

令和5年4月19日

第15回匿名医療情報等の提供に関する専門委員会

資料1

# 医療・介護データ等解析基盤（HIC）の本格運用を見据えた NDBデータの利活用の更なる促進について

厚生労働省

Ministry of Health, Labour and Welfare of Japan



## 【NDBの提供形式】

	特別抽出（本省利用）	特別抽出（第三者提供）	オープンデータ	サンプリングデータセット
内容	政策の企画立案のため、厚労省の指定に沿ってデータ抽出を行い、個人単位のデータを利用。	審議会による提供承諾を得た研究において、研究者の依頼に応じたデータ抽出を行い、個人単位のデータを提供	診療行為、処方薬、健診項目等について、全国レベルで集計を行った集計表（csvファイル形式）	1ヶ月分（1月/4月/7月/10月診療分）の匿名レセプトデータから、入院診療10%、外来診療1%を抽出し、高額レセプトの削除等の匿名化処理を行ったプリセットデータ
利用方法等	<ul style="list-style-type: none"> <li>患者個人単位の経時的な追跡など、<u>複雑な解析が可能</u></li> <li>厚労省の指定に沿って、事務局で集計表までを作成。</li> <li>審議会の審査は不要。</li> </ul>	<ul style="list-style-type: none"> <li>患者個人単位の経時的な追跡など、<u>複雑な解析が可能</u>（NDBデータに一定の見識を有した研究者が利用）</li> <li>利用者希望に応じ、事務局で集計表までを作成し提供（NDBのデータ分析ノウハウのない研究者に有用）</li> <li>提供日数は平均330日（2021年度）。提供件数は年間約30~50件。</li> </ul>	<ul style="list-style-type: none"> <li>厚労省HP上に公開。利用者は審議会による審査を経ずに自由に即時ダウンロード・利用可能</li> <li>初診料・再診料、特定の処置、手術等の診療行為、利用頻度の高い医薬品等の利用傾向把握に有用。</li> <li>全国レベルの集計に加え、診療月・性年齢・都道府県・二次医療圏別の層別集計結果も公開。より細かい傾向の把握も可能。</li> <li>オープンデータへのアクセス件数は、年間約44万回（2021年度）</li> </ul>	<ul style="list-style-type: none"> <li>審議会の承諾が必要。<u>プリセットデータをそのまま提供するため、提供日数が短縮。</u></li> <li>提供件数は、年間約10件。承諾された研究目的の範囲内で、<u>探索的研究を行うことが可能であり、自由度は高い</u></li> <li>提供日数は、データセット再作成やNDBクラウド化による抽出停止期間の影響を受けた期間は、平均109日（2020年度）と平均182日（2021年度）であったが、これらの影響のない期間は平均52日（2019年度）。</li> </ul>

## 【NDB抽出件数等】

	平成31年度	令和2年度	令和3年度	令和4年度	令和5年度予算
抽出件数	212件 本省利用 140件 第三者提供 72件	188件 本省利用 144件 第三者提供 44件	292件 本省利用 218件 第三者提供 74件	277件(見込み) 本省利用 208件 第三者提供 69件	—
抽出量	計149TB 本省利用 77TB 第三者提供 72TB	計151TB 本省利用 92TB 第三者提供 59TB	計271TB 本省利用 136TB 第三者提供135TB	計174TB(見込み) 本省利用 75TB 第三者提供 89TB	—
事業費 ※R4年度～予算ベース	2.9億円	3.2億円	4.6億円	9.2億円 ※令和4年4月からクラウド化に伴う利用料を含む	9.6億円

## (参考) NDB利活用の促進に係る指摘について

### ○デジタル田園都市国家構想総合戦略(令和4年12月23日閣議決定)

- レセプト情報等を活用した分析の体制整備の推進・保健医療に関するビッグデータの効果的な利活用を推進するため、NDB(レセプト情報・特定健診等情報データベース)や、NDBと介護DB(介護保険総合データベース)の情報の連結解析を可能とするために構築した医療介護連結解析基盤(HIC)について、ポータル機能や探索的利用環境の運用開始といった機能の充実や利便性の向上等を図る。

### ○規制改革推進に関する中間答申(令和4年12月22日規制改革推進会議)

#### II 重点分野

- 「社会課題の解決」をフロンティアとして経済成長を実現すべく、会議として来夏の答申の取りまとめに向けては、(1)デジタル時代の規制改革、(2)GX推進のための規制改革、(3)人口減少等に対応した規制改革、(4)地方の活性化を図る規制改革の4つの分野において重点的に取り組み、これを突破口として規制改革の展開を図る。
- (略) こうした考えの下、以下の事項を突破口にデジタル時代の規制改革を進める。
  - (略)
  - 医療データの利活用促進

### ○規制改革推進会議 医療・介護・感染症対策WG(令和4年11月7日)

※出席者：佐藤座長(一橋大学教授)、武井座長代理(西村あさひ法律事務所)、杉本委員(日本大学教授)、中室委員(慶應義塾大学教授)、菅原委員(経済同友会常務理事)、印南委員(慶應義塾大学教授)、大石委員(メディヴァ社長)、大浦委員(おとなの学校取締役)、大橋委員(東京大学副学長)、佐々木委員(悠翔会理事長)、落合委員(渥美坂井法律事務所)、戸田委員(日本電気常務理事)、林委員(桜坂法律事務所)、説明者：宮田俊男氏(早稲田大学理工学術院教授)、他

議題：医療データ等の利活用の促進について

#### (1) NDBの利活用容易化について

厚生労働省、日本製薬工業協会、健康保険組合連合会からヒアリング

## (参考) 医療・介護・感染症対策WG議事録(抜粋)

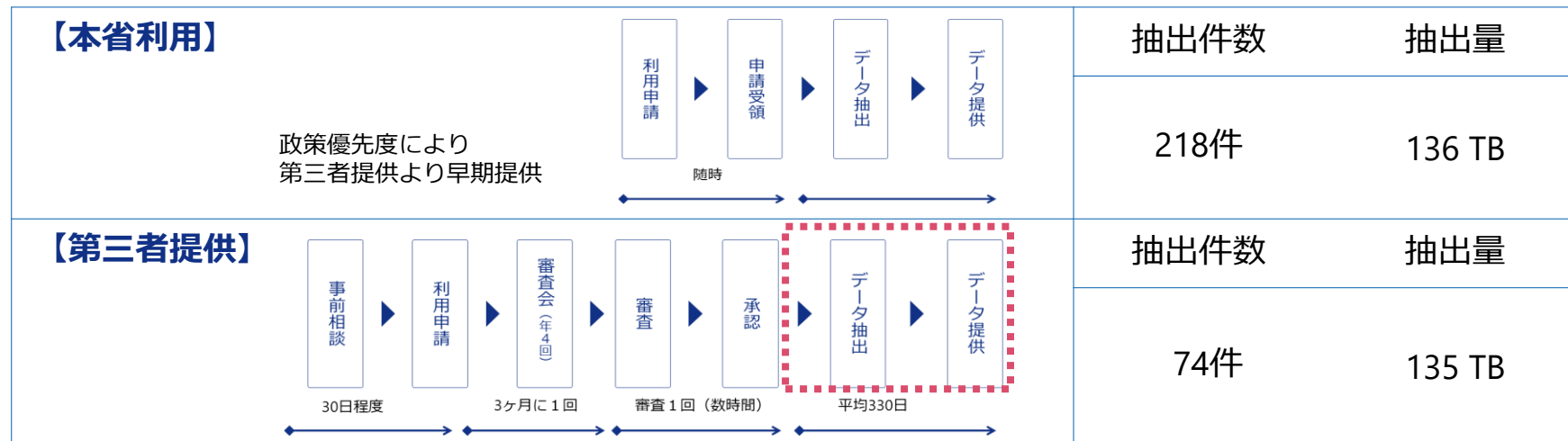
- 宮田教授 各種ヒアリングで提供までに1年ぐらいの時間を要するというので、これは世界的に見ても遅いと思います。そこはできるだけ時間が短縮されるように、何かしら知恵を絞って頑張ってくださいと思います。(知財について) その辺りも民間事業者のデータを使う場合は、きっと使えると思うのですが、日本は医薬品開発が世界にかなり遅れているものもあると思うので、何かしら工夫をしてやってほしいということと、実際、使用用途目的についても、医薬品開発というのは、どうしても営利的な側面もあると思うのですが、その線引きは難しいですし、最終的には公衆衛生にも寄与できるものになると思うので、今の健保組合や製薬協さんのプレゼンのとおり、そこも柔軟にできるようにされるといいと思います。今、製薬業界は、希少疾患とか、難病の製品開発をすごく強く進めている一方で、例えば各県に1人しかいないような希少疾患の患者さんのプライバシーが同定されるかもしれないという部分もリスクとして行政は考えていると思うのですが、ただ、実際に難病の患者さんなどに私が話を聞くと、そもそも薬がないから、そうしたリスクは多少あっても、そういうものを使ってほしいという話もあります。ゼロリスクは何事も難しいと思うので、その辺りのベネフィットとリスクのバランスを取って、その部分もなるべく柔軟に使えるように工夫してもらえるといいと、常日頃、私も思っています。
- 中室委員 申請からデータを使えるようになるまで330日かかって、これから改善をしても250日というのは、正気の沙汰とは思えない時間がかかっていると思います。当然そこから分析するのに時間がかかるので、例えばコロナの状況などを分析しようと思ったら、データが来るまでに1年かかって、そこから仮に半年なり、1年なり分析に時間がかかったら、2年かからないとコロナの分析はできないということなので、日本の研究者がコロナの分析を2年もたって発表していたら、世界中からあはだと思われるのではないかと思います。なので、仮に改善が行われたとしても、250日というのは、全く論外の時間のかかり方だと思います。そう考えると、海外はもっと短い時間で、多くの件数で出せているわけなので、現状の分析自体が間違っているというか、方向が違っているのではないかと思います。私どものほうで調べてみたら、例えばイギリスでは、イギリス政府が運営しているNHS Digitalなどが提供しているようなものと、2010年からの利用実績は3,000件以上ある。しかも、日本よりも短いスピードで出せているわけなので、どうして海外と同じようなスピード感と規模でやらないのかということについて、お伺いしたいと思います。
- 印南専門委員 オープンデータ、サンプリングデータ、ダミーデータでも何でもいいのですが、これらは時系列的にはくっつかない仕組みになっていませんか。悉皆でなくても、コホートデータみたいに一定の人たちを匿名化した情報で時系列的に連結してあれば、割と短期間に分析できて、政策提言などに間に合うと思います。そうでないと、理論的にくっつけることができなかつたりすると、使い勝手が悪いと思います。政策提言に向けたデータ利用に対しては、毎年ランダムにサンプルされてしまったりすると、むしろ意味がないと思います。パネルデータでないという意味がないのです。
- 落合専門委員 インセンティブがないということだと、どうしても企業活動の中でそこにリソースを割くことが難しくなってくるかと思います。これは企業だけではなくて、研究者の方々であったとしても、何らかの可能性があると重要だと思います。この部分については、政策的にどういうふう考えていくか、適切に研究開発を促進していくような取組をしていくのかどうかの判断に大きく関わるところだと思います。インセンティブ設計の点について見直しも考えていただきたいと思っています。もう一つは、要件に関する部分です。利用できる要件が厳しいということで、利用されていない部分があるのではないかと考えております。特に公益性について、判断ができるかどうか分からないことは重大な問題なのではないかと考えております。例えば創薬等の初期の探索的な研究から始まって、開発前のニーズ調査であったり、副作用の分析、治療のアドヒアランスといったような、自社の医薬品の研究開発、安全性にわたるようなことも、どういう場合に利用できるのかをより明確にして、利用しやすくすることが重要だと思います。

## (参考) 医療・介護・感染症対策WG議事録(抜粋) その2

- 落合専門委員 公益性だけではなく、セキュリティの部分も利用要件としてかなり厳格な設定がされているのではないかと思います。なぜこういうセキュリティの要件が設定されているのかですが、そもそもデータ自体が漏えいしてしまう、何のために行っているのかによって、どういう形で合理化をしていくべきかの考え方が変わってくると思うので、趣旨を伺いたいと思います。データの漏えいを防止することを主に考えられているのか、それとも実際にデータを抽出して、それを利用することによって個人を選別したり、弊害が起きることを防止するためなのか、ほかのことにあるのであれば、それはそれでいいのだと思いますが、それを示していただくことが必要だと思います。いずれにしても、これがどういう趣旨なのかということを踏まえ、セキュリティの部分は、特にクラウド化で使いやすくすることをしっかり行っていべきかと思っております。その際に対策として何をすべきなのかを考えるに当たって、ここは決定的に重要だと思います。例えば法的な義務であったり、もし法改正をされるのであれば、不適正目的での利用であれば、罰則までつけるということすら考えられるかもしれません。何をすべきなのかという前提としては、何を心配するのかということによると思いますので、そこは相当の公益性とセキュリティを踏まえ、なぜそういう厳しい要件にできてきたかという2点について、明確にさせていただきたいと思っております。
- 戸田専門委員 今の新しいシステムの仕様要件として、要求仕様の確定から抽出が終わるまでのターンアラウンドはたしか5日という資料があったかと思うのですが、その性能は十分に出ているのでしょうかという確認が一つです。それから、今、申請側と運用事業者側との間のすり合わせで、ものすごく時間がかかっているということで、これは抽出条件をSQLに落とし込むためのやり取りに時間がかかっているということだと思っておりますけれども、これを増員で対応されるというお話ですが、人手をかけて効果が出るのでしょうかということを疑問に思っています。例えば、申請する側でトライ・アンド・エラーというか、いろんなシミュレーションを実施できる環境を準備してあげて、抽出条件の精緻化をもっと効率化させる取り組みも必要ではないかと思うのですが、いかがでしょうか。
- 大石専門委員 本件は莫大な手間と時間と労力をかけてつくったのだと思います。国民皆保険制度を持っている日本のデータは、ずり落ちていく日本の国力を上げるかもしれない。最終的にアウトプットは論文でもいいし、薬でもいいのですけれども、いずれにしても、最大限に活用すべきものだと思います。その状況から鑑みたときに、海外と恐ろしく時間差があることについて、やはり納得ができない。なので、一度、海外はどうやっていて、日本はどうやっていて、その差は何で発生しているのかということを明確に指し示していただきたいと思っております。これが1点です。二つ目は質問なのですが、もしかしたらということなのですが、いわゆるこれはビッグデータなので、普通のデータベースのエンジニアでは多分対応できなくて、要はビッグデータをちゃんと扱えるエンジニアであるとか、もしくはNoSQLみたいに、ビッグデータ専用のデータの持ち方、システムとか、そもそもやっていることが違うのではないかとかと思っております。あと、先ほど審査会が3か月に一度しかない、かつ審査に数時間かかるという話があったのですが、そもそも業務の見直しも必要なのではないかかと思っております。ですから、海外との差分の発生の理由、また、データベースの持ち方等のシステム上の課題、業務フローの課題等を1回見直していただきたい。そういうことにたけた専門家の目をきちんと入れていただいたほうがいいと思います。
- 林専門委員 NDBの提供の流れのところですが、それぞれ短縮する、または並行してできるものは、並行して行うという形で、効率化していただきたいと考えております。例えば事前相談の30日程度というところは、申込みから1週間以内に厚労省から回答する。そして、審査会はオンラインで毎週1回なり、委員が回答を送るという形で、別にみんなで合議しなくても、持ち回りでもよいと思われるので、そういうことをやって、ここを迅速化していただきたいと思っております。少なくとも改善しても250日というのは、あり得ないと思います。先ほどどなたかが笑ってしまうとおっしゃっていましたが、こんなことを許してはいけないと思っております。
- 佐藤座長 データは非常に貴重な資源だと思いますので、資源の有効利用という視点はあつてしかるべきだという気がしました。厚労省さんには短くしていただいていますし、努力していただいているのは分かるのですが、根本的に考え方を変えないと、もしかしたらこの問題は解決しないのかもしれませんということでしょうか。どうもありがとうございました。

# NDBデータの提供について

## 【利用申請からデータ提供まで】



※令和3年度実績値

## 【第三者提供における承諾からデータ提供の内訳】

※令和3年度に承諾を受けて翌年度(12月時点)に提供した4件の実績

	研究者の書類提出 ※1	①抽出条件の確定		②抽出待ち ※2	抽出作業 ※3	手数料納付等 ※4	媒体発送 ※5	合計
		コーディネータ	DB技術者					
平均日数	32	151	9	55	44	37	3	331

### 課題:

- ・研究者とデータベース技術者(SE)をつなぐコーディネータの不足
- ・研究者のNDBデータリテラシーの不足

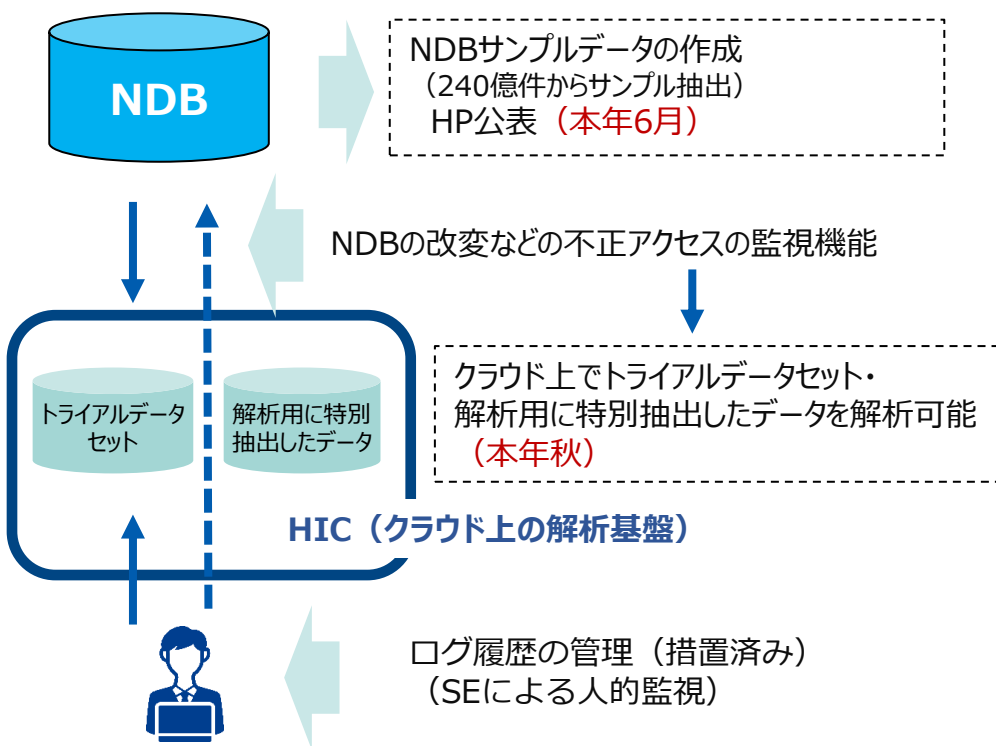
- ・物理的制限(サーバー数)
- ・人的制限(技術者(SE)の数)

- ※1 研究者の書類提出(データ抽出依頼書・誓約書)
- ※2 抽出作業まわりの滞留期間
- ※3 プログラムの動作点検・修正対応を含む
- ※4 手数料通知の発送 + 研究者からの手数料納付
- ※5 R5年度中のHICの運用開始により、クラウド上での利用が可能に

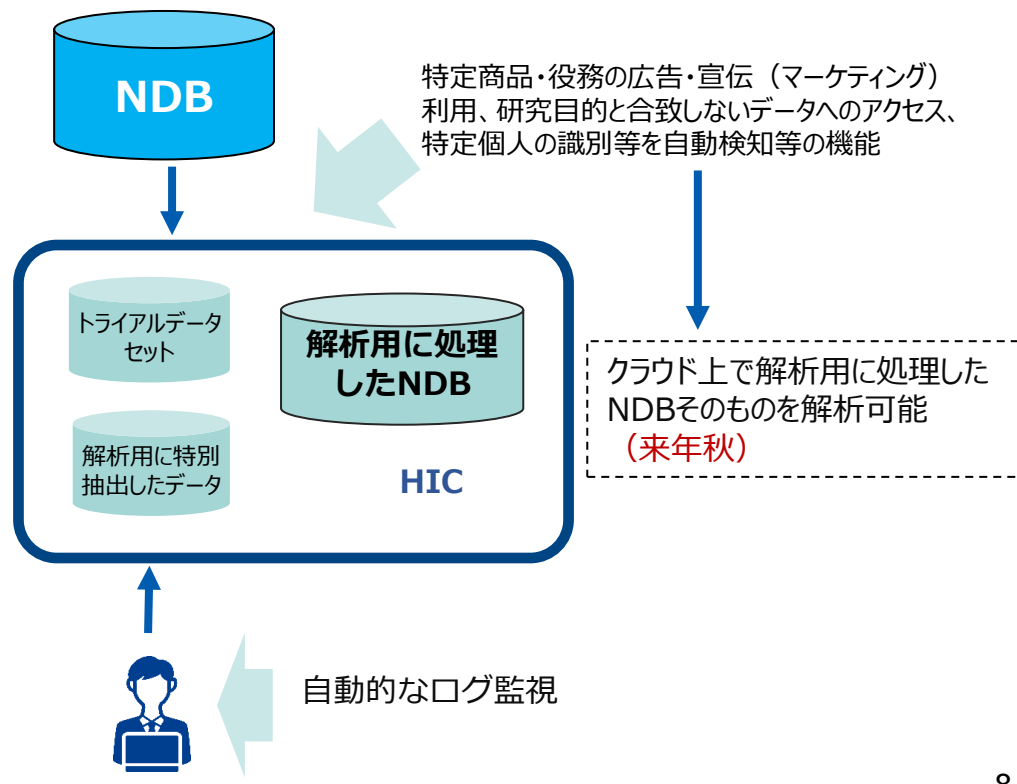
# NDBデータ提供の抜本的見直し（概要）

- **直ちに**、サンプルデータの作成、トライアルデータセットの作成、不正アクセス監視機能の実装に取り組み、
    - ・ **本年6月**、NDBサンプルデータを厚労省HPに公表
    - ・ **本年秋**、リモートアクセスでトライアルデータセット・解析用に特別抽出したデータを解析可能
  - さらに、不適切利用等の監視機能やポータルサイトの機能拡充を開発・実装の上、
    - ・ **来年秋**、リモートアクセスの解析データを拡大
    - ・ 申請からデータ提供まで平均390日の現状に対し、申請~~メ~~切を毎月設定し、申請から**最短7日**で処理
- ※申請が月5件程度であることを踏まえ、当面月1回を設定するが、今後申請件数が増えれば複数回設定する

## 【不正アクセスの監視機能の実装】



## 【不適切利用の監視機能の実装】





# NDBサンプルデータ（データ見本）の提供（案）

NDBのスムーズな利用申請を支援するため、データ構造の理解を促進する目的で、数MB程度のサンプルデータ（データ見本）を厚生労働省HP上で公開する。

医科入院外レセプトの例

NDBの格納項目



高確法における匿名加工

- ①提供の都度、ハッシュ化のアルゴリズムを変える
- ②特異な情報を削除等  
→100歳以上一括りなど
- ③機微な情報の加工  
→医療機関コード等は匿名で提供等



レセプトデータ（イメージ）

1,2,0,IR,1,13,1,9999913,,サンプル病院,202005,00,03-9999-9999  
1,3,0,RE,9,1117,202004,サンプル  
丸,1,19490208,,20200304,,,29,,0009,,,,,  
10,,,,,  
1,4,0,HO,06132013,1 2 3 4 5 6 7,  
1,3,82255,,6,3990,,,57600,,,  
1,6,0,SY,1509003,20151102,1,,,01,  
1,7,0,SY,8841698,20151110,1,,,  
1,8,0,SY,2500015,20151110,1,,,  
1,9,0,SY,1509003,20151217,1,8048,,01,  
1,10,0,SY,8842624,20151222,1,,,01,  
1,11,0,SY,1101003,20151222,1,,,  
1,12,0,SY,6869044,20160530,1,,,

診療年月	医療機関コード	保険者番号	保険医療機関名	被保険者証記号番号	氏名	生年月日	性別	傷病名	診療行為	請求点数
平成 22年 4月分	13	9999913	1 医科 1 社保 1 単健 8 高外一	06132013	サンプル 79	19990303	男	糖尿病(主)	12 01 内診	1.619
								糖尿病(主)	12 02 内診	1.619
								糖尿病(主)	12 03 内診	1.619
								糖尿病(主)	12 04 内診	1.619
								糖尿病(主)	12 05 内診	1.619
								糖尿病(主)	12 06 内診	1.619
								糖尿病(主)	12 07 内診	1.619
								糖尿病(主)	12 08 内診	1.619
								糖尿病(主)	12 09 内診	1.619
								糖尿病(主)	12 10 内診	1.619
								糖尿病(主)	12 11 内診	1.619
								糖尿病(主)	12 12 内診	1.619
								糖尿病(主)	12 13 内診	1.619
								糖尿病(主)	12 14 内診	1.619
								糖尿病(主)	12 15 内診	1.619
								糖尿病(主)	12 16 内診	1.619
								糖尿病(主)	12 17 内診	1.619
								糖尿病(主)	12 18 内診	1.619
								糖尿病(主)	12 19 内診	1.619
								糖尿病(主)	12 20 内診	1.619

- 診療年月
- 医療機関コード
- 保険者番号
- 保険医療機関名
- 被保険者証記号番号
- 氏名
- 生年月日
- 性別
- 保険者番号、被保険者番号記号番号、生年月日、性別をハッシュ化
- 氏名、生年月日、性別をハッシュ化
- 傷病名 → 傷病行為コード、転帰
- 診療行為 → 診療行為コード  
12 再診  
13 医学管理  
14 在宅  
20 投薬 等
- 請求点数

郵便番号  
市町村コード

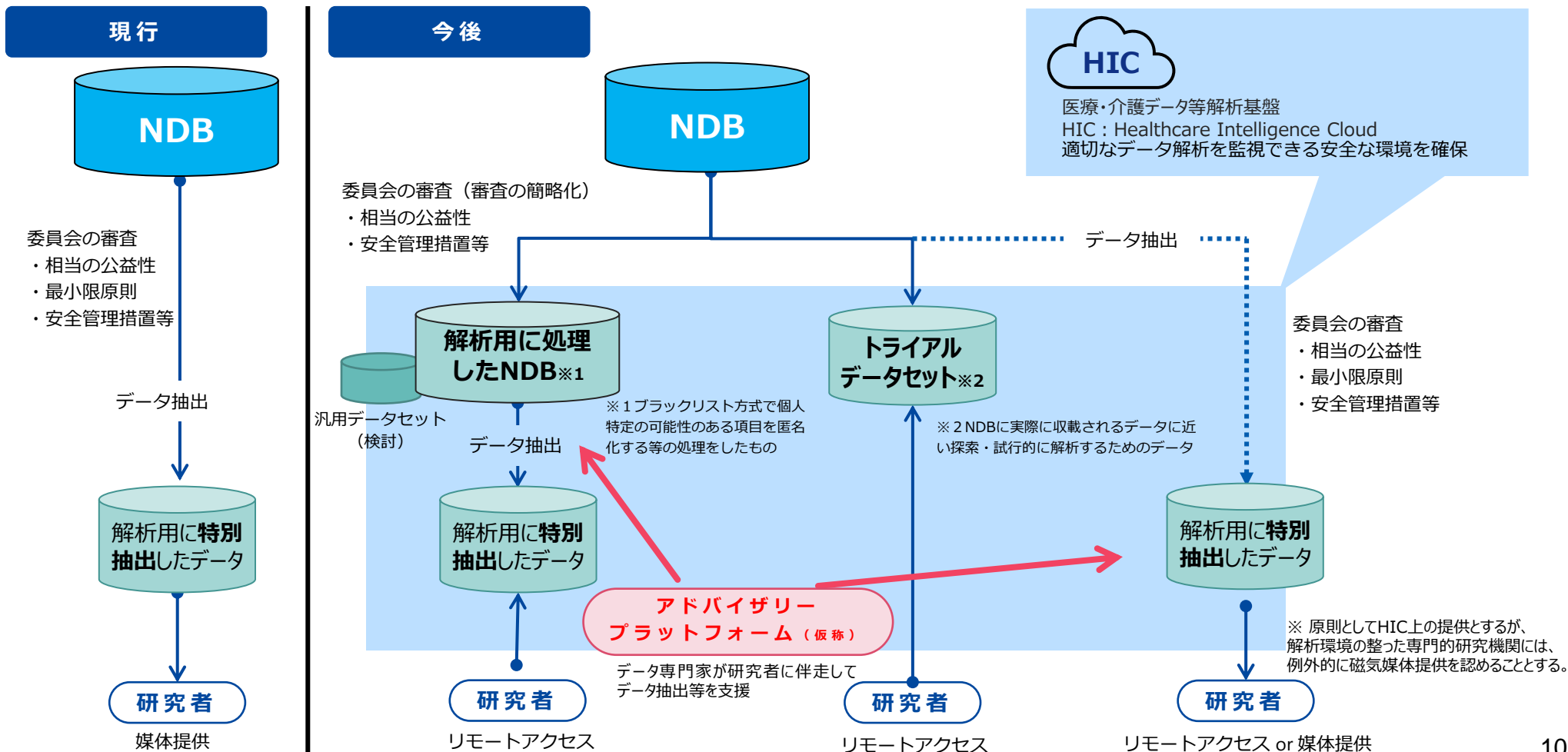
高額療養費  
自己負担限度額  
(適用区分)

公費負担医療  
レセプト  
公費負担者番号

レセプト情報がどのようにNDBに格納されているのかがイメージしにくく、利活用の障壁となっている

# 医療・介護データ等解析基盤（HIC）の機能拡充（案）

- ▶ クラウド環境で整備中の医療・介護データ等解析基盤（HIC）の機能を拡充し、解析用に処理したNDB（ブラックリスト方式で個人特定の可能性がある項目を匿名化する等の処理をしたもの）にリモートアクセスする仕組みを構築。
- ▶ NDBの解析用の処理内容、HIC上での適切なデータ解析を監視できる安全な環境の確保、研究者が負担するHIC利用料等について検討のうえ、システム改修を行い、令和6年秋に試行運用を開始



# 令和5年度厚生労働科学研究について

- NDBの利活用促進のために必要な機能、具備すべき安全管理措置、研究者への支援のあり方など技術的課題を検討するため、令和5年度中に、厚生労働科学研究事業「NDBの迅速提供に向けたスキーム再構築に資する研究」を実施。
- NDBオンサイトセンターの利用者、NDBで研究実績がある研究者、審査実務に精通した研究者で体制を構築し、レセプトデータの取扱経験のある者へのヒアリング等を通じて、データユーザーの観点、審査実務の観点からNDB利活用促進に向けた検討を進める。

## <主な検討事項>

以下について、ヒアリングやこれまで集積された知見等を踏まえ、具体的な提言を行う。

- HICにおける効率的なデータ提供のあり方
- 新たな提供体系において必要な安全管理措置
- データ活用を促進するための必要な支援のあり方

## <研究体制>

◎山本 隆一	医療情報システム開発センター
今中 雄一	京都大学
今村 知明	公立大学法人 奈良県立医科大学
松田 晋哉	産業医科大学
康永 秀生	東京大学
満武 巨裕	医療経済研究機構

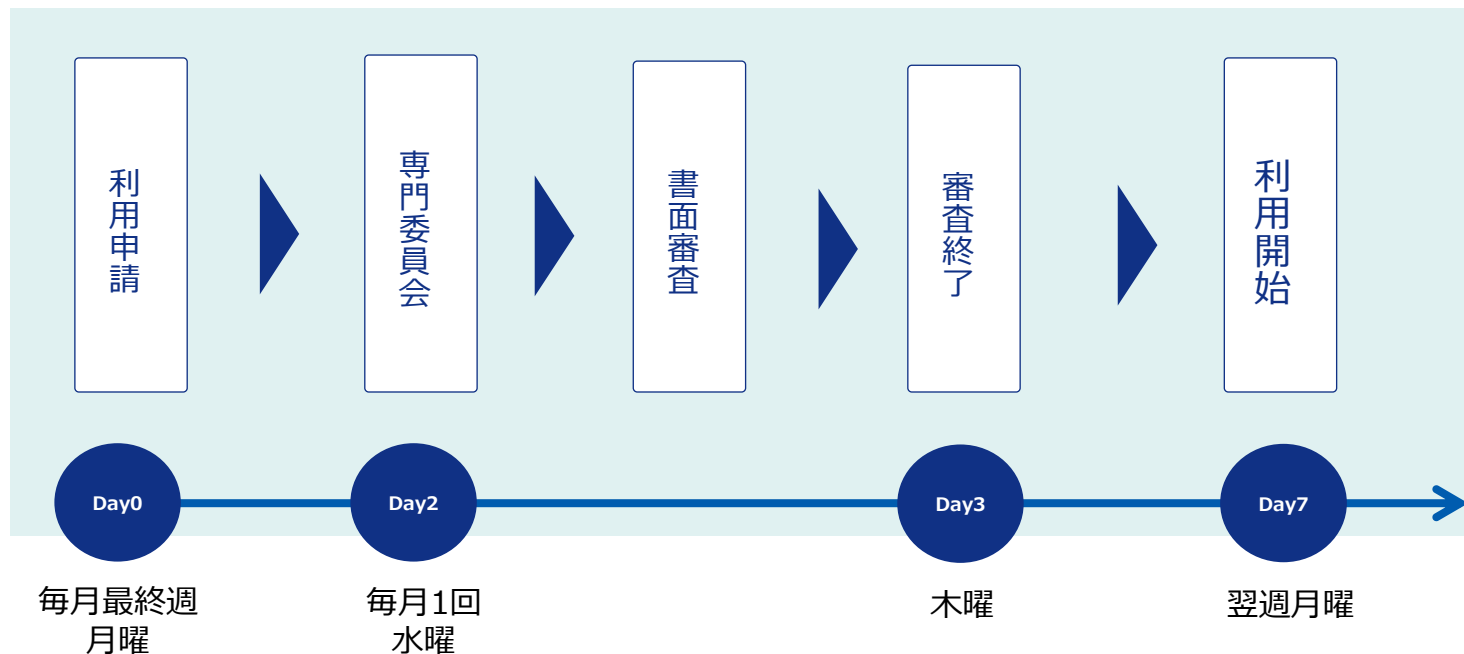
◎：研究代表

# HIC機能拡充に伴う審査方法の見直し（案）

- 申請からデータ提供まで平均390日の現状に対し、申請〆切を毎月設定し、申請から**最短7日**で処理  
※申請が月5件程度であることを踏まえ、当面月1回を設定するが、今後申請件数が増えれば複数回設定する

## 最短7日のイメージ※1

※1 研究者側の都合に要した期間は除く



- ポータルサイト上で申請

### 【承諾の場合】

- ポータルサイト上でID/PW提供
- 手数料の納付が必要

### 【不承諾・継続審査】

- ポータルサイト上で理由を通知
- 研究内容が不明確で判断できない場合等は対面審査

※ 例外的に磁気媒体の利用審査を行う場合は、対面審査

# (参考) NDB利用の安全管理措置

## 1 NDBの収載

- 厚生労働省は、高齢者の医療の確保に関する法律（以下「高確法」）に基づき、医療費適正化計画の作成・実施・評価に資するため、保険者等からデータの提供を受け、NDBに収載している。
- 厚生労働省は、医療費適正化計画の作成・実施・評価に資する目的で、自ら調査・分析を行うとともに、国民保健の向上に資する目的で相当の公益性を有する分析等を行う者に対して提供することができる。

## 2 NDBの安全管理措置

- NDBは、高確法に基づき、個人特定できないよう氏名等を削除し、匿名化した形でデータを収集している。
- NDB利用者は、高確法に基づき、他の情報と照合等の禁止義務、利用後のデータ消去、安全管理措置、不当な目的利用等の禁止などの義務が課されている。
  - ※ NDB利用の成果物は、公表前に、個人情報保護の観点から、希少疾患など個人が特定されないか（最小集計単位が10未満となっているか）を確認などを講じている
- 厚生労働省は、法令違反などの疑いがある場合には、高確法に基づく立入検査・是正命令を行うことが可能。
  - ※ 不当な目的利用等の禁止義務や是正命令に違反した者には、罰則あり（1年以下の懲役・50万円以下の罰金）
  - ※ このほか、NDBの不適切利用が発生した場合、厚生労働省は、ガイドライン・利用規約に基づく利用停止等の措置を行うことが可能