事業）「発達障害の原因、疫学に関

する情報のデータベース構築のための研究」[1]においては、ナショナルデータベース（NDB）を活用し、自閉スペクトラム症、注意欠如多動症、知的障害の診断を受けた患者を抽出したうえで、都道府県別・性別・年齢階級別の有病率の算出を行った。この際には、事前に集計条件を設定し、3次元集

計表としてデータ提供を受ける方式を採用した。

その結果、自閉スペクトラム症の累積発生率[2] や注意欠如多動症の診断動向[3] など、重要な疫学的知見が得られた。一方で、知的障害については想定を下回る有病率が算出され、医療費請求と直接関連しない診断名では診療報酬請求データへの登録漏れ（偽陰性）が生じやすい可能性が示唆された。また、集計表方式では事前に詳細な集計条件を指定する必要があるため、事後的な探索的分析や条件変更が困難であるという方法論的な制約も明らかとなった。

これらの課題を踏まえ、本研究（2年間）では、発達障害の疫学情報をより多角的かつ経時的に検討するため、個別の時系列追跡が可能となる「特別抽出」方式によるNDBデータの活用を計画した。研究2年目となる本年度は、初年度（昨年度）[4] に実施した利用申請に基づき受領したデータを用いた疫学調査およびデータ処理一連の手順の実装を行う。このアプローチにより、発達障害に関するデータ収集の精度向上と、将来的な持続的データ収集体制の実現可能性を検証する。

B. 研究方法

本研究では、初年度に行った利用申請（特別抽出）に基づき、2009～2023年度に自閉スペクトラム症または注意欠如多動症と診断された者を対象としたNDBデータを受領した。抽出対象となったデータの条件は以下の通りである。

① 傷病名レコード（SY）：自閉スペクトラム症と注意欠如多動症の傷病名を抽出条件として設定し、出力対象とした。な

お、疑い病名は除外するため、疑い病名フラグ=0を条件に加えた。出力項目には、診療開始日、レセプト通番、レコード順序、レコード識別情報を含めた。

② 傷病レコード（SB）（DPCレセプト情報のみ）：自閉スペクトラム症および注意欠如多動症の傷病名を抽出条件とし、疑い病名フラグ=0であるものを対象とした。傷病名レコード（SY）と同様に診断名で絞り込むが、SYレコードとの組み合わせはOR条件とした。出力項目には、診療年月、レセプト通番、レコード順序、レコード識別情報を含めた。

③ レセプト共通レコード（RE）：レコード順序、レコード識別情報、レセプト通番、ID1、ID2、診療年月、男女区分、年齢階層コード2、満年齢階層コード2を出力対象とした。

④ 医療機関情報レコード（IR）：レコード順序、レコード識別情報、レセプト通番、都道府県を出力対象とした。

入手したデータファイルは、本年度整備した解析環境下において、Python (version 3.14) を用いて解析可能な形式に結合・統合し、各種集計処理を実施した。具体的なデータ処理手順は以下の通りである。

- ・ 調査対象疾患の傷病名が記載されたレセプトの診療開始日を確認する。
- ・ 同一IDが複数回出現する場合は、同一人物とみなし（ID1またはID2のいずれかの一致）、疾患ごとに診療年月が最も古いレセプト情報のみを残す。
- ・ 診療開始日、都道府県、診療年月、年齢階層コード2、満年齢階層コード2を抽出する。
- ・ 医科レセプトおよびDPCレセプトのそ

それぞれで上記手順を実施し、両方に同一 ID が存在する場合は、診療開始日が早い方のデータを採用する。

以上の集計処理を経たうえで、持続的データ収集体制の実現可能性、ならびに課題や限界について検証した。

(倫理面への配慮)

本研究は、信州大学医学部生命科学・医学系研究倫理委員会の承認を得た上で実施している。本研究は、開始前より個人を特定できない匿名化情報のみを利用するため、対象となる被験者からの個別同意取得を行うことはできない。したがって、インフォームドコンセントおよびオプトアウトは実施しない。

厚生労働省より提供された NDB データ保存媒体、ならびに NDB データおよびその中間生成物を保存した外付け記録媒体は、施錠保管を行う。データは、施錠された室内に設置された特定の端末内への保存、バックアップ用外付け HDD への保存、または打ち出し帳票による保存以外の方法では保管しない。

当該端末へのアクセス時には、二要素認証方式を用いて取扱者の識別と認証を行い、端末はインターネット、学内 LAN、院内 LAN など外部ネットワークには一切接続しない。また、台帳管理された外付け HDD や USB 等の記録媒体以外の外部記録媒体を端末に接続しない。端末本体についても、物理的な持ち出しを防止するため、施錠したチェーンにより固定する。

研究終了後は、厚生労働省から提供された保存媒体内の NDB データを消去するとともに、端末内および外付け HDD 内に保

存された NDB データについては、専用のデータ消去ソフトを用いて完全消去を行う。また、打ち出された帳票はシュレッダー処理により適切に廃棄する。

C. 研究結果

初年度（2024年9月）に行った NDB の特別抽出データの利用申請が審査を通過し、本年度（2025年12月）にデータを受理した。結果として、利用申請から提供まで約15か月を要した。

提供されたデータは2009年度から2023年度までの月別データベースファイルであり、各月につき7種類のファイル（医科レセプト：SY, RE, IR；DPCレセプト：SY, SB, RE, IR）に分割されて配布された。総計1,260ファイル、データ容量は約19GBであった。本年度はこれらのファイルを解析可能な形式に結合・統合し、発達障害に関する各種前処理および集計を完了した。

集計の結果、人口10万人あたりの自閉スペクトラム症および注意欠如多動症の診断数は、年度を追うごとに増加する傾向が認められた。成人の診断数については、2019年度から2020年度にかけて一時的な増加の停滞がみられたものの、2021年度以降は再び増加に転じた。とくに成人期の注意欠如多動症の診断数は急増しており、20歳以上の診断数は2012～2013年頃に自閉スペクトラム症の診断数を上回った。

さらに、女性においては20代前半の注意欠如多動症の診断数が近年顕著に増加しており、従来診断のピークであった児童期に匹敵する規模で20代前半に診断されていることが明らかとなった。

D. 考察

①NDB を用いた継続的なデータ収集体制の実現可能性

発達障害に関する疫学情報を継続的に解析する体制の実現可能性を評価するため、データ受領から前処理までを実際に行った結果、以下の実務的な課題が明らかとなった。

- ・ 申請手続きの煩雑さ
- ・ データ利用申請から提供までの長期化（今回は約 15 か月）
- ・ 厳格なセキュリティ要件の順守
- ・ データ構造の複雑さと前処理にかかる負荷

しかし、実際に作業を完了した結果、これらの課題はいずれも対応可能であることが確認された。

申請手続きや前処理に要した作業負荷については、初回は手続きや前処理の方法を新たに整備しながら進める必要があったため、一定の負荷が生じた。しかし、今回確立した一連の手順を次回以降にそのまま適用できることから、作業効率は大幅に向上する見込みである。また、データ提供までの期間についても、定期取得を前提として早期にスケジュールを設定することで、運用上十分に対応可能であることが確認された。セキュリティ要件に関しては、データ保管専用の空間を確保できれば、その他の要件は必要な備品の整備により対応可能である。以上の検討結果から、NDB を用いた継続的なデータ収集体制の構築は、一定の初期準備を要するものの、実務上は十分に実現可能であると結論付けられた。

②データ自体の限界について

解析を進める中で、保険診療情報に固有の以下の限界も確認された。

- ・ 保険病名によるバイアス

医療費請求を目的とした「保険病名」が含まれるため、診断の実態と完全に一致しないケース（偽陽性）が生じる可能性がある。ただし、発達障害では診療報酬上の処方薬等の選択肢が限定的であるため、他の疾患と比較して偽陽性の頻度は低いと推測される。一方で、知的障害のように医療費請求に直結しない病名では登録漏れ（偽陰性）の懸念が残る。

- ・ ID リンクの限界

集計においては、被保険者番号等に基づく「ID1」と、氏名・生年月日・性別に基づく「ID2」の2種類のハッシュ化 ID を用いて個人を識別する。ID2 の特性上、同姓同名・同一生年月日・同性の別人が統合される理論的リスクがあるが、発生確率は低く無視できる水準である。しかし、保険者変更（ID1 の変化）と婚姻等による姓の変更（ID2 の変化）が同時に発生した場合、同一人物が別人として扱われ、新規患者として重複計上される（発生率の過大推定）可能性がある。日本では婚姻に伴う改姓の約 95%が女性であり、特に 25～34 歳層の女性においてこの誤分類リスクが相対的に高くなる点には、結果の解釈において留意が必要である。

③構築したシステムで算出可能な指標

本研究期間中に構築した前処理手順により、今後以下の発達障害に関する主要な疫学指標の算出が可能となる。

- ・ 出生年度別・都道府県別の 9 歳になる年

度までの累積発生率

- ・ 年度別・年齢層別の診断率
- ・ 自閉スペクトラム症と注意欠如多動症の併存診断数

これらの指標は、発達障害の診断動向を多角的に把握するうえで重要な基礎データとなる。詳細な疫学的解析結果については、本手法に基づく集計が完了次第、学術論文等にて順次報告する予定である。

④今後に向けた課題と展望

本研究（2年間）を通じて、データ収集プロセスの第一段階である NDB 利用申請から、データ処理環境の整備、実際のデータ結合・統合までを達成し、データ収集の基盤を確立することができた。今後は、収集したデータの本格的な解析をさらに進めるとともに、継続的な診断動向の把握に向け、抽出条件の適正化や再評価を行っていく予定である。また、定期的なデータ提供を実現するための運用体制について、厚生労働省との実務的な協議を進めていく必要がある。

⑤総合的評価

保険診療情報を用いた疫学調査は、全国規模の悉皆的なデータを比較的簡便に把握できるという極めて大きな利点を有する。一方で、保険病名特有の性質、ID リンクの限界、居住地情報の欠如、人口移動の反映不足など、いくつかの制約を伴う。しかし、これらの限界を適切に認識しつつ継続的なモニタリング体制を構築すれば、発達障害に関する地域支援体制の整備や政策立案に資する、非常に強力な基盤データとなる。地域ごとのニーズの違いやその経時的変化を的確に把握するためにも、本研究で構築した

システムを基盤とし、恒常的に保険診療情報を活用できる体制の維持・運用が強く求められる。

E. 参考文献

- [1] 厚生労働科学研究費補助金障害者政策総合研究事業:発達障害の原因、疫学に関する情報のデータベース構築のための研究—平成 30 年度～令和元年度総合研究報告書, 2020.
- [2] Sasayama D, Kuge R, Toibana Y, Honda H. Trends in Autism Spectrum Disorder Diagnoses in Japan, 2009 to 2019. JAMA Netw Open. 2021 May 3;4(5):e219234.
- [3] Sasayama D, Kuge R, Toibana Y, Honda H. Trends in Diagnosed Attention-Deficit/Hyperactivity Disorder Among Children, Adolescents, and Adults in Japan From April 2010 to March 2020. JAMA Netw Open. 2022 Sep 1;5(9):e2234179.
- [4] 厚生労働科学研究費補助金障害者政策総合研究事業:発達障害への地域支援に資する継続的な情報収集・活用方法・体制整備に向けた研究—令和 6 年度総括・分担研究報告書, 2025.