

厚生労働行政推進調査事業費補助金
(政策科学総合研究事業 (政策科学推進研究事業))
臨床疫学に活用可能なNDB等データセットの作成に関する研究
分担研究報告書

NDBデータと介護DBデータ及び連結データの実分析およびデータセット作成にかかる技術支援に関する研究

研究分担者 野田 龍也 奈良県立医科大学公衆衛生学講座 准教授
研究分担者 今村 知明 奈良県立医科大学公衆衛生学講座 教授
研究協力者 明神 大也 奈良県立医科大学公衆衛生学講座 助教

研究要旨

本分担研究は2つの目的を置いている。第一の目的は、NDBと介護データベース（以下、介護DB）の連結可能なサンプルデータの作成を見据え、主にKDBを用いてデータベースの試行的な処理を行うことである。第二の目的は、NDB・介護DBのHIC試行的利用により、操作可能性を予備的に検証することである。本目的を検証するため、医療レセプトと介護レセプトの連結可能性と連結時の課題を整理する基礎として、某県のKDBデータを用い、サンプル抽出を行って課題を整理した。また、HICにNDBデータを格納し、利用可能になるまでの手順を予備的に検証し、課題を抽出した。

KDBを用いた医療と介護の結合は、小サンプルで試行した限りでは、一定の名寄せが可能であるとの予想が得られた。介護DBには給付実績情報連結キーが付与されるとの情報があり、それを用いることができれば、さらに精緻な結合が可能になると思われる。NDB・介護DBのHIC試行的利用による操作可能性の予備的検証では、マニュアルの不備などの課題が明らかとなった。

今後、データベースの連結を見据えた課題克服と利活用をより一層進める。

A. 研究目的

本分担研究は2つの目的を置いている。
第一の目的は、NDBと介護データベース（以下、介護DB）の連結可能なサンプルデータの作成を見据え、主にKDBを用いてデータベースの試行的な処理を行うことである。
第二の目的は、NDB・介護DBのHIC試行的利用により、操作可能性を予備的に検証することである。

B. 研究方法

第一目的に関連し、医療レセプトと介護レセプトの連結可能性と連結時の課題を整理する基礎として、某県のKDBデータを用い、サンプル抽出を行って課題を整理した。
第二目的に関連し、HICにNDBデータを格納し、利用可能になるまでの手順を予備的に検証し、課題を抽出した。

(倫理面への配慮)

本研究では完全に匿名化された個票を用い、個人情報や動物愛護に関わる調査・実験は行わない。研究の遂行に当たっては、各種法令や「人を対象とする医学系研究に関する倫理指針」を含めた各種倫理指針等の遵守に努める。また、厚生労働省保険局を始めとする関係各所の定めた規定・指針等を遵守し、必要な申請を行う。また、実施にあたっては、奈良県立医科大学医の倫理審査委員会の許可を得た。

C. 研究結果

C-1. NDBと介護DBの連結可能なサンプルデータの

作成を見据えた上での、主にKDBを用いたデータベースの試行的な処理

- ・介護DBの別添8に含まれる情報の特定
要介護認定情報 (D_NINTEI) 1シート
- 給付実績情報 (DT1111_H1、DT1111_D1~DF、DT1111_T1) 18シート
- 給付管理票情報 (DT1121) 1シート
- 台帳情報 7シート

このうち、KDBでは給付実績情報のみが入手可能であった。

後述するD4テーブルとDBテーブルは、テーブルが存在しつつもデータが入っていないことが確認された。残りの16シートのうち、データ数が最小であるD2テーブルから10 ID (KDB個人番号に近い) をランダムに抽出した。

その際、1つのIDは4か月分、1つのIDは2か月分あったので、合計14行で構成されることとなった。この10個のIDに一致するデータを、他のテーブルから抽出した。

以下にテーブル名と概要を示す。実施年月は、10 IDを抽出するもととなったD2テーブルにおける実施年月である。

H1:もともと1575万行、10IDで1735行、実施年月との一致で68行

D1:もともと4800万行、10IDで6236行、実施年月との一致で270行
D2:もともと600行、そこから10ID抽出、実施年月でも分けると14行になった
D3:もともと35万行、一致するIDなし
D4:もともと0
D5:もともと610万行、10IDで539行、実施年月を対象年月に変更すると25行
D6:もともと4.5万行、一致するIDは10行未満、実施年月との一致はなし
D7:もともと2.8万行、一致するIDは10行未満、実施年月との一致はなし
D8:もともと150万行、10IDで113行、実施年月との一致で10行未満
D9:もともと185万行、10IDで228行、実施年月との一致で32行
DA:もともと5万行、一致するIDなし
DB:もともと0
DC:もともと2.8万行、一致するIDは18行、実施年月との一致で10行未満
DD:もともと2.1万行、一致するIDなし
DE:もともと2500行、一致するIDなし
DF:もともと2000行、一致するIDなし
T1:もともと1177万行、10IDで1386行、実施年月との一致で54行

上記のうち、D3、DA、DD～DFには一致するIDが確認できなかった。実施年月も一致させると、抽出データが非常に限定されることが明らかとなった。

C-2. NDB・介護DBのHIC試行的利用による操作可能性の予備的検証

HIC上でデータを受領してからのタイムラインを示す：

2023年1月13日：
厚生労働省から当研究班への説明会

2023年2月1日：
関係大学実務者によるミーティング

2023年2月：
最終的にEC2環境まで入って取り込みを行い、S3からのデータコピーの手順に苦労しつつ、コピーを作成した。

D. 考察

D-1. NDBと介護DBの連結可能なサンプルデータの作成を見据えた上での、主にKDBを用いたデータベースの試行的な処理

某県のKDBを用いた医療と介護の結合は、小サンプルで試行した限りでは、一定の名寄せが可能で

あるとの予想が得られた。介護DBには給付実績情報連結キーが付与されるとの情報があり、それを用いることができれば、さらに精緻な結合が可能になると思われる。

D-2. NDB・介護DBのHIC試行的利用による操作可能性の予備的検証

2023年3月下旬の段階で、全部のデータベースの環境は作ることができていない状況にあるが、一連の流れで、以下の課題が明らかとなった。

そもそものマニュアルがなく、IDの証明書だけを参考に、手探りで進める必要があった。また、マニュアルはworkspacesの環境に置かれていたが、Excelが入っておらず、所与の環境で閲覧することができなかった。

Dドライブには16TB領域があったものの、データのcsvファイルが合計で12TBほどあり、そのままデータベースを作ろうと試みると容量不足となった。そのため、データを一つずつコピーしてデータベースに入れ、元のデータを削除する処理を繰り返す必要があった。データベースを展開する際に一時的に必要な大容量が確保されていないため、可能な範囲ごとに作業を行う点は課題と考えられた。

また、多人数が同時に（または同時期に）大きなサイズのファイルを利用する場合の容量不足が予想される点も課題として挙げられる。

上記の課題を回避するためには、たとえば、各施設の研究者がデータベースの展開を必要としないような、一定のレディメイドのデータセットを準備することなどが考えられる。

E. 結論

NDBと介護DBの連結可能なサンプルデータの作成を見据えた上での主にKDBを用いたデータベースの試行的な処理及びNDB・介護DBのHIC試行的利用による操作可能性の予備的検証において種々の課題が明らかとなった。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし