

令和2年度 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))

医療・介護のデータの利活用の推進のための、
NDB・介護DBの連結可能性および活用可能性の評価に関する研究
分担研究報告書

NDB・介護DB連結データマートの構築および試行集計

研究代表者 加藤 源太 (京都大学医学部附属病院 准教授)
研究分担者 黒田 知宏 (京都大学医学部附属病院 教授)
研究分担者 大寺 祥佑 (京都大学医学部附属病院 助教)
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 康永 秀生 (東京大学大学院医学系研究科 教授)
研究分担者 田宮 菜奈子
(筑波大学医学医療系/ヘルスサービス開発研究センター 教授)
研究分担者 杉山 雄大 (国立国際医療研究センター 医療政策研究室長)
研究分担者 中山 健夫 (京都大学大学院医学研究科 教授)

研究要旨

【研究目的】

本研究は、匿名レセプト情報等データベース (NDB)、および匿名要介護認定情報・匿名介護レセプト等情報 (介護DB) の両者のデータ連結について、大きく分けて、①両データの連結可能性の評価、②連結データの利用可能性の評価を目的として行われるものである。①連結可能性の評価として、介護DBの概要と構成要素について整理し、これまでの医療・介護データの連結解析の議論や、両データを連結して行った先行研究を参考に、連結解析が容易になるような新しいテーブルを作成した。②の利用可能性の評価として、作成したデータマートを用いて先行研究を参考に簡易な連結解析を行い、その作業過程を評価した。

【研究方法】

上述の研究目的に基づき、2014年度、2015年度分のNDBデータおよび介護DBデータを利用して、実施した。

【研究結果】

(連結可能性の評価)

介護 DB は匿名要介護認定情報と匿名介護レセプト等情報で構成されているが、要介護認定情報は利用者の病態や ADL に関連する情報を多く含んでおり、行った医療・介護行為の請求情報が中心である医療および介護のレセプト情報と連結することで、データの価値を向上させることが期待できる。しかし、レセプト情報が基本的にサービス利用時に毎月発生するのに対し、要介護認定情報は調査時にしか登録されないため、対応するレセプトと連結して分析することは容易でない。そこで、要介護認定情報が各月ごとに確認でき、調査内容も参照できるよう、新たなテーブルを作成するとともに、そこに含める情報は「ID」「性、年齢階級」「年月ごとの要介護度」「要介護度の認定有効期間（開始年月日、終了年月日）」に限定した。また、要介護認定情報のその他の情報を簡易に参照できるよう、要介護認定情報を一意に指定できる「認定通番」を本研究にて独自に導入し、介護レセプト情報にも加えて、両者の連結解析が容易に実行できるようにした。

（利用可能性の評価）

作成した新たなデータマートを用いて、連結解析を行っている 2 つの先行研究に着目し、3 つの集計の再現を試みた。1 つ目は、匿名要介護支援情報の要介護度情報を使い、要介護度別にみた傷病分類の集計を NDB の傷病名レコードを使用し、簡素化した形式で実施した。2 つ目はその逆で、傷病分類ごとに要介護度の分布を明らかにする集計を実施した。もう 1 つの研究は、特別養護老人ホームおよび介護老人保健施設の入所者における、避けられたかもしれない入院（Potentially Avoidable Hospitalizations, PAH）をアウトカムとし、PAH とそれ以外の入院、病院での死亡との関連要因についての施行集計を行った。NDB の RE と BU の情報から入院年月日の同定を行い、匿名要介護認定情報からは、経管栄養の有無の情報を入手し、匿名介護レセプト情報からは施設入所の有無を給付実績情報（明細情報レコード）DT1111_D1 のサービス種類コードから、入所年月日を給付実績情報（基本情報レコード）DT1111_H1 の入所（院）年月日情報から入手して行った。いずれの集計も、NDB を加工する作業に数十分程度を要したが、介護 DB データと突合する作業に要したのはそれぞれ数分程度であった。介護 DB データとの突合作業が短時間で済んだのは、介護情報を突合させる段階で対象人数が大幅に減少することもその要因となっていると考えられた。

A. 研究目的

本研究は、匿名レセプト情報等データベース（NDB）、および匿名要介護認定

情報・匿名介護レセプト等情報（介護 DB）の両者のデータ連結について技術的検証を行うとともに、連結データが保健医療分野における政策や研究のエビデ

ンスを導き出せる質を備えることができるのか、その場合はどういった課題やテーマに対して強みを発揮できるのかを評価するものである。

当研究では、以下の2点を評価すべき課題として、研究事業を進めてきた。

- ① 連結可能性の評価：NDB と介護DB とを連結させることが技術的に可能か否か、どの程度の作業負荷が発生するのか
- ② 利用可能性の評価：NDB と介護DB とが連結された場合、そのデータは、具体的にどういった分析が新たに可能となり、どういった政策の立案に寄与できるのか

この2つの評価課題に対し、まず①連結可能性の評価として、介護DBの概要と構成要素について整理を行った。そのうえで、これまでの医療・介護データの連結解析の議論や、両データを連結して行った先行研究を参考に、NDBデータと介護DBデータの連結解析を容易にするデータ構成を検討し、連結解析に使用できるような、新しいテーブルの作成を行った。

次に、②の利用可能性の評価としては、作成したデータマートを用いて先行研究を踏まえつつ、それをトレースする形式で非常に簡単なレベルの試行集計を行った。その際に、どの程度の作業時間を要したかについても、確認を行った。

B. 研究方法

今回の研究では、上述の研究目的に基づき、2014年度、2015年度分のNDBデータおよび介護DBデータを利用して実施した。両データを解析する基盤については、本研究報告の末尾にて紹介する通りである。

倫理面への配慮

NDBデータ、介護DBデータいずれに対しても提供依頼申出を行い、使用に際しては「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」及び「要介護認定情報・介護レセプト等情報の提供に関するガイドライン」を遵守した。またデータ提供を受けるに際しては所属機関における倫理審査での承認が必要であるため、京都大学医の倫理委員会にて申請を行い、承認を得た（R1119-2）。

C. 研究結果

【連結可能性の評価】

本研究報告に先駆けて、まず、介護DBの概要と構成を紹介する。

介護DBの概要と構成

介護DBの保有情報は、匿名要介護認定情報と匿名介護レセプト等情報がある。それぞれ収集経路が異なっている。匿名要介護認定情報は実質的な保険者である市区町村から、匿名介護レセプト等情報については介護サービスを提供する

事業所から国保連合会を経て格納されている¹。国保連合会は、介護レセプト情報に対し審査・支払を行う機関としても機能している。

介護保険が利用されるに先立ち、まず市区町村が保険者として存在しており、利用者ごとの認定調査を経て要介護度の評価が行われ、この要介護度に基づいて介護サービスの枠が定められている。そして、その枠の範囲で利用者が介護サービスを利用すれば、サービス事業者が国保連合会に介護報酬の請求を行う流れとなっている。このほかに市区町村と国保連合会が委託関係にあり、国保連合会が審査支払業務を担っていることから、審査支払業務に必要な台帳情報を市区町村と共有する必要があり、この台帳情報が国保連合会に保管されている。

介護 DB を構成する情報として、まず、各月の請求のたびに格納される情報として、介護給付実績情報、いわゆる介護レセプト情報がある。介護給付費明細書の情報が格納されており、介護サービス事業者ごと、利用者ごとに毎月の情報が格納されている。ある意味で、NDB のレセプト情報と扱いはほぼ同様と考えてよい。また、匿名介護レセプト情報の各レコードには個人 ID が含まれているので、ここに NDB と同一の ID が付与されていれば、特段の加工をせずとも連結解析は可能である。格納されているテーブルは多数あるが、基本的に明細情報レコード (DT1111_D1) が明細情報に相当するので、この情報を用いれば多くの事例において対応が可能ではないかと

考えられた。更に詳しい情報が必要であれば、他のテーブルの情報も必要になるが、その都度 ID で連結させることが可能である。

一方、要介護認定情報は、登録のたびに格納される情報、つまり毎月に入っていない情報である点で、介護レセプト情報等は仕様が異なっている。テーブル内の ID の格納場所も介護レセプトとは異なっている。介護 DB におけるテーブル名は「D_NINTEI」である。台帳情報は、受給者だけではなく、サービスを行う事業所や保険者などの基本情報が含まれている。保険者は、台帳情報に変化があれば、そのつど国保連合会に異動情報の提出を行っており、その都度データが登録されることになっている。介護 DB におけるテーブルとして利用される可能性が高い情報は受給者台帳情報

(DT5341) であり、この情報以外には事業者情報などは含まれているものの、個人 ID は格納されていない。

介護 DB 情報の活用の態様

日本における介護 DB に相当するデータを用いた研究を我々でいくつか確認したところ、大きく分けて以下 2 つの使用の態様が認められた。

- ① 得られた情報を患者のアウトカムとして使用する
- ② 得られた情報を対象者の状態像を示す情報として使用する

このうち、①は要介護度の変化（患

化)、または悪化までの期間をアウトカムとして評価している事例を多く認めた。また、給付費をアウトカムとして使用する研究も認められた。他の例として、特定の介護サービスを利用しているかどうかや、そのサービスの提供量に着目した研究がみられた。また、自宅におられる方が介護保険施設に入所したかどうか、または入所までにどのくらいの期間を要したかを評価したものや、施設の方が自宅へと退所されたか否かを評価した事例もみられた。その他として、死亡時の場所を特定するための情報として介護レセプトを使用する事例もみられた。

一方、②の例としては、認定調査における要介護度や、認定調査情報に含まれている詳細項目(例:立ち上がれるか、自分で食事がとれるか、など)を使用したものや、障害高齢者や認知症高齢者の日常生活自立度を評価するための指標に使用したり、その他利用したサービスから病態を推測したりする(認知症に対する加算が付いている場合に、認知症があると類推する、など)事例や、自治体レセプトを使用している研究で介護保険料区分に関する情報を用いることで経済状況を把握しようとするような研究もみられた。その他、要介護度の悪化やその期間、詳細項目、日常生活自立度といった情報も使用されていた。

これらの情報は、介護レセプト情報だけでなく、要介護認定情報や台帳情報に含まれる受給者台帳情報に記載されていることが少なくない。レセプトは基本的には行った行為に対する請求情報であ

り、対象者の状態を示す情報はそれほど多く含まれていない。この傾向はNDBのレセプト情報についても言えることであり、NDBのレセプトデータにも要介護認定情報や台帳情報を円滑に連結できれば、患者の病態やADLといった情報を付加した利用ができるようになり、データの価値が高まると考えられる。

要介護認定情報と介護レセプト等情報の違い

要介護認定情報に含まれる要介護度については、要介護認定情報だけでなく介護レセプト情報、受給者台帳情報からも算定状況などを通じて把握することが可能である。ただし、利用の面ではそれぞれ一長一短がある。たとえば、介護レセプト情報から得られる要介護度情報はレセプトの請求に紐づいているという利便性があるが、請求に基づいた情報となっているため、介護サービスを利用していない月は請求が発生せず、認定を受けていてもその事実を把握することができない。

一方、要介護認定情報に格納されている要介護度情報は、サービスの利用の有無にかかわらず記録されており、細かい項目の情報も含まれてはいるものの、認定調査が行われる際にのみ登録されるため、年月単位の情報になっていない。台帳情報に格納されている要介護度情報も、要介護認定情報の要介護度と同様、サービス利用の有無とは無関係に記録されている。基本的には認定調査が行われるたびごとに国保連合会に保険者に提出

する「異動連絡票情報」に記録される。この情報は原則的には要介護認定情報の要介護度と連動すると考えられるが、今回入手した介護 DB データを確認したところ、要介護度情報が台帳情報には存在するものの、要介護認定情報には存在していない事例がわずかに存在していた。その理由は確定できておらず、今後も精査していく必要があると考えられる。逆に、要介護認定情報に要介護度情報が入っていれば、台帳には必ず情報が含まれていた。

新しいテーブル「NI (仮)」の導入

以上のように、要介護認定情報にはその時の利用者の病態や ADL に関する情報が多く含まれており、利用者の病態を評価する情報として価値を有するものであるが、レセプト情報とは異なり、要介護認定情報は認定時にしか情報が格納されない。このため、認定調査ごと（最長 36 か月、継続事例であれば最長 48 か月）で格納されるデータを、毎月の匿名介護レセプト情報や NDB レセプト情報に連結させて分析しようとする、その都度、前後数か月、最長で合計 48 か月分の要介護認定情報を確認しなければならない。そこで、要介護認定情報を各月ごとに参照できるテーブルがあれば、介護レセプト情報のみならず NDB のレセプト情報とも連結した分析が行いやすくなると考えられた。

こうした方針に沿って、今回、要介護認定情報が各月ごとに確認でき、調査内容も参照できるよう、新たなテーブル

「NI (仮)」を作成した。その構成は以下のとおりである。

- ID
- 性、年齢階級
- 年月ごとの要介護度
- 要介護度の認定有効期間
(開始年月日、終了年月日)
- 認定通番
(本研究にて独自に導入)

NI テーブルに多くの情報を入れると使いづらくなるため、テーブルに含める情報は上記に限定しつつ、認定調査 1 回に対して一意の ID となる「認定通番」を作成し、テーブルに加えた。これにより、該当する月の認定調査の詳細情報を参照できるようした。認定通番は介護レセプトにも付与しており、認定通番を用いてその時に有効な要介護認定情報の中の認定情報を円滑に参照できるようにしている。

要介護認定情報を各月情報として格納するための手順として、まず認定開始年月日を 1 回の認定とみなして作成した。要介護認定情報が複数行入る場合が稀にみられたので、その場合は原則として認定期間、有効期間の開始日が最新のものを選択し、その情報を要介護認定情報として NI に格納した。認定通番は、保険者番号と被保険者番号と認定有効期間を連結させて作成しており、これで一意性を確保している。

【利用可能性の評価】

今回の研究では、医療レセプト、要介護認定情報、介護レセプト情報間で連結解析を行っている2つの先行研究に着目し、その再現を試みた。集計自体は、合計3通り実施した。以下にその結果を報告する。

事例 1a：要介護度別にみた疾患の集計

厚生労働省主催の第1回医療・介護データ等の解析基盤に関する有識者会議（2018年5月16日開催）において、有識者会議構成員でもある松田構成員より、医療介護総合データベースを活用した分析事例の紹介が行われた。ここで紹介されている事例のひとつに、介護保険を使用している患者ごとにどういった病態を有していることが多いのかを把握するため、医療と介護のレセプトを連結し、要介護度別に疾病の出現傾向を評価した結果が示されている。たとえば、女性の医科外来レセプトに着目したところ、要支援1であれば筋骨格系疾患や循環器疾患を有する事例が多くなっており、要介護度が上がるにつれて脳卒中や神経変性疾患など、ADLの悪化を伴うことの多い神経系疾患などが増加していることを明らかにしている²。これは医療レセプト情報に要介護認定情報を連結させて集計を行ったものである。この調査を、構築したデータベースを用い、簡易化させたいうで試行的に再現を行った。

まず、NDBのREレコードで性別情報および患者を一意に定めることを行い、次に、SYレコードを使ってカテゴ

リー化した傷病名を特定する作業を行った。本来であればSYレコードの傷病名コードをICD10などで臓器別に区分化して行うのが望ましかったが、今回は介護DBにおけるIDがランダムに付与されており、結果を解釈する余地がなかった。このため、集計に際してのパフォーマンスを確認することを優先課題とし、主傷病、つまりSYレコードにある主傷病フラグが1である傷病を用い、傷病名コードの下1けた（0～9）に基づいて10種類の傷病カテゴリーに分類する、原始的な試行的な傷病分類を作成した。また、介護DBからは新規作成したNIを用いて、この期間のなかで最新の年月における要介護度を取得し、分析に使用した。集計結果は、末尾に示す図の通りとなった。

事例 1b：傷病分類別の要介護度集計

次に、事例1aの分析とは逆の発想で、傷病分類ごとに要介護度の分布を明らかにすることにも意義があると考え、連結集計を行った。今度は、NDBのSYレコードに高血圧、虚血性心疾患、脳卒中の傷病コードが含まれるレセプトを抽出し、それらの事例の要介護度を、最新の年月におけるNIテーブルの要介護度情報とかけあわせて試行集計を行った。結果は、末尾の図の通りである。事例1a、事例1bの集計いずれも、介護DBのIDはランダムな番号とし、集計結果の解釈ができないという前提において介護DBの全データの入手を了承されたという経緯があったことから、特段の傾向

などを見出すことは出来なかったが、集計結果を得ることには成功した。

事例 2：認定調査情報を用いた集計

もう 1 つの事例として、先行研究 ” Potentially avoidable hospitalizations, non-potentially avoidable hospitalizations and in-hospital deaths among residents of long-term care facilities.”³ をもとに、試行集計を実施した。この研究は、特別養護老人ホームおよび介護老人保健施設の入所者における、避けられたかもしれない入院 (Potentially Avoidable Hospitalizations, PAH) をアウトカムとし、PAH とそれ以外の入院、病院での死亡との関連要因を、ある単一市区町村の医療介護レセプト連結データから明らかにしたものである。この研究のなかで、アウトカムとして用いられている変数のひとつとして、匿名要介護認定情報の調査項目に含まれている経管栄養の有無に関する情報を利用していることに着目し、経管栄養の有りの無しのフラグを作成して、試行集計を行った。

入院開始日について、NDB の RE と BU の情報から入院年月日の同定を行った。正確な入院年月日の特定は医科レセプトでは同一疾病で再入院した場合も当初入院日が記載されるなどするため、正確な入院年月日を反映できているとは限らないが、今回はこの情報を入院年月日として設定することで集計を進めた。匿名要介護認定情報からは、経管栄養の有無の情報を入手し、匿名介護レセプト情報からは、特別養護老人ホーム (特養)

入所者を給付実績情報 (明細情報レコード) DT1111_D1 のサービス種類コード (申出依頼テンプレート項番 9) が 51 (介護福祉施設サービス) の情報から、介護老人保健施設 (老健) 入所者を同じくサービス種類コードが 52 (介護保健施設サービス) の情報から特定するとともに、入所年月日を給付実績情報 (基本情報レコード) DT1111_H1 の入所 (院) 年月日 (申出依頼テンプレート項番 35) の情報から入手した^{4,5}。集計は、入所年月日より入院年月日が後であったか否かで評価し、集計した。本来であれば入所後入院までの期間で場合分けすることが望ましいが、今回は入院年月日が後か否かでのみ判断することとした。経管栄養を実施していた者を NI から認定通番を活用して抽出し、人数と割合を集計してクロス表を作成した。

結果は、末尾の表に示した通りである。ID がランダム付与されているので、得られた結果に対する解釈を行うことは行っていない。

これらの試行集計に要した作業時間

集計 1 の実施に際しては、既存の Hadoop 基盤に格納されている NDB から RE レコード、SY レコードを使用して傷病毎のファイル (仮の傷病分類: 傷病コードの下 1 桁による分類、および傷病グループ (高血圧、虚血性心疾患、脳卒中、その他)) を作成するのに要したのが、約 23 分であった。当研究で使用している分散処理基盤で得られた結果であり、データ基盤によってはさらに時間

を要することが想定された。しかし、一度傷病毎の分類を作成すれば、その後当研究で作成したテーブル（NI）と突合して年月ごと、患者ごとの処理を行う作業、IDで集計し傷病グループごとに集計表を作成する作業に要したのはそれぞれ数分程度であった。作業時間の減少については、介護情報を突合せさせる段階で対象人数が大幅に減少することもその要因となっていると考えられた。

集計2に際しては、NDBから対象者の入院年月日を抽出しファイル作成する作業に約10分を要した以外は、何れも1～3分程度で処理を終えることができた。匿名介護レセプト情報から施設入所状況を確認し、NIの認定情報から更に要介護認定情報の詳細にアクセスして経管栄養に関する情報を取得し、最終的には入院の有無、経管栄養の有無などのフラグを付与してNDB介護情報としてファイル出力を行っている。

今回の2つの試行集計において、介護DBのデータはカラム数が多く、全てのデータが必要とは思われなかったため、インデックスとしてNIのテーブルを用意して、そこから各研究において必要なデータを取り出す、という手法を採用した。元々のHadoop基盤の性能によるところも大きいですが、この手法をとったことで、連結データにおいても作業速度を遅延させることなく処理できたのではないかと考えている。

D. 考察

匿名要介護認定情報と匿名介護レセプト等情報の性質の相違に基づく、両者の連結の難しさ

本研究では、匿名要介護認定情報を各月の匿名介護レセプト等情報に新たなテーブルとして付与する手法を採用したが、この方法には一つの限界がある。それは、「利用するデータの対象期間」の、特に開始時期を、匿名要介護認定情報と介護レセプト等情報で合わせることに難しいことである。たとえば、2014年4月から2015年3月までの匿名要介護認定情報、匿名介護レセプト等情報の提供を申し出たとする。この場合、匿名介護レセプト等情報については、サービスの提供を受けている限り基本的に毎月情報が格納されるので、2014年4月分から2015年3月分までの情報の提供が受けられる。一方、匿名要介護認定情報については、認定調査の機会にしかデータが格納されない。このため、2014年4月に認定調査が行われている事例であれば問題ないが、例えば2013年9月に認定調査を受け、その後状態の変化等に伴って2014年7月に再度認定調査を受けた事例の場合、2014年4月、5月、6月の匿名介護レセプト情報には、参照できる匿名要介護認定情報が存在しないこととなる。この課題を解決するには、

- ① 匿名介護レセプト情報の入手初月分以前の匿名要介護認定情報を、参照用に入手する
- ② 匿名介護レセプト情報の月に対応する匿名要介護認定情報がな

い場合は、匿名介護レセプト等情報から確認できる要介護度情報を活用する

の2通りの解決策が考えられる。但し、①の場合には認定調査の期間が更新認定の場合最長で36カ月（状態区分が同じ場合は最長で48カ月）となるため、匿名要介護認定情報を匿名介護レセプト等情報よりも3年分余分に入手し、処理を行わなければならない。一方、②の場合は、要介護度情報等はある程度の信頼性を期待できるものの、認定調査に関する項目の詳細情報を参照できないため、これら項目を変数として研究利用することが不可能となる。①、②いずれの対応であっても課題を残すこととなる。実際の匿名介護情報等の第三者提供においてはこの問題はまだ顕在化していないが、今後、利活用が進むにつれて、登録型の匿名要介護認定情報と毎月の業務から発生する匿名介護レセプト等情報という、データ収集の仕組みが異なる2つのデータを以下に連結させて利用するか、という問題は顕在化していくのではないかと考えられた。

因みに、匿名要介護認定情報における調査項目の情報は認定調査時の調査員が入力した情報に、自治体が要介護・要支援など介護度の認定審査を行う際には、主治医意見書の記載内容と照らし合わせて、相違がある際には主治医意見書の情報に照らして情報を格納しているが、匿名要介護認定情報と主治医意見書情報とのバリデーションも、今後の課題の一つ

であると考えられた。

「受給者台帳情報」の「受給資格喪失」情報を活用した死亡の類推

匿名介護レセプト等情報のひとつである受給者台帳情報（テーブルID：DT5341）のなかに、受給者台帳の「異動年月日」「異動区分コード」（1: 新規、2: 変更、3: 終了）および「異動事由」（01: 受給資格取得、02: 受給資格喪失、03: 市町村異動、04: 合併による新規、99: その他異動）の項目が含まれている⁴⁶。このうち、「異動区分コード」で「受給資格喪失」として扱われている事例では、死亡年月日として代用できる可能性があり、先行研究でもそのような活用がなされている事例がある。このことを踏まえて、今回の新しいテーブルNIにおいて、この情報を付与することを行った。この情報単独で、あるいは他の介護レセプト等情報やNDBと連結させることで、死亡日を特定することがさらに容易になることが期待される。

一方で、施設系のサービスであれば死亡したか否かを施設側が把握できるためほぼ確実に異動事由を入力できていると思われるが、通所系サービスなどでは通所の途絶と死亡との関連がわからない場合があるため、この情報から必ずしも死亡の事実を類推できるとは限らない。あくまでも他の情報と連結させることで活用可能性が高まる、という位置付けとするのが良いと判断した。この点、NDBでも死亡個票との連結が議論されているところであり、そちらでの議論のさらな

る進展が待たれるところである。NDBと介護DBの連結の本意が「データの価値の向上」であるならば、精度の高い死亡情報の付加も重要な課題であると思われる。

医療費・介護費に関する研究

今回の試行集計では費用負担に関する項目は入っておらず、医療費と介護費の集計は実施していないが、今後連結分析が定着すれば、医療費と介護費の合計金額に焦点を当てた研究も進むと思われる。その都度、研究者ごとに各月のNDBのHOレコードや介護DBの単位数に当たるのではなく、今回のNIテーブルに新たに合計金額の項目を付加するなどすれば、利用者の利便性が高まることが想定された。今回の研究では実現できなかったが、今後の課題といえる。

ランダムなIDで連結させたことに起因する、得られた結果に対する解釈の限界

今回実施した試行集計において、幸い、新たに得た匿名介護情報等に付与したIDをNDBと連結させてデータ処理を行う作業は実際に「実施可能」であることは確認でき、またHadoop基盤という特殊環境のもとではあるものの、どの程度の時間を分析に要したかも、評価することはできた。一方で、既存の先行研究と照らして得られた結果を解釈したりするなど、データの精度評価・バリデーションを行うには至らなかった。今回の研究を行うにあたり、あくまでもNDBと介護DBの連結解析の可能性を評価す

ることが目的であったため、その前提でランダムなIDで集計を行うことと引き換えに、介護DBの網羅的な提供を得ることができた。NDBと介護DBの連結解析が作業として実施できたことを踏まえて、今後は、得られた結果を適切に解釈できるノウハウを充実させていくことが求められる。NDBも介護DBも、そもそも、単独のデータであっても利用経験者が必ずしも多くない中で、医療レセプトと介護レセプト・要介護認定情報などを連結させて分析した経験を有する者は、極めて限られている。今後は、NDBと介護DBの連結解析の可能性を研究者間で積極的に論じ、実際にデータを利用する過程で幅広く共有していくことが求められる。

E. 結論

匿名レセプト情報等データベース(NDB)、および匿名介護情報等データベース(介護DB)の両者のデータ連結について技術的検証を行うことを目的に、2014年度、2015年度のNDBデータと介護DBデータを、連結キーを用いてランダムに連結し、そこからデータマートを作成する作業を行った。2020年度は匿名要介護認定情報のいくつかの情報を含んだ新しいテーブルを設け、これをNDB、介護DBのレセプト情報と連結解析させる基盤を構築し、幾つかの試行集計を実施した。これらの研究を通じて、NDBと介護DBの全数レベルでの連結解析が実施可能であることを確認す

ることができた。

参考文献

1. 厚生労働省、要介護認定情報・介護レセプト等情報の提供に関する事前説明会資料1「要介護認定情報・介護レセプト等情報について」
(https://www.mhlw.go.jp/stf/shingi/2/0000198094_00004.html).
2. カ：厚生労働省、第1回医療・介護データ等の解析基盤に関する有識者会議「医療介護総合データベースを活用した分析事例の紹介」
(<https://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000206674.pdf>)
3. Boyoung Jeon, Nanako Tamiya, Satoru Yoshie, Katsuya Iijima, Tatsuro Ishizaki, Potentially avoidable hospitalizations, non - potentially avoidable hospitalizations and in - hospital deaths among residents of long - term care facilities. *Geriatrics Gerontology International*, 2018 Aug; 18(8): 1272-1279. doi:<https://onlinelibrary.wiley.com/doi/full/10.1111/ggi.13458>
4. 厚生労働省、匿名介護情報等の提供について、「別添8 申出依頼テンプレート（抽出）」
(https://www.mhlw.go.jp/stf/shingi/2/0000198094_00033.html)
5. 独立行政法人福祉医療機構、介護保険事務処理システム変更に係る参考資料
(https://www.wam.go.jp/gyoseiShiryou-files/documents/2021/0331154520695/20210331_023.pdf).
6. 厚生労働省、第2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
(<https://www.mhlw.go.jp/content/12301000/000331608.pdf>)

F. 健康危険情報

なし

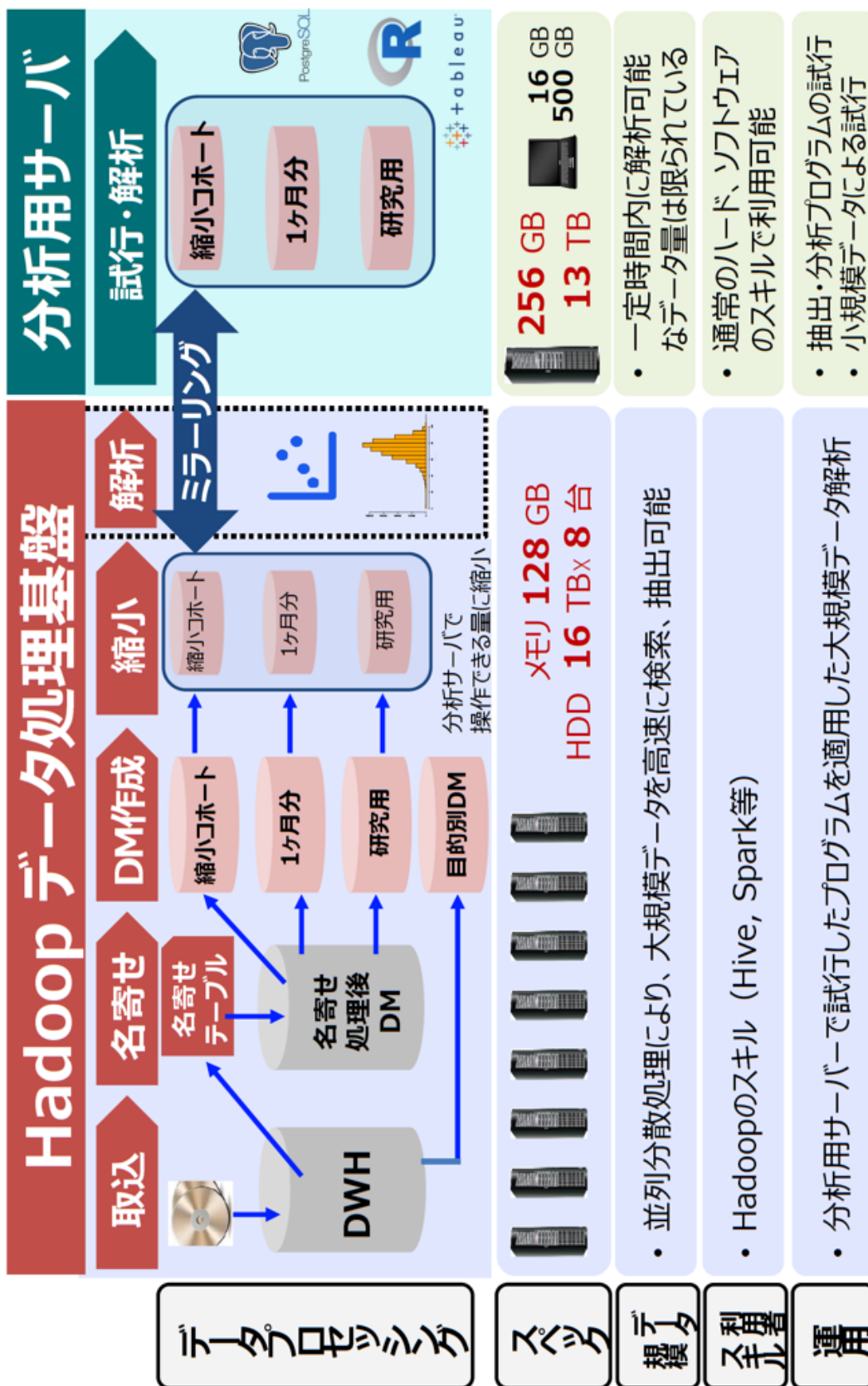
G. 研究発表

なし

H. 知的財産権の出願・取得状況

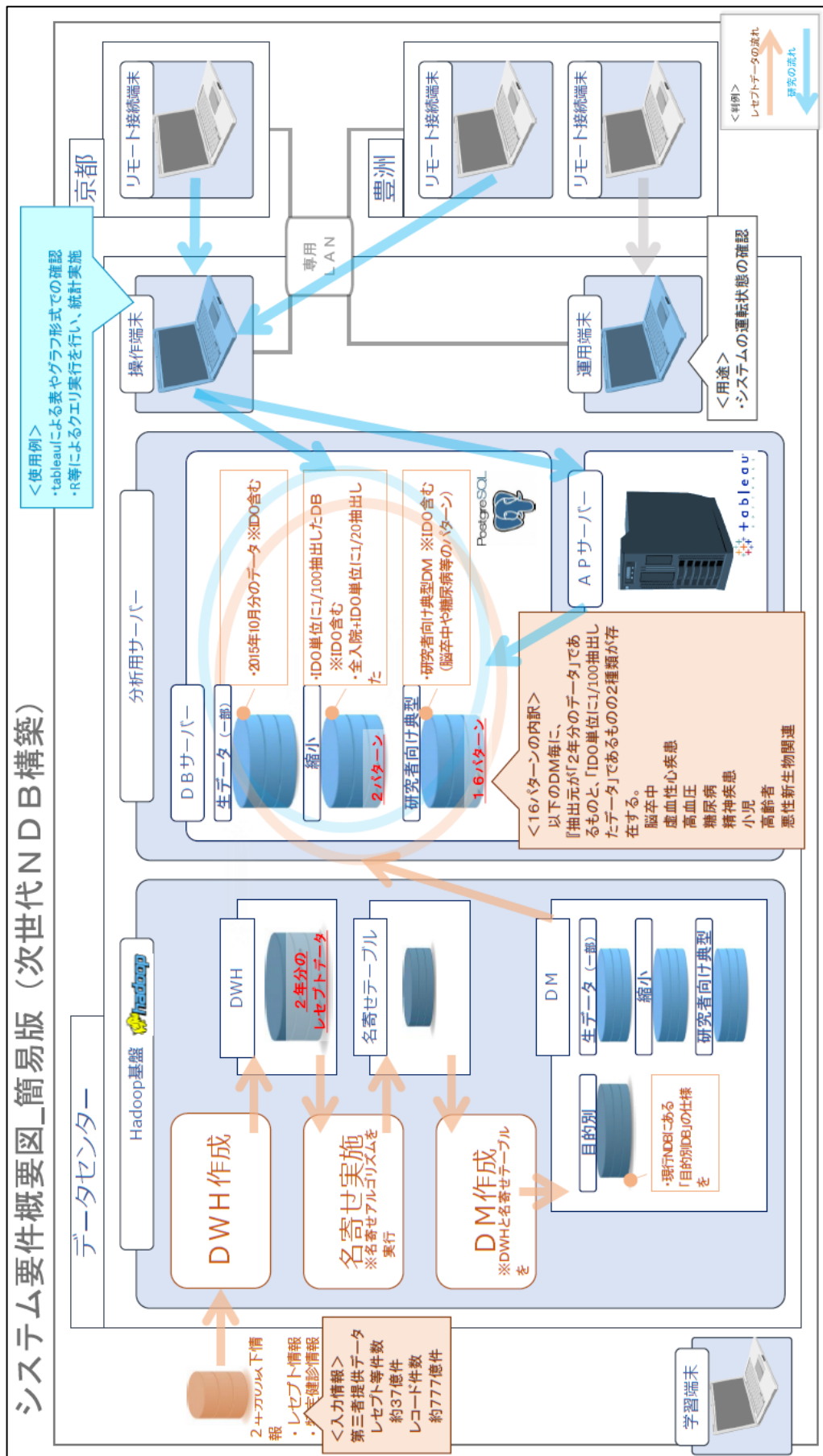
なし

(本研究にて使用する、NDB 並列分散処理基盤の構成)

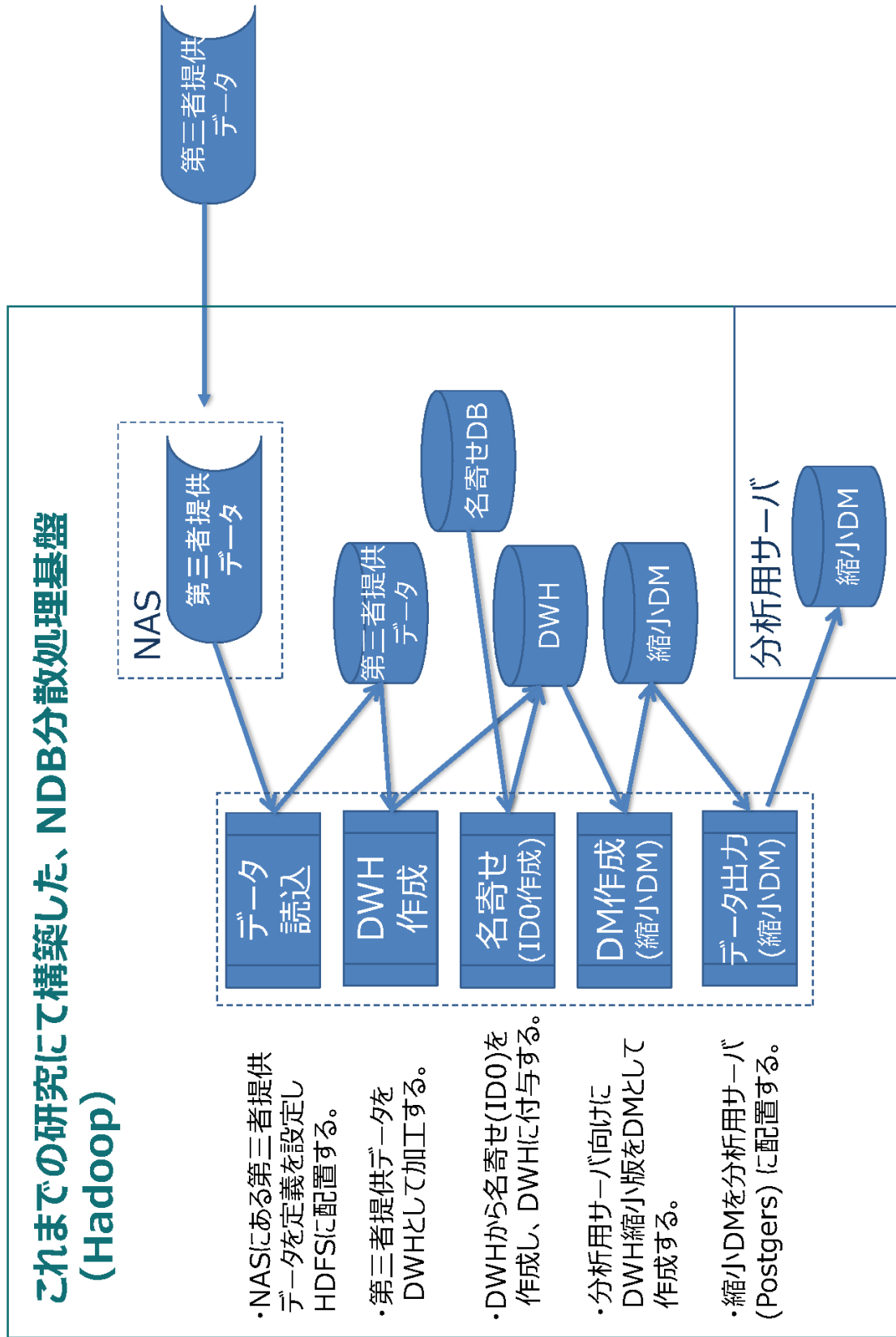


データロギング
スプレッド
データ
利権
運用

(NDB 基盤：システム要件概要図)



(参考：NDB および介護 DB 取り込みのプロセス)



NDB分散処理基盤 (Hadoop) に介護DBを追加

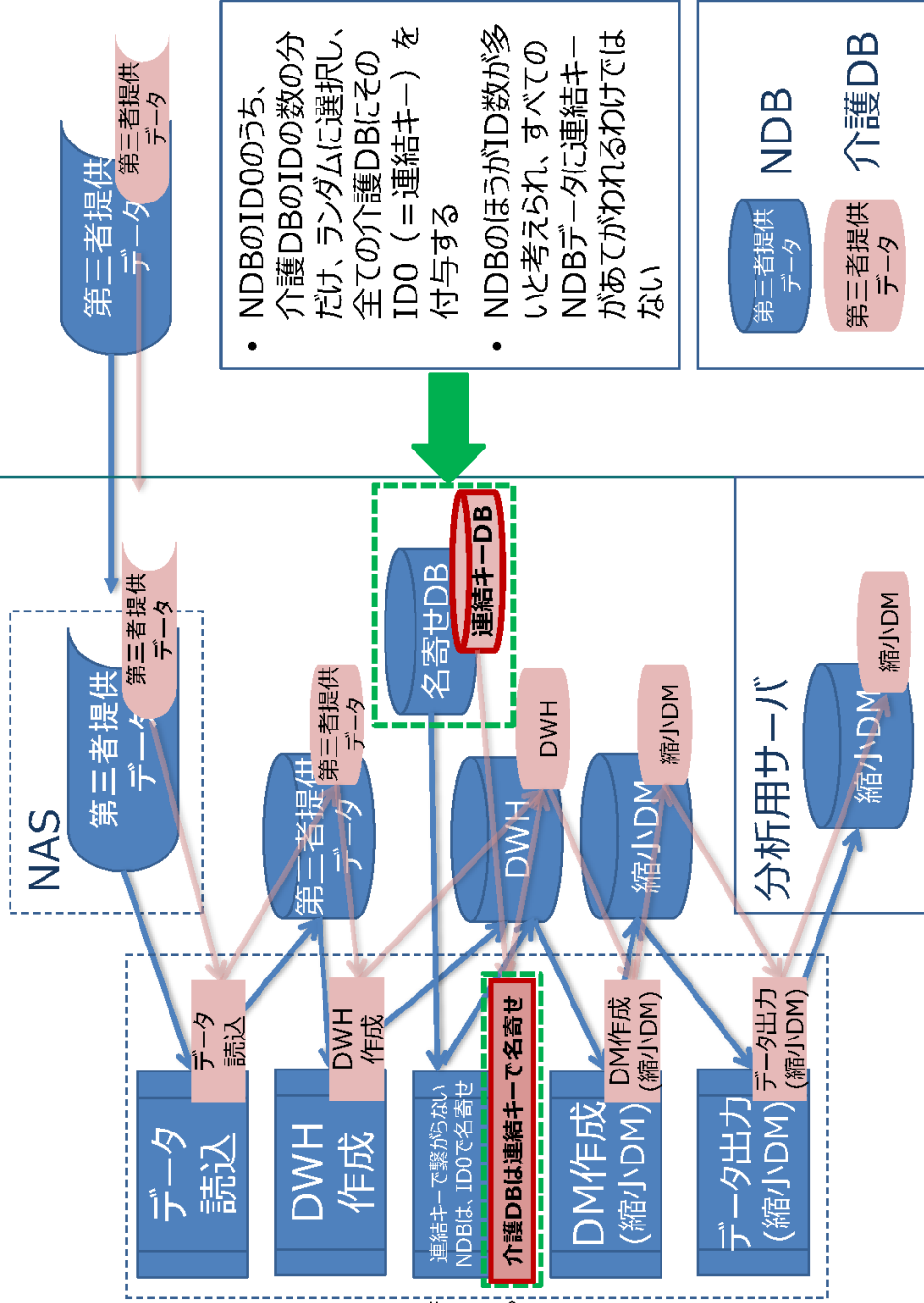
• NASにある第三者提供データを定義を設定しHDFSに配置する。

• 第三者提供データをDWHとして加工する。

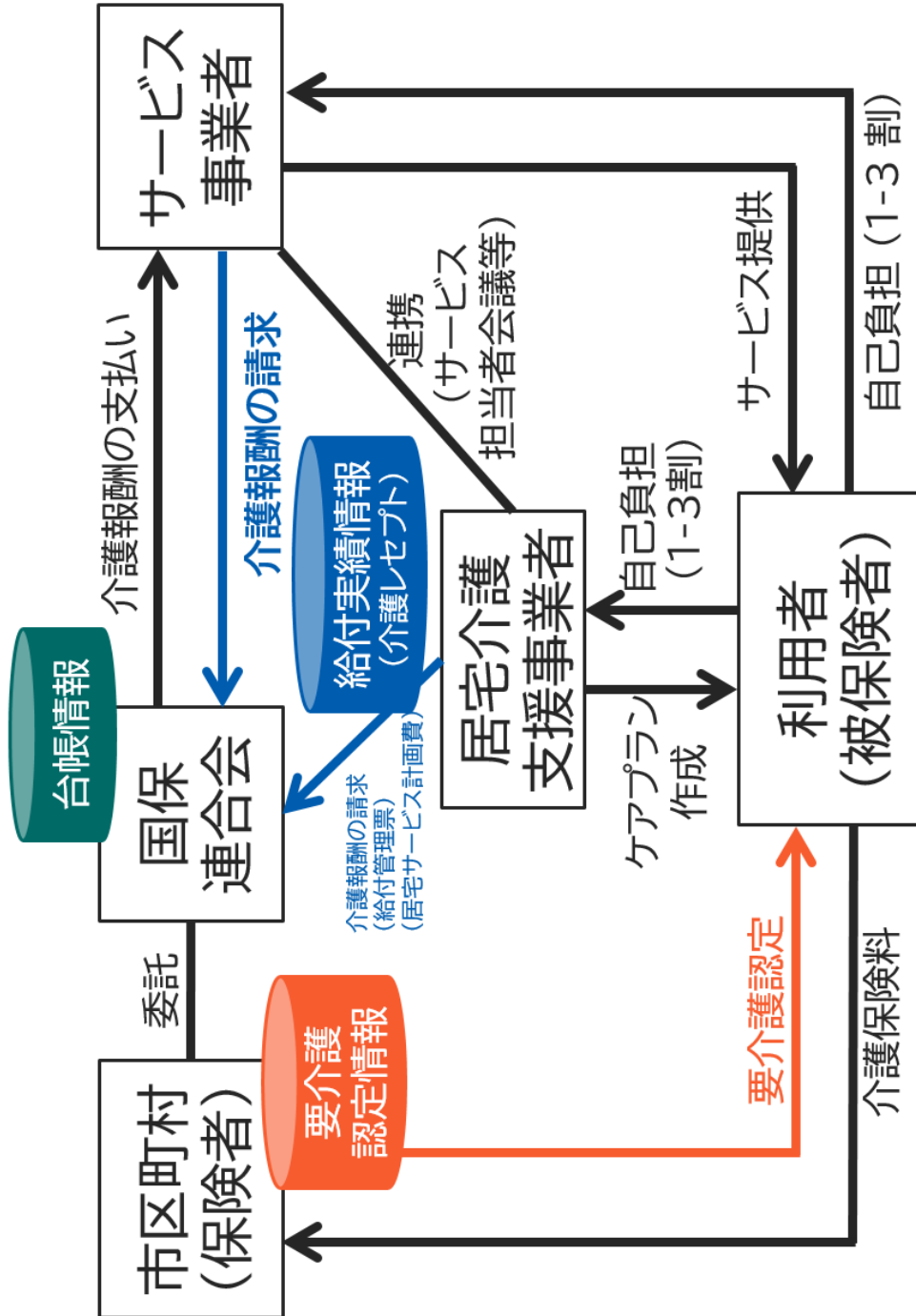
• レセプトデータと介護データを連結させるキーを疑似的に作成し、レセプトおよび介護データに付与。

• 分析用サーバ向けにDWH縮小版をDMとして作成する。

• 縮小DMを分析用サーバ(Postgers)に配置する。



(介護保険制度の流れと介護 DB 情報)



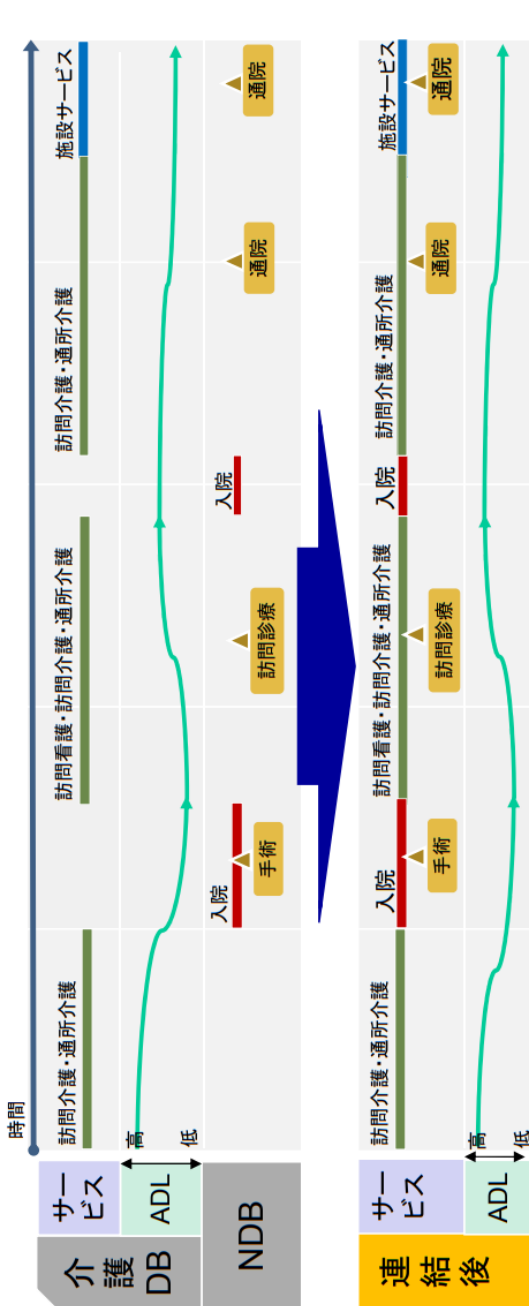
保険者である市区町村が所有する要介護認定情報、事業者から提出される介護レセプト（給付実績）情報、国保連合会が所有する台帳情報から介護 DB は成立している。

(NDB との連結でデータの質の向上が期待できる介護 DB 項目)

介護DBとNDBの連結

介護DB（要介護認定等情報・介護レセプト等情報）の課題

- 介護DBだけでは、高齢者が利用している介護サービスの種類・量・費用と要介護度・ADL等しか分からないため、それらの変化に影響したイベントが把握できない。
- NDBと連結されることで、少なくとも医療的なイベントと介護サービスの種類・量・費用と要介護度・ADL等との関係性が把握できる可能性がある。
- NDBにとっても、患者のアウトカムが分からない特性を一定程度補える可能性がある



第2回医療・介護データ等の解析基盤に関する有識者会議 資料 2-1
<https://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000209437.pdf>

(今回作成したデータマートのイメージ)

要介護認定情報

NINTEI

要介護認定データのクレンジング

＜要介護認定情報のポイント＞ ※データクレンジング

- ① 取下区分が有効でないものについては除外
- ② 次の有効期間(開始)より、有効期間(終了)を判断し、正しい終了日を作成。
- ③ 以下を条件に除外。
 - ③-1 認定有効期間(開始)が同日だった場合、二次判定日が最新のものを採用。
 - ③-2 さらに二次判定日も同日だった場合、申請日が最新のものを採用。

以降は月別展開。

認定申請日	取下区分	二次判定日	二次判定結果	認定有効期間(開始)	認定有効期間(終了)	認定有効期間新規(終了)
2014/08/01	1(認定有効)	2014/08/15	13(要支援2)	2014/09/01	2014/12/31	2014/12/31
2014/12/08	1(認定有効)	2014/12/20	13(要支援2)	2015/01/01	2015/03/31	2015/03/31
2015/02/10	2(却下)	2015/02/20	21(要介護1)	2015/03/11	2015/06/30	2015/06/30
2015/03/11	1(認定有効)	2015/03/22	13(要支援2)	2015/04/01	2015/06/30	2015/04/20
2015/04/21	1(認定有効)	2015/04/30	21(要介護1)	2015/04/21	2015/08/31	2015/08/31
2015/07/14	1(認定有効)	2015/07/22	21(要介護1)	2015/09/01	2015/11/30	2015/11/30
2015/07/22	1(認定有効)	2015/07/28	21(要介護1)	2015/09/01	2015/11/30	2015/11/30

NTT DATA


© 2020 NTT DATA Corporation

61

(「認定通番」のイメージ)

要介護認定データのクレンジング

要介護認定情報



認定申請日	取下区分	二次判定日	二次判定結果	認定有効期間 (開始)	認定有効期間 (終了)
2014/08/01	1(認定有効)	2014/08/15	13(要支援2)	2014/09/01	2014/12/31
2014/12/08	1(認定有効)	2014/12/20	13(要支援2)	2015/01/01	2015/03/31
2015/02/10	2(却下)	2015/02/20	21(要介護1)	2015/03/11	2015/06/30
2015/03/11	認定申請日	取下区分	二次判定日	認定有効期間 (開始)	認定有効期間 (終了)
2014/08/01					
2014/12/08					
2015/03/11					
2015/04/21	1(認定有効)	2015/04/30	21(要介護1)	2015/04/21	2015/08/31
2015/07/22	1(認定有効)	2015/07/28	21(要介護1)	2015/09/01	2015/11/30

年月情報、保険者番号、被保険者番号、認定有効期間 (開始) を連結

年月	保険者番号	提供先番号 (被保険者番号)	認定有効期間 (開始)	認定有効期間 現 (終了)	二次判定結果	認定通番
201409	131101	xosge93dkxa	2014/09/01	2014/12/31	13(要支援2)	131101xosge93dkxa 20140901
201410	131101	xosge93dkxa	2014/09/01	2014/12/31	13(要支援2)	131101xosge93dkxa 20140901
201411	131101	xosge93dkxa	2014/09/01	2014/12/31	13(要支援2)	131101xosge93dkxa 20140901
201412	131101	xosge93dkxa	2014/09/01	2014/12/31	13(要支援2)	131101xosge93dkxa 20140901
201501	131101	xosge93dkxa	2015/01/01	2015/03/31	13(要支援2)	201501131101xosg e93dkxa20150101
...

NINTEIに合わせたキーを付与。(左記は同一キー)

12/14向け修正要

© 2020 NTT DATA Corporation

6Z

NTT DATA

(介護レセプト側への「認定通番」の付与)

納付実績情報 (基本情報) H1

介護DM (認定情報) NI

被保険者情報有効期間〇〇年月日を使用。(「開始年月日」、「終了年月日」ではない。)

一意になるか確認要。

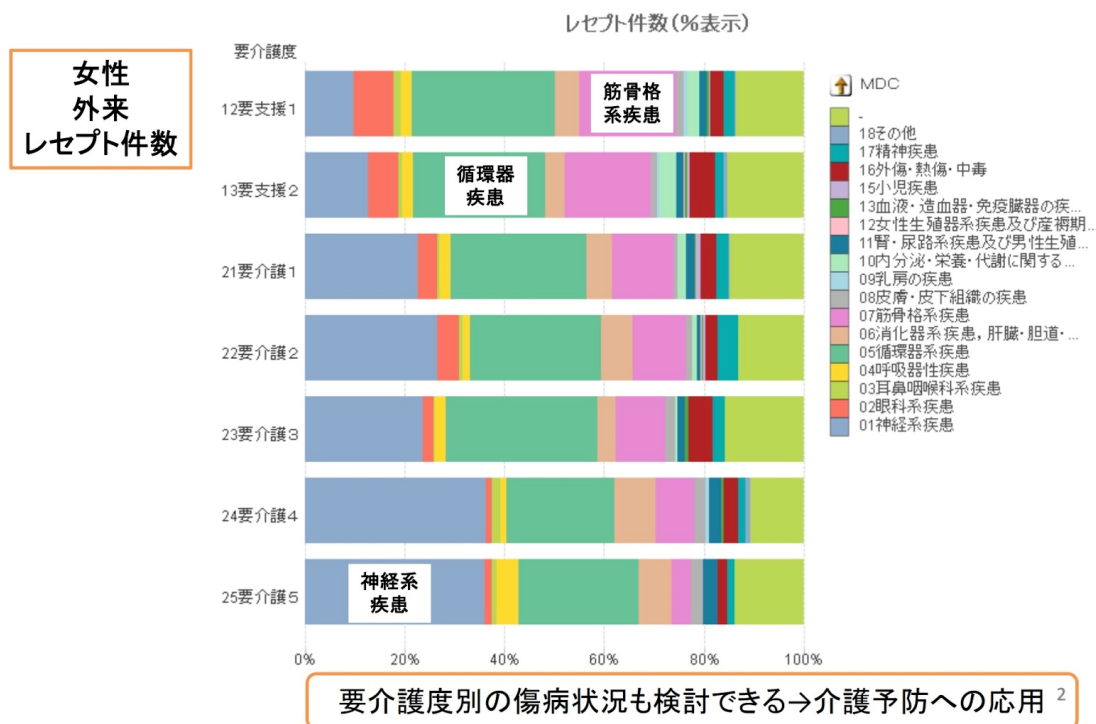
年月、開始年月日、終了年月日、id0 (id特定キー) をもとにNIからH1に認定通番を付与する。

NI用に作成した新規有効期間(終了)ではなく、もと認定情報にあった有効期間(終了)を使用する。(NINTEIとH1に同様の情報が設定されていると想定。)

サービス提供年月	被保険者情報有効期間開始年月日	被保険者情報有効期間終了年月日	id0	認定通番
201409	2014/09/01	2014/12/31	aabbcc0000	131101xosge93dkx a20140901
201410	2014/09/01	2014/12/31	aabbcc0000	131101xosge93dkx a20140901
201411	2014/09/01	2014/12/31	aabbcc0000	131101xosge93dkx a20140901
201412	2014/09/01	2014/12/31	aabbcc0000	131101xosge93dkx a20140901
201501	2015/01/01	2015/03/31	aabbcc0000	131101xosge93dkx a20150101
...
年月	保険者番号	認定有効期間(開始)	認定有効期間(終了)	認定通番
201409	131101	2014/09/01	2014/12/31	131101xosge93dkx a20140901
201410	131101	2014/09/01	2014/12/31	131101xosge93dkx a20140901
201411	131101	2014/09/01	2014/12/31	131101xosge93dkx a20140901
201412	131101	2014/09/01	2014/12/31	131101xosge93dkx a20140901
201501	131101	2015/01/01	2015/03/31	131101xosge93dkx a20150101
...

(事例1の参照事例)

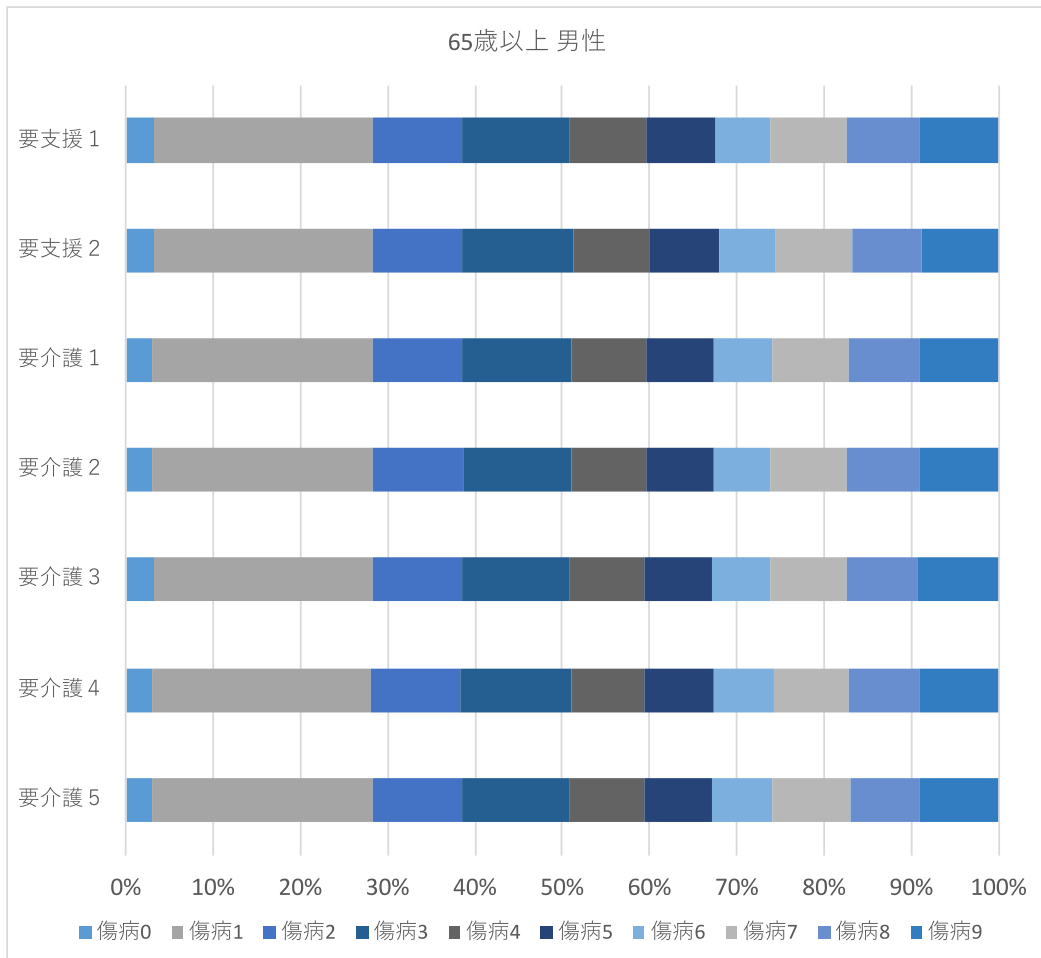
医療・介護レセプト連結分析の例(1)

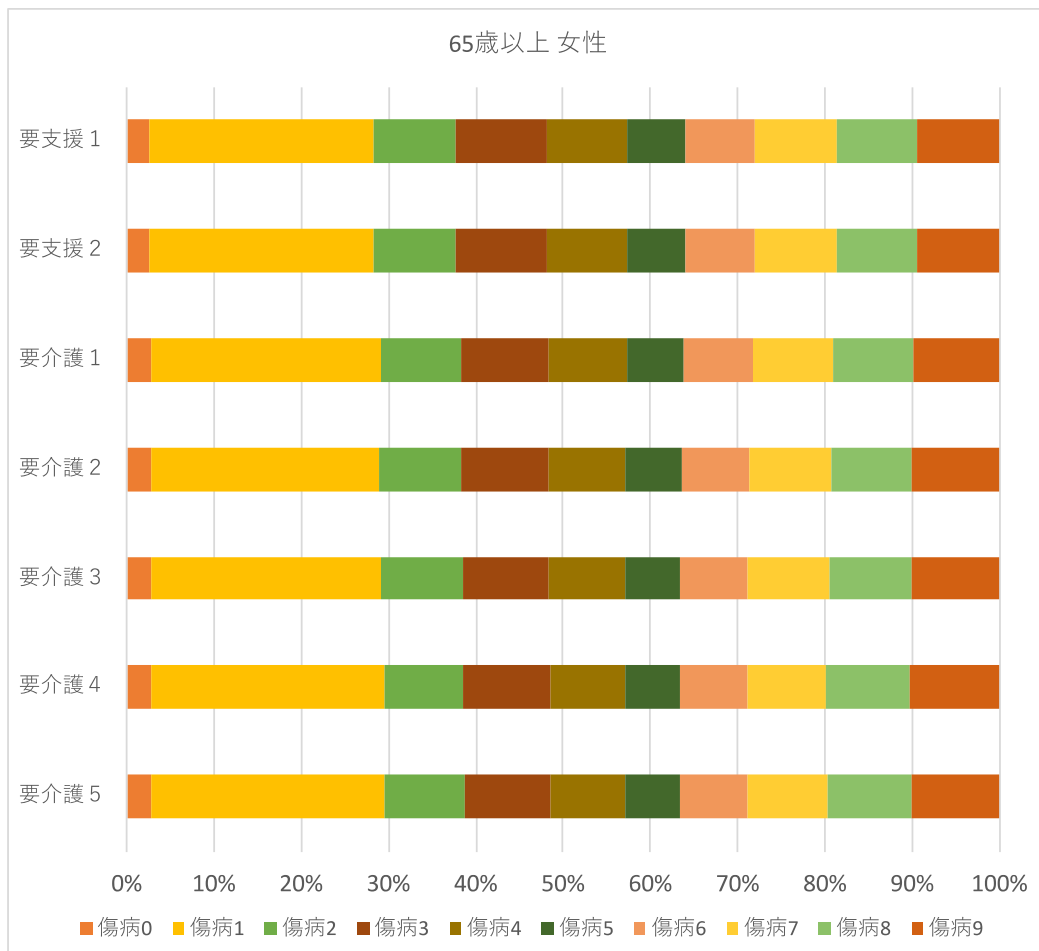


第1回医療・介護データ等の解析基盤に関する有識者会議 資料3

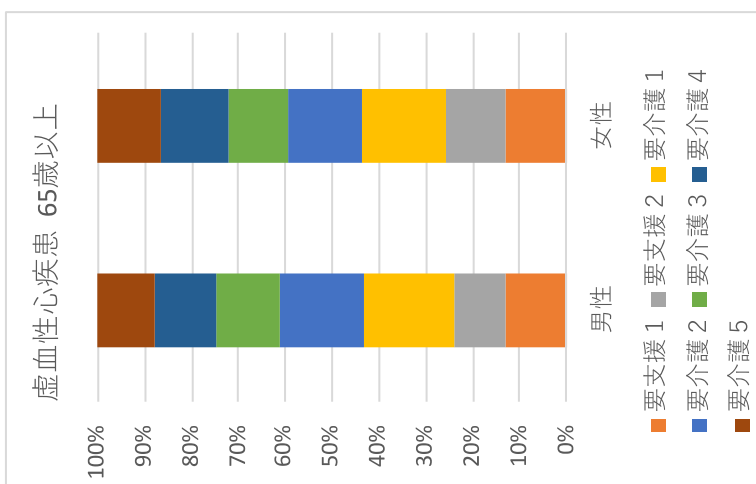
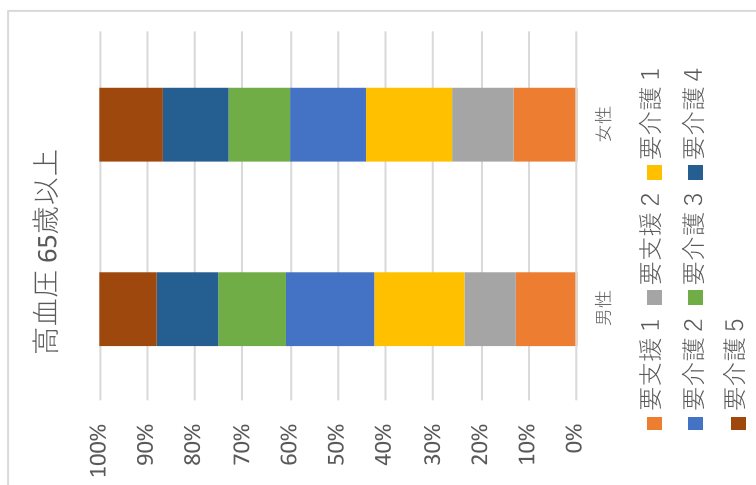
<https://www.mhlw.go.jp/stf/shingi2/0000206551.html>

(事例1：集計(a)の結果)

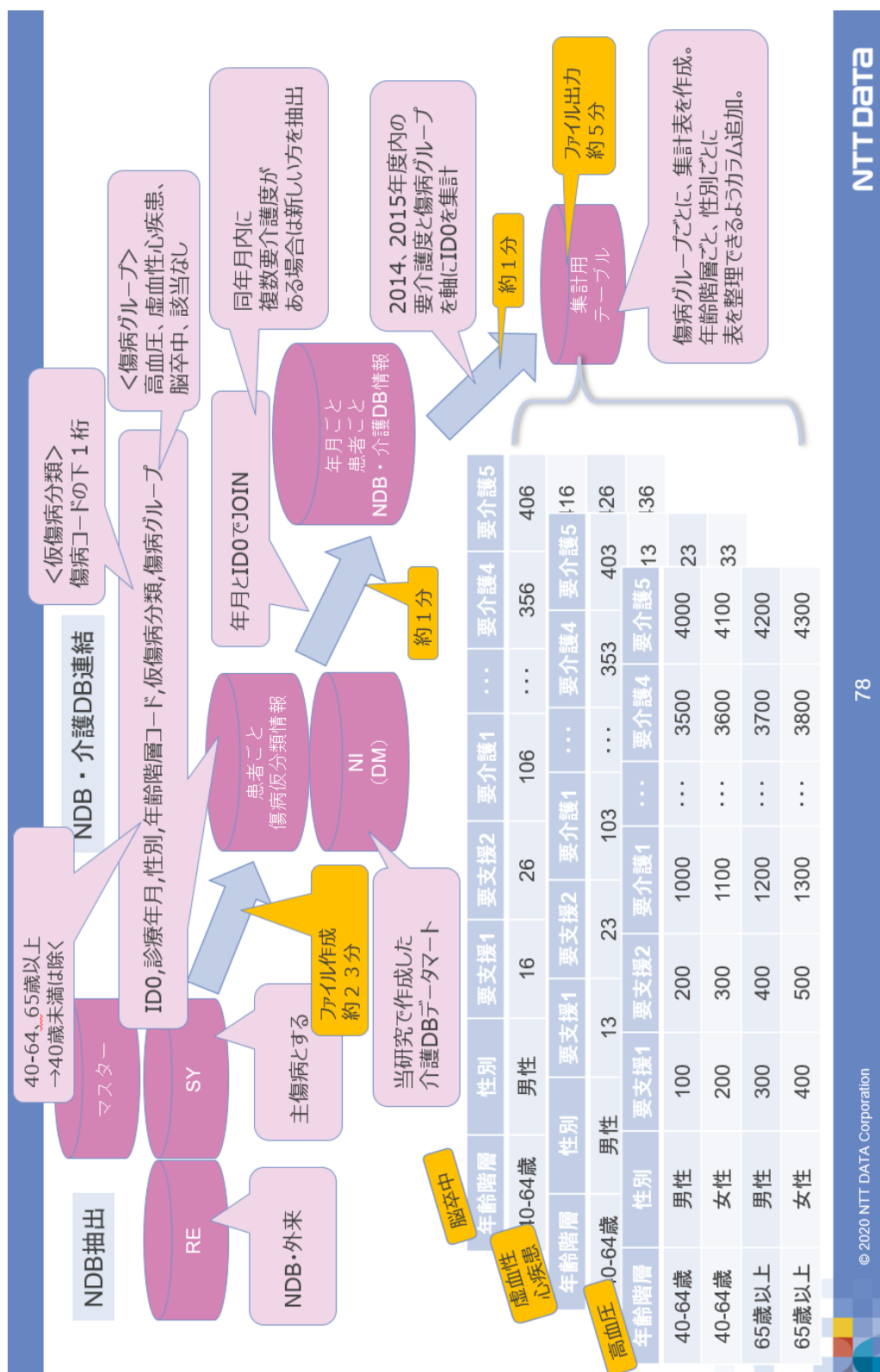




(事例1：集計 (b) の結果)



(事例1：集計における手順及び所要時間)



(事例 2 : 結果)

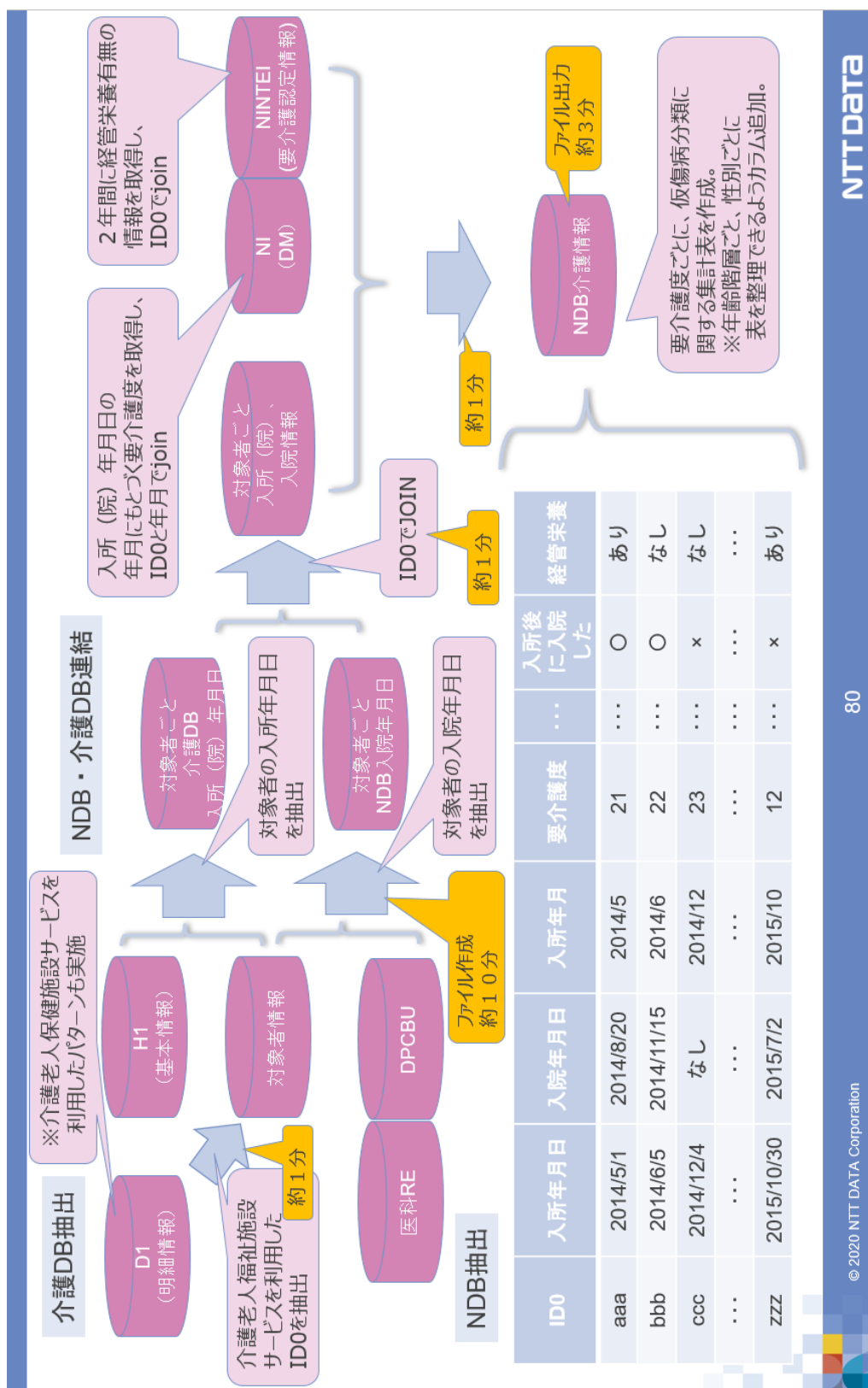
特別養護老人ホーム

	入所後入院なし		入所後入院あり		合計
経管栄養なし	168,000	(70.9%)	69,000	(29.1%)	237,000
経管栄養あり	8,000	(72.7%)	4,000	(36.4%)	12,000
欠損	1,000		1,000		2,000
合計	176,000	(71.0%)	73,000	(29.5%)	249,000

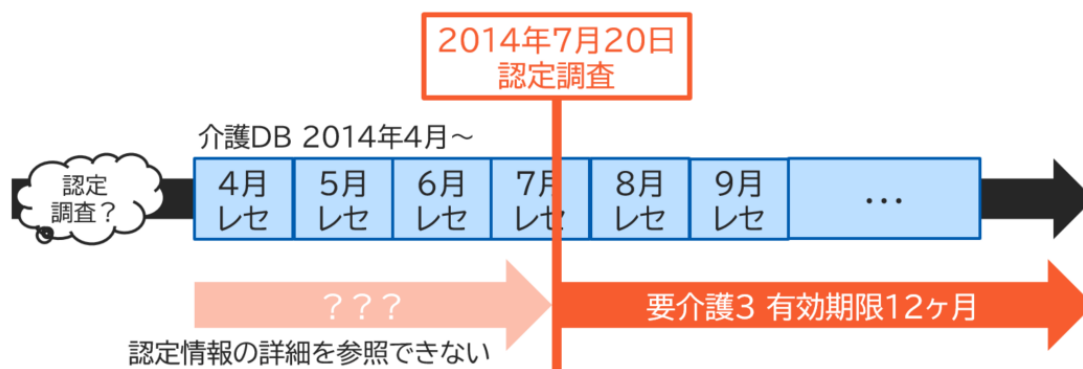
介護老人保健施設

	入所後入院なし		入所後入院あり		合計
経管栄養なし	285,000	(73.1%)	106,000	(27.2%)	391,000
経管栄養あり	13,000	(72.2%)	5,000	(27.8%)	18,000
欠損	1,000		1,000		2,000
合計	299,000	(73.1%)	111,000	(27.2%)	410,000

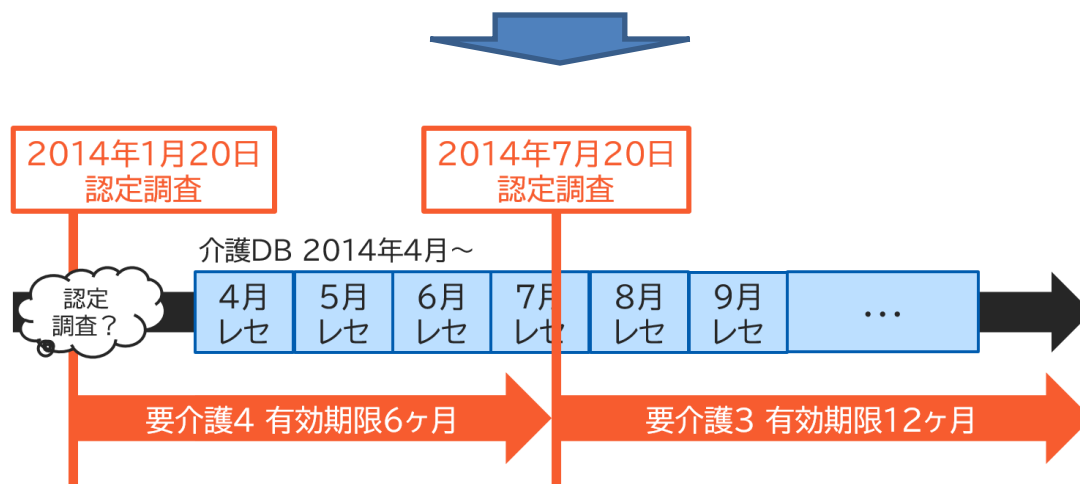
(事例 2 : 集計における手順及び所要時間)



(補足：匿名要介護認定情報の参照)



上記の例のように、匿名要介護認定情報と匿名介護レセプト等情報のデータの対象期間の起点を同一（つまり2014年4月）にしてしまうと、それ以降、最初の認定調査が仮に2014年7月20日に行われていたとすると、それ以前の2014年4月、5月、6月の匿名介護レセプト等情報に対して参照しうる匿名要介護認定情報を得ることができない。



上記の例のように、匿名介護レセプト等情報に結び付け得る匿名要介護認定情報は匿名介護レセプト等情報以前の認定調査を参照する必要があるため、データ利用時には注意が必要である。

(NI テーブルへの DT5341 (受給者台帳) の資格喪失年月日の付与)

