

別添3

令和5年度厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業） 総括研究報告書・分担研究報告書

NDBの迅速提供に向けたスキーム再構築に資する研究

研究代表者 山本 隆一（一財）医療情報システム開発センター・理事長
研究分担者 松田 晋哉 産業医科大学 公衆衛生学教室・教授
研究分担者 今村 知明 奈良県立医科大学 医学部・教授
研究分担者 今中 雄一 京都大学 大学院医学研究科・教授
分担研究者 康永 秀生 東京大学 大学院医学系研究科・教授
分担研究者 満武 巨裕 医療経済研究機構・副部長

研究要旨 NDB（匿名レセプト情報・特定健診等情報データベース）は平成25年度から第三者提供を実施し、令和2年には法制化され民間事業者へも提供が拡大した。また、介護DBとの連結整備も行われ、医療・介護サービスの利用状況の複合的な解析が可能となった。令和4年度からはDPCDBとも連結解析可能となり、NDBのクラウド化によりデータ取込や抽出の効率化を可能とし、国の保有するDBの利便性の向上が図られた。また、他の公的DBとの連結や死亡情報との連結の検討、クラウド上で解析環境を提供する医療・介護データ等の解析基盤（HIC）の開発も進められ、研究の利活用が益々促進されている。他方で、NDBデータ提供に関しては、提供審査の承諾後も、提供データが研究者に届くまで一定時間を要しており、研究の新規性の喪失等の学術的不利益のみならず、研究費執行期限内にデータ利用ができない等の課題が生じている。要因は様々あるが、研究者のNDB構造の知識不足により抽出条件確定までに時間を要する場合も多い。本研究では、次世代医療基盤法のDB等との連結利用でニーズの高まりも予測される状況を勘案し、限られた資源を有効活用し、医療費適正化計画の作成等の政策の活用及び研究利活用を共に推進するため、本研究班では、現状のNDB提供遅延の要因を検討し、出来る限り迅速なデータ提供を目指すべく課題解決への提言を行う。なお、主任研究者・分担研究者は一体として研究を行ったため、すべての成果を総括報告書で一体として報告する。

1. 背景と研究目的

匿名レセプト情報・特定健診等情報データベース（以下、NDB）は、医療費適正化計画の作成、実施及び評価の調査や分析に用いるデータベース（以下、DB）として、平成20年にレセプト情報等の収載を開始した。平成25年度から第三者提供を実施し、令和2年には法制化され民間事業者へも提供が拡大し、介護保険総合データベース（以下、介護DB）との連結規定を整備し、医療・介護サービスの利用状況の複合的な解析が可能となった。令和4年度からはDPCのDBとも連結解析可能となり、NDBのクラウド化によりデータ取込や抽出の効率化を可能とし、国の保有するDBの利便性の向上が図られた。また、他の公的DBとの連結や死亡情報との連結の検討、クラウド上で解析環境を提供する医療・介護データ等の解析基盤（HIC）の開発も進められ、研究の利活用が益々促進されている。

他方で、NDBデータ提供に関しては、提供審査の承諾後も、提供データが研究者に届くまで一定時間を要しており、研究の新規性の喪失等の学術的不利益のみならず、研究費執行期限内にデータ利用ができない等の課題が生じている。要因は様々で、厚労省側の資源の制約等もあるが、研究者のNDB構造の知識不足により抽出条件確定までに時間を要する場合も多い。他DBとの連結等でニーズが予測される中、限られた資源を有効活用し、医療費適正化計画の作成等の政策の活用及び研究利活用を共に推進するため、現状のNDB提供遅延の要因を検討し、課題解決への提言を行う。具体的には、知見のない研究者に対し、NDB分析の知見や経験を蓄積し、能動的かつ伴

走型でアドバイスする支援体制の構築と支援内容の公共財化や、提供実態の分析等を通じた最適な提供方法の検討、クラウド化での効率的なデータ提供体制の構築等が考えられる。

本研究の研究者らは、平成23年当初より「レセプト情報等の提供に関する有識者会議」審査分科会の構成員を務めており、本研究を主導し取り纏めを行う研究代表者は、現在NDBデータの提供の審議を行う「匿名医療情報等の提供に関する専門委員会」で令和2年発足当時からの委員長であり、NDBデータの利活用方針の議論にも携わってきた。研究分担者は何れもNDBを最も活用する研究者で、ユーザー側の視点も保有する。厚労省でNDBの提供業務に従事した経験を有する研究者が研究協力者として参画することで、ユーザー側、提供側各々の課題の知見を有する研究班体制を構えた。

2. 研究方法

以下の2つの方法にて研究を進める。

2-1. NDB研究者支援体制の構築および提供データの検討

研究班にて、以下の事項について、過去の使用実績等を踏まえた議論やその他のNDBのユーザー等へのヒアリング結果などを元に検討を行う。班会議や分担項目ごとの打合せを月1回程度の割合で開催し、最終的な提言にまとめる。

- ・研究支援体制に関する具体的事項。参加する研究者（アドバイザー）の要件・資質や支援内容、具体的な支援の方法公共財とすべき知見の具体的内容や研究論文へのauthorship等のルールの整備、アドバイザーの具体的なメリットのあり方等

・研究者に提供するデータの新たなデータ提供方法や形式の検討(例:活用しやすい解析用データベースの提供、簡易な分析が可能とする体制(ダッシュボード機能やオープンデータの拡充等))

2-2. 諸外国のデータ利活用の調査

NDBの抱える課題解決のために、諸外国におけるデータ利活用についてのヒアリング調査を行った。調査対象先は、我が国と比較的保険制度の似通った台湾および韓国、先進的なデータ利活用の事例としてフィンランドの3か国の調査を実施した。

3. 結果

3-1 NDB研究者支援体制の構築および提供データの検討

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。

3-1-1 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。分析目的もよるが、一般的には特定の疾患群について入外のレセプト情報を結びつけて分析するニーズも多いことを考えると入院レセプトを起点として前後の一定期間のレセプトを連結したパネルデータが多くの利活用者のニーズを満たす可能性が高い。

また作成するパネルデータセットの信頼性を確保する観点から1人1番号化された被保険者番号から導出されるID5を用いて連結することが望ましく、当面はID5導入後のデータを対象にすることが妥当と考えられる。

また一方で、多少の精度の低下は許容し

ても、包括的な分析をおこないたい要望にも応えたほうが良い可能性もある。このような目的にはサンプリングではなく、全数を対象とすることが必要になる。このような利活用は個人特定のリスクの判定が難しく、これを実現するためには、個人特定につながる可能性のある項目を比較的広範にマスクする必要がある。HIC上で比較的大きなワークスペースが必要で、分析コストの上昇は避けられないが、ニーズが存在するのであれば対応すべきであろう。

3-1-2 利活用者支援

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。

なお以上の成果の骨子は令和5年12月に開催された第18回匿名医療情報等の提供に関する専門委員会で報告した。

3-2 諸外国におけるデータ利活用の調査報告

諸外国におけるデータ利活用の調査報告は、別紙①、別紙②、別紙③に記す通り。

4. 考察

これまでのNDB利用申請を考えると、一定程度マスタが整備された通年パネルデータセットを作成し、安全性を高めるために1%程度にサンプリングしたデータセットを作成し、簡易審査で提供することが多くのニーズを満たして上で最短7日程度で提供できることが考えられた。一方でオープンデータでできない概要的な調査を行うためには全レコードアクセスが必要になるが、これを短時間で提供するためには個人特定に繋がりうる項目をかなり多数マスキングする必要がある。その上HIC上に大きなワークスペースが必要で、利用は限定されると思われるが、対応するために、安全性を高めた全データセットも用意する必要があるかも知れない。さらには申請手続きの簡素化や利活用者支援も必要で、申請手続きに関してはポータルサイトの構築が進められているようであるが、利活用者支援は今後も継続的に検討する必要がある。

海外の状況として、台湾、フィンランド、

韓国を調査したが、いずれもそれぞれ工夫を凝らした医療情報の提供を行っている。申請から提供までの時間はいずれも現状のNDBよりは短い傾向にあるが、最短でも1.5ヶ月かかっているようである。またある程度のデータクリーニングが行われているためもあるが、データ発生から利用可能になるまでの時間はNDBに比べて長い。

5. 結論

利活用申請から提供までを最短7日で行うことは海外事例の調査から見てもチャレンジングであるが、可能性のある手法として0.1~1%程度にサンプリングした通年パネルデータセットの提供と、大幅に安全性を高めた全数データセットが考えられた。ただしHIC利用を前提とするために全数データセットは利活用経費が高騰することは避けられない。

ポータルサイトの整備は計画されているが、マスタ等分析ノウハウの蓄積は今後の運用に依存し、利活用者支援とともに今後期待したい。