

「NDB の迅速提供に向けたスキーム再構築に資する研究」中間報告（案）

○はじめに

令和5年6月の規制改革実施計画において、厚生労働省はNDBについて、利用申請から申請者が実際にデータの利用を開始し得るまでに要する期間について、平均で390日を要する現状から原則7日(研究者等側の都合に要した期間は除く。)とすることとされ、NDB提供体制の抜本的な見直しが進められる。本研究は、NDBデータの利活用の推進のために、令和5年度において対応が必要な政策・技術的課題の検討と、将来的なNDBのあり方も含めた課題解決に向けた具体的提言を行い、厚生労働行政に直接的に寄与することを目的としている。

NDBデータの提供については、現在運用が開始されつつあるHICでの提供の実現によって、新たにNDBを用いた研究を始めようとする者の初期投資を大幅に削減することが期待され、今後規制改革実施計画においてとりまとめられたような迅速なデータ提供が実現すれば、劇的に研究促進につながることを期待される。他方、単なるアクセス向上のみを目的とするのではなく、NDBが有する個人特定性等の課題を踏まえつつ、研究に必要な情報を迅速に提供し、結果として、公衆衛生の向上や厚生労働行政に資するNDB利活用を実現させることが重要である。そこで、本研究班では、短期的なアクセス向上への対応にとどまらず、NDBが政策や研究の発展に資するために長期的に目指すべき方向性となることを目的として、まず、NDBの特徴を踏まえた利活用のあり方を検討し、効率的な提供に資するデータの仕様及びそのための提供体制を議論し、更に、現在の審査体制の課題等について、厚生労働省において提供業務に携わったことのある複数の研究協力者に参画頂き、検討を行っている。本中間報告においては、NDBの特徴を踏まえた利活用のあり方や令和6年度に開始するデータ提供内容についての具体案について、専門委員会での議論に資するよう報告する。

○NDBの特徴を踏まえた利活用のあり方

NDBは膨大で悉皆的なデータを保有しているという利点が最大の特徴であり、患者重症度等の調整が困難であるという弱点はあるものの、この利点を活かした厚生労働行政の評価や政策利用の手法開発等を推進することが必要である。また、NDBは診療報酬請求のため集められた各月のレセプト毎のデータベースであり、必ずしも研究のための利便性を考慮して構築されたものではない。このため、一般的なデータベースハンドリングや統計解析、疫学等の知識に加えて、レセプト構造の理解、コードの解釈のための診療報酬の知識が必要である。次項で詳述するがNDBを研究活用のためのパネルデータとして活用するためには、レセプト単位で作成されるデータを個人単位化する等の前処理が必要となる。このようなNDBの利便性を向上させるための前処理の一部は、これまでも厚生労働省において行われてきているが、NDB分析に関する知見は今も発展段階であり、今後広く利活用するため

には、こうした個人単位化するために必要な情報の付与やデータベース構造の見直し等のデータベースを扱いやすくする不断の見直しが必要である。そのためには、研究者は利活用を通じて得られた NDB の分析に関する知見を厚生労働省に還元し、厚生労働省はその知見に基づき、次項で提案するような迅速提供のためのデータの整備を優先的に行いつつ、今後にも新たに得られた知見を優先度に応じて発展させるようなアジャイル型のデータベースの発展させていくことが望まれる。

○規制改革実施計画の対応に向けたデータの仕様と提供体制

NDB を用いた多くの研究は、レセプト単位ではなく個人単位で行われるため、NDB データが提供された初期段階において、個人単位のパネルデータにするために必要な情報の付与やその他研究に必要なデータを作成するために、データベース構造を再成形する処理等の前処理が実施されることが一般的である。迅速かつ効果的な NDB の政策活用や研究利用を発展させるためには、現時点で妥当と考えられる情報の事前の付与を行いつつ、将来的には、利用しやすい提供用のデータベース構造を設計可能か等も検討を進めることが望まれる。

原則 7 日での提供にむけては、これまで、研究に必要な NDB データのみを抽出したデータセットが研究者に提供されていたため、抽出から提供までの時間がかかる要因となっていた。他方、提供期間短縮のために、全ての研究者が全てのデータに実質的な制限なくアクセス可能とすることは、個人特定性のリスクやインフラ維持のコストを考えたときに現実的なものとは言えない。研究利用の観点においては、全ての研究において必ずしも NDB に格納された十数年分の全てレセプトデータが必要ではなく、一定期間のデータについて、1/1000 を抽出した場合でも外来入院合わせて年間で延べ 200 万レセプト程度のデータとなるため、一定の探索的な解析を含め、多くの研究には十分な数となる。この十分量のデータについて、事前のデータ処理や HIC 上での提供を合わせて行うことで、個人特定性のリスクを減じることができ、コストも低減できると考えられる。従って、症例の縦断的追跡を可能となるような情報を付与・加工し、1～数年分の十分なデータ件数を確保したデータの提供を広く行い簡易な審査によって提供することで、短期間で安全性の高い NDB データ提供が実現するものと考えられる。前述の加工済みデータよりも多くのデータが必要な研究については、追加的な審査の上で、提供された環境に追加的なデータ利用を可能とすることにより、初回の利用から切れ目無い研究環境を構築できると考えられる。

○その他

このほか、一部の研究者は頻回に NDB を利用しているが、研究を始める際に様々な課題があり、新規で NDB 研究を行おうとする者は多くはない現状がある。新規参入者を増加させるためにも、利便性の高い分析環境の構築とともに、研究者への支援体制も必要である。研究を始める上で特に困難となるのは、アクセスする NDB のデータ構造に関する理解や、そ

れに基づく前処理手法、マスタ整備等がある。こうした課題に対しては、これまで複数の分担研究者が NDB に関する e-learning 等を開発し、データベースを再構築するための提案等が行われてきている。こうした NDB に関する分析等の知見を一元的に集約したプラットフォーム環境の構築が期待される。このほか、上記のような仕組みを前提とした申請・審査体制の再構築も必要である。複数のデータベースの連結解析に係る申請の際に同一の書類を複数の事務局に提出しなければならない、必ずしも個人特定性のリスクが低いデータセットの利用にも厳格に審査している、等の課題があり、データ利用における安全性確保、特に HIC 特有の解析環境の課題やアクセス向上による新たなリスクへの対応等も前提としつつ、合理的でない申請業務は書類審査の実現等、研究者の負担軽減の観点に立った運用改善が必要である。