

厚生労働科学研究費補助金(循環器疾患・糖尿病等生活習慣病対策総合研究事業)
分担研究報告書

National Database (NDB)等の電子レセプト情報の研究応

研究分担者 中山 健夫 京都大学医学研究科 健康情報学分野

要旨: レセプト情報・特定健診等情報データベース(NDB: National Database of Health Insurance Claims and Specific Health Checkups of Japan)は、特定健診・特定保健指導情報(2008年度～)とレセプト情報(2009年度～)から構成されている厚生労働省管理のデータベースである。2017年時点で日本における保険診療の95%以上の請求情報が含まれており、2011年より医療サービスの質の向上等を目指した正確なエビデンスに基づく施策の推進等を目的とした分析・研究に対し、有識者会議での審査を経て政策立案者や研究者が用いることが可能となった。慢性期における脳卒中を含む循環器病診療の質の評価において、NDBがどのような価値を持ち得るか検討を行うため、初年度はNDBの現状と利用可能性を概観した。

A. 目的

レセプト情報・特定健診等情報データベース(NDB: National Database of Health Insurance Claims and Specific Health Checkups of Japan)は、特定健診・特定保健指導情報(2008年度～)とレセプト情報(2009年度～)から構成されている厚生労働省管理のデータベースである。2017年時点で日本における保険診療の95%以上の請求情報が含まれ、2011年より医療サービスの質の向上等を目指した正確なエビデンスに基づく施策の推進等を目的とした分析・研究に対し、有識者会議での審査を経て政策立案者や研究者が用いることが可能となった。

初年度はNDBの現状と利用可能性の基礎的検討を行った。

B. 研究方法

文献的検討、データベース研究。

C. 研究結果

1. データベースに格納されている情報

NDBは大きく分けて「レセプト情報」と「特

定健診・特定保健指導情報」によって構成されており、レセプトには氏名や生年月日・性別などといった患者情報、医療機関や保険者、被保険者等の情報、保険診療に関連する傷病名や治療内容、投薬等の情報が含まれている。また個人を特定するIDとして、主に保険者、被保険者関連情報を元にしたハッシュ値(ID1)と、主に氏名情報を元にしたハッシュ値(ID2)の、2種類のハッシュ値が個々のレセプト情報に付与されている。このID1,2を用いることで、医療機関、各月ごとに発行されているレセプト情報を同一人物の情報として特定することが可能となる。特定健診情報は、40歳以上75歳未満の被保険者・被扶養者を対象とする内臓脂肪型肥満に着目した問診結果や生活習慣病に関連した測定項目の結果、あるいは特定保健指導については保健指導レベルや支援形態などの情報が含まれており、レセプト情報と同様にID1,2が付与され格納されている。格納されているデータの詳細については、厚生労働省のホームページから参照できる。

2. 利用形式

NDB データは様々な利用形式が設けられているが、実際に利用者がデータを操作する形式としては、大きく分けて主に「特別抽出」、「サンプリングデータセット」の2つがある。研究者が必要と考えるデータすべての提供を申出する形式は、「特別抽出」と呼ばれている。有識者会議(審査分科会)において承諾されれば、申出者は希望するNDB データを入手できる。事前に限定された目的に対する完成度の高い研究プロトコルの準備と、合理的な根拠に基づいたデータ項目の指定が必須となっている。また縦断研究を行えるのは特別抽出のみであるが、ID で統合し、分析できるデータ形式を再構築するなど分析の前処理が必要なため、データ分析に関する人材や物理的分析環境が整っていることが不可欠であり、現実的にはレセプトデータの分析に相当習熟した研究者、研究チームに利用者は限られよう。

特別抽出では探索的な研究が原則として認められていないことから、そうしたニーズに対応出来るよう整備されたデータが、「サンプリングデータセット」である。これは、単月分のレセプトに対し、性別および5才刻み年齢別に、入院で10%、外来および調剤で1%の抽出を行い、出現回数の少ない傷病名や診療行為、医薬品情報のダミー化など、匿名性を強化する一定の処理が行われたデータである。サンプリングデータセットは単月分の情報しか含まれていないため、縦断研究が出来ないという欠点があるが、特別抽出に比べるとデータ量が限られており、ID で紐付けする作業も不要であることから、比較的現実的な環境下で操作することが可能である。厚生労働省ホームページにこれまでの利用者による体験報告も紹介されており、各疾患の出現頻度や治療実態を横断的に評価することを検討している研究には比較的利用しやすいと言える。

3. NDB オンサイトリサーチセンター

NDB の利用にあたっては十分なセキュリティ環境の確保も求められているが、その環境を誰しも確保できるとは限らない。そこでセキュリティが確保された環境下で多様な研究者がNDB データにアクセスできることを目的として、NDB オンサイトリサーチセンター(以下「センター」)が2015年に設置された。センターは東京大学と京都大学に設置されるとともに、両機関はNDB データの利活用の推進・普及・啓発に関し厚生労働省との連携協力機関として協定を締結することとなり、NDB データの利活用を推進する役割を担うこととなった。

センターでのNDB データ利用はまず両機関において試行的に利用が開始され、2016年は主にセンターが備えるシステムの性能を評価するパフォーマンステストが行われた。センターの端末には様々なアプリケーションや統計解析ソフトが用意されているが、目的とする集計・分析によってアプリケーションの適性が異なっていたり、ローカル環境で統計解析する場合とサーバー上で統計解析する場合とで所要時間等に大きな差が出たりするなど、システム性能に様々な特性があることが明らかとなった。現在、具体的な研究課題の実行可能性を評価中であり、今後は両機関以外の研究者等もセンターを利用できるようにするため、活用方策の検討や諸規程の整備が行われる予定となっている。

4. NDB データを研究に使用する際の課題

NDB データの第三者利用は2011年以降徐々に進んでおり、申出件数も2017年3月時点で125件の研究申出に対してデータ提供が承諾されている。一方で、NDB データの更なる利活用を考えた場合、様々な課題が浮かび上がってくる。まず、レセプトは月ごと、医療機関ごと、レセプト種類ごとに発行されるため、特別抽出で研究を行う場合、患者単位でデータを統合する作業が必要と

なるが、この作業に際しての定形化された ID 統合ロジックが未だ確立されていない。現在の ID1 と ID2 は、就職先の変更や氏名の変更、記入の揺れなどによって変化が生じるため、結婚や就職などといったライフイベントのたびに NDB データの ID の精度が減じる。ID 統合ロジックが確立されていないため、研究者によって得られる結論が異なるという懸念が付き纏っている。このことは、縦断研究のハードルを高くしていると共に、データ分析に多大な年月を必要とすることとなるため、理論上診療後 2~3 か月でデータベース化されるはずの NDB データの鮮度を結果的に落としてしまうことになっている。

またレセプトデータの構造や特性、効果的な分析手法や活用方法も、研究目的での二次利用という観点から見て、未だ知見が網羅的に蓄積されているとは言い切れない。患者単位でのデータ統合ロジック以外にも、いわゆる「レセプト病名」の存在など、傷病名の精度を揺るがせる要因もあり、適切にデータを抽出する手法が未だ確立されていない。特定の疾患の患者を抽出する際には、傷病名情報以外に、行われた診療行為の情報や処方された医薬品の情報を活用することで一定程度バリデートすることが可能ではあるが、それら手法についても確立されたものは存在しておらず、各研究者が手探りでやっているところである。研究目的で NDB データを利用してきた者が、今後切り拓いていかなければならない課題であろう。

米国では Medicare, Medicaid データの研究利用が盛んに行われているが、データ利用時は、データを管理する CMS と契約関係にある Research Data Assistance Center (ResDAC) という組織が、利用申請手続きに関する相談や研究者からの学術的質問を受け付けている。日本では一部業務は外注されているものの、利用者に向けたデータ利用を支える学術的支援体制が十分に確

立されているとは言い難く、現時点では一部公開されている関連資料を頼ったり、あるいは既に研究利用の知見を有している研究者と共同でデータ利用を行ったりするなど、現実的には分析のノウハウを何らかの形で確保しなければ、NDB データの十全な利活用は容易ではない。

D. 考察 & E. 結論

現時点ではさまざまな課題を持つ NDB であるが、日本のほぼ全人口を対象とした特定健診とレセプトの最大規模のデータが集積されている貴重な情報源であることは確かである。

これまで地域集団を対象にコホート研究が進められてきた循環器疫学の領域において、NDB がどのような価値を持ち得るか、引き続き検討を進めたい。

研究事例

レセプトを用いた循環器病診療の質評価：壮中年期の虚血性心疾患患者における心臓リハビリテーション

虚血性心疾患 (Ischemic heart disease: IHD) の再発予防治療の一つに心臓リハビリテーション (以下、心リハ) がある。心リハとは医学的評価に基づく運動療法を中心に、入院中あるいは退院後から約 2-3 か月間、運動や食事など生活全般にわたる指導・教育を行う包括的プログラムである。心リハの有効性については、これまで壮中年期患者を中心に多くのランダム化比較試験が実施され、コクランレビューで死亡リスクと再入院リスクの有意な減少が確認された。その効果は冠動脈バイパス術 (Coronary artery bypass graft surgery: CABG)・経皮的冠動脈形成術 (Percutaneous coronary intervention: PCI) のいずれの手術を受けた患者においても認められ、国内外の複数の

診療ガイドラインでIHD患者に対する心リハの実施が推奨されている。一方、心リハはその高い有効性にもかかわらず普及が進んでいないことが各国共通の課題となっている。本邦では一定の施設基準を満たした認定施設のみが心リハを提供可能であるが、認定施設自体が不足している。本研究ではNDBではなく、健保レセプトを用いて、壮年期の虚血性心疾患患者に対する心リハ実施状況の評価を行った。

2006年4月から2013年8月の間に、IHDに対しCABGまたはPCIの手術を受けた20歳以上65歳未満の患者で、IHD患者の抽出には、ICD-10コードのI20-25「虚血性心疾患」を対象とした。I201「記録された攣縮を伴う狭心症」、I241「ドレSSLラー症候群」、および「疑い」病名は除外した。またCABGおよびPCI患者の抽出には診療点数早見表区分コードを用い、K546-552-2を対象とした。さらに、レセプトの転帰情報を用い、当該入院中の死亡症例を除外した。期間中に複数回の手術があった場合は初回の手術のみを対象とし、同一入院内に両方の手術を行った場合はCABGを採択した。また、当該入院月から2ヶ月以内に再度手術を行っている場合は一連の治療とみなした。主要なアウトカムは、心リハ実施の有無とした。心リハ実施は当該入院月以降、退院後2ヶ月以内に、心大血管疾患リハビリテーション料(以下、心リハ料)IまたはIIを1回以上算定したものと定義した。また、CABGおよびPCIを受けた医療施設が心リハ実施施設か否かも評価した。本研究で使用したデータベースは各医療施設の心リハ施設認定に関する情報を含まないため、心リハ実施施設の定義には対象期間中の各医療施設における心リハ料算定状況を利用した。各施設で心リハ料算定を認めた場合、初回の心リハ料

算定月以降を心リハ実施施設と定義した。さらに、退院後の外来心リハ実施の有無についても評価を行った。外来心リハの定義は、心リハ実施者のうち、入院外レセプトによる心リハ料の算定が1回以上あったものとした。データベースから最終的に1860名が抽出され、心リハが実施されていたのは440名(24%)であった。心リハ実施施設で手術を受けた907名(49%)のうち、心リハ実施者は434名(48%)、外来心リハ実施者は77名(9%)であることが明らかとなった。

G. 研究発表

- 中山健夫. National Database:厚生労働科学・戦略研究の経験から. 日本臨床疫学会第1回年次学術大会 シンポジウム1 ビッグデータを用いた臨床疫学研究(東京大学) 2017年9月30日
- 加藤源太, 中山健夫. レセプト情報・特定健診等情報データベース(NDB)を用いた研究の動向: NDBデータを用いた臨床研究. *Progress in Medicine*. 2018;38(2): 123 -126.
- Kanazawa N, Ueshima K, Tominari S, Nakayama T. Underuse of Cardiac Rehabilitation in Workers With Coronary Artery Disease - Claims Database Survey in Japan. *Circ J*. 2017;81(10):1424-1431.

H. 知的所有権の取得状況

なし

ナショナルデータベース (NDB)

- 厚生労働省 (2009年～) が構築・運営する特定健診データとレセプトデータによるデータベース
- レセプト情報・特定健診等情報データベース
National Database of Health Insurance
Claims and Specific Health Checkups of Japan
(NDB)
- レセプト (診療報酬明細)
年間約20億件、累積100億件超

1

NDBの「潜在的価値」

超高齢者を含む「人間」に対し
国民皆保険制度の下で
どのような医療が行われているか
一億人規模の人口を擁する国レベルで
解明できる
現時点で世界で唯一
そして最大のデータベース

2

NDB利用の方法

• 特別抽出

- 申し出からデータ提供に時間を要する
- テーマによるとデータ量が膨大になるためハンドリングが難
- 時系列の連結には今後の共通ID導入に期待

• サンプリング・データセット

- 各年の10月分の入院1/10、外来1/100 のサンプリング
- 横断研究に適する
- 都道府県のデータなし

• オーダリング集計

• NDBオープンデータ

- 主な診療行為・処方の都道府県別の集計データ（2016～）

• オンサイトリサーチセンター（東京・京都）

3