

令和4年12月7日

第13回匿名医療情報等の提供に関する専門委員会

資料3

# トライアルデータセット（仮称）の仕様について

厚生労働省 保険局医療介護連携政策課  
保険データ企画室

ひと、暮らし、みらいのために



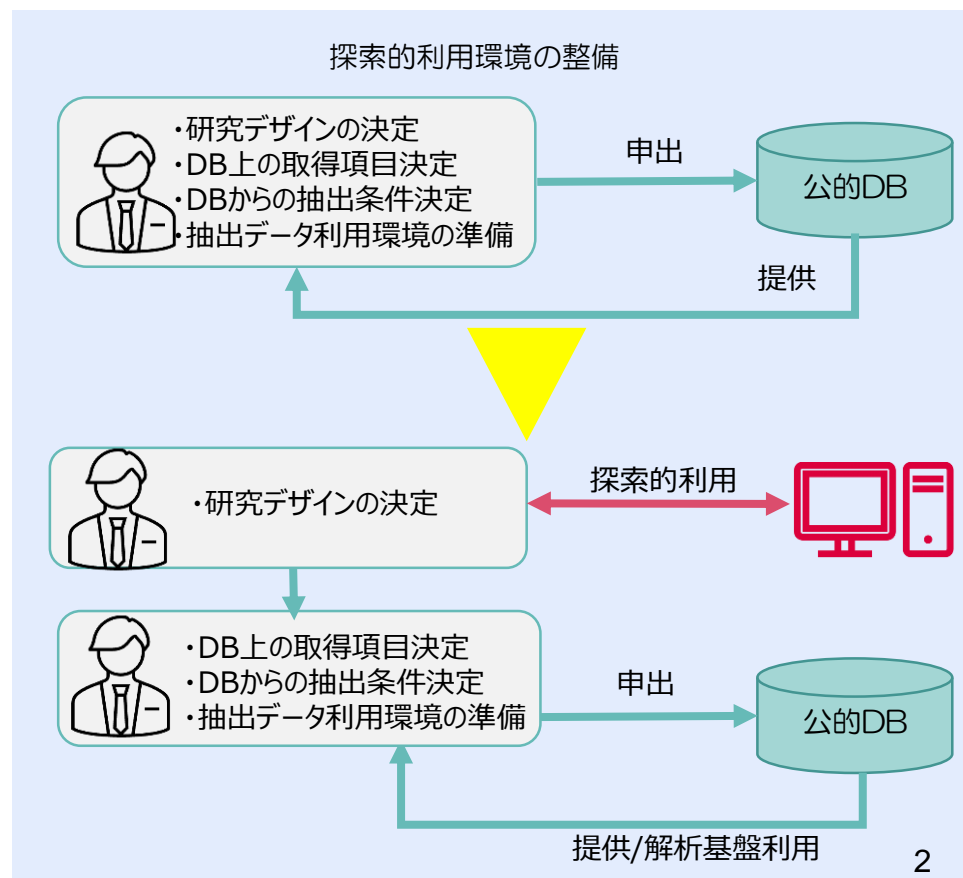
厚生労働省  
Ministry of Health, Labour and Welfare

# 探索的利用環境整備の目的・意義

- 現行の申出（申請）制度は、事前に利用するデータを絞り込む必要がある等、公的DBに関する知識を有することを前提としており新規参入や普及促進のハードルとなっている。そのため、多くの利用主体が臨床等のエビデンス研究に取り組めるよう、連結解析を容易とする解析環境において探索的な利用をできるようにする。
- データベース研究をしたことはあるが、NDB・介護DB等を利用したことがない研究者等がトライアルデータ（仮称）を用いて探索・試行的に分析するための環境を提供する。

## 探索的利用環境でトライアルデータ（仮称） 提供の意義

- ・ NDB利用申請前に、NDBに実際に収載されているデータに近いトライアルデータ（仮称）の探索的な利用を可能とすることで、研究者がNDBのデータ構造、研究テーマ関連の実態及び研究計画の実現可能性を把握し、リサーチクエスチョン、研究デザインを決定することを支援する。
  - ・ 研究者の希望に沿って、SQLなどの知的財産はダウンロードを可能とし、解析環境の準備や抽出条件の決定などについて円滑に進められるようにする。
- ↓
- ・ データ構造の理解、研究デザインの決定のため、NDBに実際に収載されているデータに近いデータを提供する必要がある。
  - ・ サンプリングデータセットを元に、さらに個人特定のリスクを軽減したデータセットを提供する。



# サンプリングデータセット

## サンプリングデータセット：1か月分の抽出データ

対象レセプト	<ul style="list-style-type: none"> <li>➤ 各年・「医科入院レセプト」、「医科入院外レセプト」、「DPCレセプト」、「調剤レセプト」のレセプト毎に用意しております。</li> <li>➤ ○2011年：1月、4月、7月、10月 ○2012年：1月、4月、7月、10月 ○2013年：1月、4月、7月、10月 ～ 2014- 記載省略 ～</li> <li>➤ 1月、4月、7月、10月の対象レセプトは以下の通り（調剤レセプトのみ翌月分の対象に含まれます） <ul style="list-style-type: none"> <li>1月：当該年1月診療分、かつ2月、3月、4月に審査されたレセプト</li> <li>4月：当該年4月診療分、かつ5月、6月、7月に審査されたレセプト</li> <li>7月：当該年7月診療分、かつ8月、9月、10月に審査されたレセプト</li> <li>10月：当該年10月診療分、かつ11月、12月、1月に審査されたレセプト</li> </ul> </li> <li>➤ 調剤レセプトに限り、①調剤単体分(PHA)／②医科入院外に紐付く調剤分(PHA外)の2種を提供しております。 (※②は、医科入院外を抽出後、ハッシュ値1で紐付けた調剤レセプトを抽出)</li> </ul>
抽出方法	<ul style="list-style-type: none"> <li>➤ 「医科入院」「DPC」は、レセプト全体の10%を抽出しています。</li> <li>➤ 「医科入院外」「調剤」は、レセプト全体の1%を抽出しています。</li> <li>➤ 性別、および5歳刻みの年齢層において、母集団と構成比率を変えないように抽出されています。</li> </ul>
空欄化 もしくは削除 した主な情報	<ul style="list-style-type: none"> <li>➤ 「医科入院」においては700,000点以上、「医科入院外」「調剤」においては50,000点以上のレセプトは削除しています。</li> <li>➤ 「都道府県」に関する情報、「保険者」に関する情報は削除しています。</li> <li>➤ 「DPC」レセプトにおいては、医療機関別係数は空欄化、その係数の影響を受ける情報は削除しています。</li> </ul>
匿名化処理	<ul style="list-style-type: none"> <li>➤ 「傷病名コード」「医科診療行為コード」「医薬品コード」「診断群分類」において出現頻度の低いものは匿名化処理を行っています。</li> </ul>

# 探索的利用環境で提供するトライアルデータセット（仮称）の作成

医療データ利活用人材育成に資するデータ利用法の開発

京都大学医学部附属病院 医療情報企画部 黒田 知宏教授

## 研究概要（申請書抜粋）

- 文部科学省「医療データ人材育成拠点形成事業」において、医療データの利活用推進に資する人材育成プログラムを実施しており、本研究はその一環としてNDBデータ分析の初学者が分析手法を習得するために必要な教育プログラムの開発を行う。具体的にはサンプリングデータを用いて、NDBデータを分析する際のデータハンドリングの定型化について検討する。また、開発された手法について「医療データ人材育成拠点形成事業」に参加する大学院生を対象に有用性についての検証を行う。
- この検証にあたり、利用者申請をしていない学生が実習可能なデータセットが必要となるため、サンプリングデータセットを加工し、「レセプト情報・特定健診等情報の提供に関するガイドライン」の公表基準に準拠したデータセットを作成する。



当該研究で作成したデータセットと同じ仕様で、探索的利用環境で提供するトライアルデータセット（仮称）を作成してはどうか