

分担研究名 NDB データ研究の実践と人材育成プログラム開発・検証

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:

【目的】 NDB サンプルングデータセットの使用経験を通して、特別抽出と比較した NDB サンプルングデータセットの意義、課題、限界を考察し、NDB データ研究の実践に必要なリソース、知識、スキルを検討した。

【方法】 平成 24 年、25 年、26 年、各年 10 月の、65 歳以上の死亡患者の医科入院、DPC レセプトデータを用い、終末期高齢者の診療内容を解析した経験を通し、1) サンプルングデータの代表性、2) 横断研究の実施可能性、3) 利便性、4) NDB 解析に必要な知識、スキルを検討した。

【結果】 1) 代表性: サンプルングデータから 65 歳以上高齢者の死亡の約 80% を把握可能であり、終末期医療の実態を国レベルで把握するのに十分足るデータと考えられた。2) 横断研究の実施可能性: 死亡日から一定期間遡った診療内容の把握が可能であった。しかし、7 日間以上の長期間の診療内容を把握出来るサンプル数は限られていた。また、医療供給側要因、1 ヶ月以上前の診療経過を考慮した解析には限界があった。3) 利便性: 特別抽出に比べてセキュリティ要件が緩和された環境で、探索的分析が実施可能であった。4) NDB 解析には、診療報酬制度、データベース、レセプトデータに関する基礎的な知識が必要であった。

【結論】 NDB サンプルングデータを活用し、国レベルで終末期高齢者の診療実態を解明する横断研究が実施可能であった。今後、データ利用経験の共有等により、セキュリティ環境や大規模データベースの構築環境が困難な研究者にも、活用を推進することが重要と考えられた。

研究協力者

酒井 未知 京都大学医学研究科 健康情報学分野

加藤 源太 京都大学医学部附属病院診療報酬センター

大寺 祥佑 京都大学医学部附属病院医療情報企画部

高齢者の診療内容を解析し、サンプルングデータの、1) 代表性、2) 横断研究の実施可能性、3) 利便性、4) NDB 解析に必要な知識、スキル、リソースについて検討した。

解析対象は、平成 24 年 10 月、25 年 10 月、26 年 10 月、のいずれかの診療年月において、年齢が 65 歳以上、入院理由に「死亡」が記録された高齢者の医科入院レセプト、DPC レセプトデータとした。

A. 目的

NDB サンプルングデータセットの使用経験を通して、特別抽出と比較したサンプルングデータセットの意義、課題、限界を考察し、NDB データ研究の実践に必要なリソース、知識、スキルを明らかにする。

1) 代表性については、(A) サンプルングデータから得られた観察死亡数を、(B) 人口動態統計に記録された死亡者数と比較して検討した。(A) 観察死亡数は、1 ヶ月分のサンプルングデータで把握出来た死亡数(* 12 ヶ月で年間死亡数を推計)を、入院レセプトの抽出率 10% で除した数とした。(B) は、サンプルングデータと同年の人口動態統計における、同年齢層の医療機関内での

B. 研究方法

NDB サンプルングデータセットを用い、終末期

死亡者数とした。

C. 研究結果

NDB サンプルングデータセットを用いることで、国レベルで終末期高齢者の診療実態を解明する横断研究が実施可能であった。特別抽出データと比較した利点、限界を、(表 1. サンプルングデータと特別抽出データの比較)に示した。

C.1 代表性

本研究で対象としたサンプルングデータから得られた(A)観察死亡数は2,010,840人、(B)人口動態統計に記録された同年齢層の医療機関内での死亡者数は2,610,429人であった(1)。サンプルングデータから、65歳以上高齢者の死亡の77.0%を同定可能であることから、本邦の終末期高齢者の診療実態把握に十分活用可能と考えられた。

しかし、一部のデータは、個人識別リスクを軽減するため、匿名化や削除が行われており、セクションバイアスに留意する必要がある。例えば、サンプルングデータでは、700,000点以上の医科入院レセプトは削除されている(2)。そのため、終末期医療費の解析ではセクションバイアスの限界がある。また、出現頻度が0.1%未満の傷病名、診療行為、医薬品のコードは匿名化されており、該当する疾患の患者、診療行為、医薬品処方を受けた患者を対象にする解析にも限界がある。

C.2 横断研究の実施可能性

サンプルングデータの医科入院、DPCレセプトには、患者IDに該当するデータが含まれていないため、レセプト単位の分析のみ可能である。入院日のデータは利用可能であるが、1ヶ月分のデータではレセプトに特有の入院日記録ルールに沿った検証に限界が大きいことから、入院エピソード単位の分析には不向きと考えられた。

本研究では、入院レセプトの転帰が「死亡」と記録される前の1ヶ月間を、「終末期」と定義し、レセプト単位で診療実態を評価した。平成24年

以降のレセプトには、各診療行為の実施日、医薬品の処方日が記録されている。これらの日付情報を用いることで、死亡を起点として、死亡日から1ヶ月間遡って、死亡前7日間、30日間の生命維持治療(心肺蘇生術、人工呼吸、中心静脈栄養)の実施割合、ICU入室割合とした解析を実施した。また、平成24～平成26年の経時的な診療実態の変化を検討する、repeated cross sectional studyも実施可能であった。

各年10月に死亡が記録された65歳以上全例のうち、10月中の診療実日数7日以上は約70%程度(1)、30日間の診療内容を把握出来るデータ件数は少なく、死亡前7日以上長期間の診療内容の把握に限界があった。また、終末期の生命維持治療の差し控え(withhold)と中止(withdraw)は意味が異なるが、1ヶ月以上前の診療内容が把握出来ないため、新たに開始された治療と、継続されていた治療を区別することが出来なかった。また、既往歴は重要な関連要因と考えられるが、10月のレセプトに記録された傷病名、診療行為、医薬品から把握できる情報は限られていた。なお、現状では平成27年までのデータしか提供されていないため、平成28年度以降の診療内容の検討は出来ない。

また、医療供給側因子による終末期医療のばらつきを明らかにする必要性は高いが、サンプルングデータに医療機関のID、都道府県、病床規模のデータは含まれていないため、医療供給側要因の解析には限界がある。病棟区分のデータを用いることで、医科入院の一般病床、療養病床、DPC病院の3群別の解析を行うことは可能であった。今回は実施しなかったが、入院基本料の請求コードに基づき、病棟を区分する解析も可能である。

C.3 データの利便性

サンプルングデータは、利用者に求められるセキュリティ要件が特別抽出に比べて緩和されており、探索的な解析が出来る利点がある。サンプルングデータでは、解析実施場所の入退室者の制限は必須ではない。ユーザー認証を行ったスタンドアロンのパソコン、保管場所を確保出来

ば、サンプリングデータ解析を実施可能であり、特別抽出データに比べて利便性は高い。

C.4 NDB データ解析に必要な知識・スキル

1) 診療報酬制度の理解

レセプトデータの特徴を理解するため、診療報酬制度の仕組み、診療からレセプト発行までの流れ等に関する知識が必要である。また、NDB の DPC レセプトは特殊な構造をもっており、特定入院期間(包括期間)を超えた患者のレセプトデータを用いることを想定し、DPC 制度とレセプトデータの関連に関する理解が必須と考えられた。

2) データベースの理解

データベース、データベースマネジメントに関する基礎的な知識が必要である。NDB は、レセプトの記録条件仕様に定められた第三正規形のリレーショナル・データベースであることから、リレーショナル・データベースの特性や、データベース管理に関し、理解が必要である。この形式は、大規模レセプトのデータ量を小さくして格納するには適している。しかし、解析処理の効率の点では限界がある。NDB は、特別抽出、サンプリングデータのいずれも、カラム名を含まない csv 形式で提供される。データベースマネジメントシステムを用いたデータベース構築のスキル、レセプト単位、あるいは、患者単位でのデータの名寄、検索速度を上げるために必要となる前処理等を行う、基礎的なスキルを要する。

3) NDB レセプトデータの構造の理解

NDB レセプトのデータ構造として、テーブル構造、各テーブルに含まれるデータ項目、各テーブルを連結するデータ(患者 ID1、患者 ID2、レセプト通番)とその特性に関する知識が必要である。

4) レセプトデータの特徴の理解

レセプトの各項目から得られる情報の特性、限界に関する知識が必要である。本研究では、終末期高齢者を特定するために用いた、入院転帰

データの感度、特異度に関する情報を持って解析を行う必要がある(3)。また、入院レセプトに含まれる日付情報(入院日等)の記録ルールを理解する必要がある。

C.5 サンプリングデータ解析に必要なリソース

本研究では、DELL Precision3510 (CPU: Core i7 6820HQ 2.7GHz、メモリ 32GB、HDD 1TB)の端末を用いた。ソフトウェアには、データベースマネジメントシステム (PostgreSQL、A5M2)、事務処理系ソフトウェア (Office Professional、ATOM、秀丸エディタ)、統計解析ソフト (R、Easys-R、STATA) を用い、データ解析環境を構築可能であった。

D. 考察 & E. 結論

NDB サンプリングデータセットを用いることで、国レベルで終末期高齢者の診療実態を解明する横断研究が実施可能であった。特別抽出に比べてセキュリティ環境が緩和された環境で、探索的研究を含めた様々なリサーチクエストに基く解析が可能であり、データを活用する意義は大きいと考えられた。今後、サンプリングデータのセキュリティ要件の緩和等も検討されており(4)、利便性の向上も期待される。しかし、サンプリングデータの活用に必要な知識、スキルに関する情報は十分普及していない。今後、データ利用者による利用経験の共有や普及啓発活動により、セキュリティ環境や大規模データベースの構築環境が困難な研究者にも、活用を推進することが重要と考えられた。

表 1. サンプリングデータと特別抽出データの比較

	代表性	研究デザイン、解析	利便性
サンプリングデータ			
	<ul style="list-style-type: none"> 入院レセプトの 10%、外来レセプトの 1%を抽出 高額レセプトは削除されているため、医療費の解析対象にセクションバイアスが生じる 出現頻度が低い傷病名、診療行為は匿名化されており、該当する疾患群、診療行為、医薬品を受けた患者を対象にする研究で、セクションバイアスが生じる 	<p>【分析単位】</p> <ul style="list-style-type: none"> 医科入院、DPC レセプトでは、レセプト単位のみ可 入院エピソード単位の分析には限界が大きい 医科入院外では、患者単位、レセプト単位が可 <p>【分析対象期間】</p> <ul style="list-style-type: none"> 1ヶ月全期間(10月の診療実日数が 30 日)を評価可能であるが、期間が長いとサンプル数が少ない場合がある 1ヶ月以上過去の背景、前治療の把握に限界が大きい H24～27 年の経年比較が可 <p>【診療実態の把握】</p> <ul style="list-style-type: none"> 平成 24 年以降のレセプトに記録されている実施日データ(診療行為、医薬品、調剤)を用い、1ヶ月の期間内の診療内容を把握可能 <p>【医療供給側要因の把握】</p> <ul style="list-style-type: none"> 病棟区分別の解析が可能 	<ul style="list-style-type: none"> データ利用場所の入退室制限(入室者の記録・識別、入退の制限)は必須ではない 探索的な研究が許可されている 研究者が csv 形式で提供されるレセプトデータからデータベースを構築する知識、スキルが必須である
特別抽出データ			
	<ul style="list-style-type: none"> 全請求の約 95% (平成 27 年時点)、電子化されたレセプトのほぼ 100%をカバー 	<p>【分析単位】</p> <ul style="list-style-type: none"> 患者単位、レセプト単位、入院エピソード単位の分析が可能 <p>【分析対象期間】</p> <ul style="list-style-type: none"> 1ヶ月以上の期間を対象とした分析が可能 <p>【医療供給側要因の把握】</p> <ul style="list-style-type: none"> 医療機関 ID、病床規模等の解析が可能 	<ul style="list-style-type: none"> データ利用場所の入退室制限、窃視防止対策、操作ログの記録が必須 探索的な研究は許可されていない 研究対象者数によっては、サンプリングデータに比べて大規模なデータベース構築が必要

参考文献

- (1) 酒井未知. 医療経済研究機構 2016 年度 (第 20 回) 研究助成 研究要旨. 大規模レセプトデータベースを用いた高齢者終末期医療の実態解明. Monthly IHEP 2017-2018 年 12・1 月号
- (2) サンプリングデータセットについて. 第 1 回 要介護認定情報・介護レセプト等情報の提供に関する有識者会議. 資料 4. 平成 30 年 3 月 14 日.
<http://www.mhlw.go.jp/file/05-Shingikai-12301000-Roukenkyoku-Soumuka/0000212768.pdf>
- (3) Sakai M, Ohtera S, Iwao T, Neff Y, Kato G, Takahashi Y, Nakayama T; BiDAME (Big Data Analysis of Medical care for the Elderly in Kyoto). Validation of claim data to identify death among aged persons utilizing enrollment data from health insurance union. *Environ Health Prev Med.* 2019 Nov 23;24(1):63.
- (4) サンプリングデータセットの内容充実について. 第 26 回レセプト情報等の提供に関する有識者会議 資料 2. 平成 27 年 12 月 9 日.
<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000108548.pdf>

F. 健康危機情報
なし

G. 研究発表

1. 中山健夫. レギュラトリーサイエンス財団「安全性監視計画入門コース」第 9 講: 日本で利用可能な DB の応用事例と課題 - MID-NET、民間 DB はどのような課題に適しているか、バリデーションをどう行うか - 2019 年 7 月 19 日 長井記念ホール
2. 中山健夫. 医療ビッグデータコンソーシアム 提言策定部会 2019 年 5 月 22 日 日本橋ライ

フサイエンスビル

3. 中山健夫. Recon sidering "Real World Data". DIA Japan. 2019 年 11 月 12 日 東京ビッグサイト
4. 中山健夫. 医薬品開発及び製造販売後調査・研究における医療ビッグデータの実践的利用法を考える. 日本薬学会 レギュラトリーサイエンス部会 / 第 16 回医薬品レギュラトリーサイエンスフォーラム 2020 年 1 月 10 日 長井記念ホール
5. 大寺祥佑, 加藤源太, 森由希子, 黒田知宏. NDB オンサイトリサーチセンターの利用動向: 今後の一般開放に向けて レセプト情報等オンサイトリサーチセンター(京都)の現状と今後の展望. 第 39 回医療情報学連合大会論文集. p213
6. 植田彰彦, 近藤英治, 大寺祥佑, 朝野美穂, 中北麦, 万代昌紀, 加藤源太, 黒田知宏. NDB オンサイトリサーチセンターの利用動向: 今後の一般開放に向けて 初学者による京都大学 NDB オンサイトリサーチセンターの使用経験. 第 39 回医療情報学連合大会論文集. p215
7. M Sakai, S Ohtera, T Iwao, Y Neff, T Uchida, Y Takahashi, G Kato, T Kuroda, S Ni shimura, T Nakayama, on behalf of BiDAME (Big Data Analysis of Medical care for the Elderly in Kyoto) Decrease of life-spanning treatment provided just before death for older inpatient in Japan: A trend analysis from 2012 through 2014 based on national representative samples (submitted)

H. 知的所有権の取得状況
なし