

## 2.HIC ポータル運用課題の实地検証と海外長期追跡サンプル データセットの比較調査

分担研究者 牧戸香詠子・東京大学大学院医学系研究科生物統計情報学講座・特任助教

分担研究者 明神大也・国立大学法人浜松医科大学・健康社会医学講座・准教授

分担研究者 森由希子・京都大学医学部附属病院 医療情報企画部・准教授

研究代表者 村松 圭司・産業医科大学 医学部・准教授

研究要旨：NDB 通年パネルデータセットのオンライン申請を可能とする HIC ポータルの運用実態と、海外における追跡可能サンプルデータセット提供事例を調査し、支援体制高度化の方向性を示すことを目的とした。2024 年 11 月に HIC ポータルを用いて利用申請と掲示板質問を実施し、操作性・応答時間・回答妥当性を検証した結果、一時保存不能、複数施設入力不可、質問割り振り機能の不透明さなどの課題を抽出した。併せて台湾 NHIRD (約 200 万人)、韓国 NHIS サンプル (約 100 万人)、英国 CPRD サンプル (人口の 0.02-0.13%、平均追跡 11-21 年) の提供条件を整理し、層別抽出設計、年間ライセンス制、IRB 要否などを比較した。これらの知見から、HIC ポータルには申請ドラフト機能、質問分類自動化、基礎集計資料拡充が求められ、年度追跡型サンプル構築においては代表性・コスト・追跡期間を考慮した段階的拡充が有効であると提言した。

### A. 研究目的

匿名医療保険等関連情報データベース (NDB) は、高齢者の医療の確保に関する法律第 16 条に基づき厚生労働省が公的医療保険レセプト情報等を掲載した国内最大規模のデータベースである。

厚生労働省は、平成 23 年に研究者等へ第三者提供が開始されて以降、介護保険総合データベース等との連結解析、死亡情報の掲載やクラウド解析環境「医療・介護データ等解析基盤 (HIC)」の開発等、利便性の向上を進めてきた。こうしたなか、令和 5 年度の厚生労働行政推進調査研究では、探索的な解析データセットとして縦断的追跡が可能な「通年パネルデータセット」の作成および分析の知見を集約した支援プラットフォーム

ムの構築が提案され、令和6年秋よりこれらの取り組みが本格的に開始された。

本研究では、通年パネルデータセットを用いた実地運用（申出・取得・解析・成果公表）を通じて、厚生労働省と研究者の双方の視点から、NDB提供体制における支援の在り方を検証する。本報告書は、NDBの利用に際してオンライン申請等を可能とするHICポータルへの運用上の課題を明らかにし、さらに海外における探索かつ追跡可能なデータセットの提供事例を調査することで、今後の在り方に関する知見を得ることを目指す。

## B. 研究方法

### B-1. 医療・介護データ等の解析基盤（HIC）ポータルに関連する課題について

HICポータルは、オンラインでNDBの申請手続きができる機能や申請内容やデータに関する疑問に対して、掲示板などを通じオンデマンドで回答が得られる仕組みが備えられている。本研究では、HICポータル機能の運用実態を確認するため、実際に令和6年11月に、HICポータルを用いたNDB（通年パネルデータセット）の利用申請および質問投稿を行い、申請手続きの操作性や掲示板による回答の応答・内容の妥当性を検証した。さらにNDB基礎資料等の集約機能の整備についても議論した。

### B-2. 海外における探索かつ追跡可能なデータセットの提供例について

HICの探索的解析環境に配置された通年パネルデータセットの試用を行ったが、2022年度分の単年データであった。今後、年度を跨ぐ通年パネルデータセットを作成するにあたり、海外の状況（台湾、韓国、英国、フィンランド）について主にWebベースで補足的に調査した。

## C. 研究結果

### C-1 HICポータルでのNDB申請における現状

まずHICポータルの利用時は、ユーザー登録が必要となるが、ユーザー側で登録した後に厚生労働省（研究班がユーザー登録実施した際は第三者提供窓口に委託）の承諾を受ける必要があるため、ユーザー登録に数時間～数日要する。

HICポータルでは、申出書（様式1）や運用管理規定（別添2）などを、システム上で直接入力する形式が採用されている。しかし、複数の利用施設がある場合でも、運用管理規定等は1施設分しか入力できない仕様であり、加えて一時保存機能が正常に動作しないとといったトラブルが生じた。また、申請が承認された後には、所属機関として正式な承認（決裁等）を得たうえで誓約書の提出が必要になるが、実務上は申出書等もあわせて決裁対象となることが多いため、申請書類の作成は従来通りに事前に行うか、もしくは書類として出力される機能があった方が利便性が高いと感じられた。さらに、申請は担当者が一括して行うため、同一申請に関与する他の取扱者が申請の進捗状況等を把握しづらく、後述するメッセージ機能による照会も担当者のみに通知される仕様であった。

上記以外にHICポータルを操作していた際に気になった点は以下である

- HIC ポータルログイン時に「当サイトのご利用時にはクッキーの使用が必要となります。」と、Cookie がカタカナ表記だった。(本質的ではない指摘)
- HIC ポータル上に、HIC 解析環境に関するマニュアルが掲載され、ダウンロードできる形になっていた。しかし一部抜粋のため HIC 解析環境の全体像が把握しづらく、電子証明書のインストール方法など HIC 解析環境の利用上、必須の箇所が抜けていて、結局 Workspaces 上にある完全版のマニュアルを見ないといけなかった。

## C-2 HIC ポータルでの申請やデータ等についての質問に関連する現状

HIC ポータルには、申請手続きやデータに関する質問を行うための機能として、「掲示板機能」と「メッセージ機能」の 2 種類が備わっている。「掲示板機能」は「コンテンツ機能」から利用可能で、投稿内容はすべての HIC ポータル利用者が閲覧できる。これにより、よくある質問への検索・参照が可能となり、共通の課題に関連する情報共有が効率的に行われた。一方、個別の申請内容等に関連する質問は「メッセージ機能」が用いられる。この機能は、回答が投稿した HIC ポータル利用者のみに表示される運用となっている。また、新たな回答が投稿された際には、登録されたメールアドレス宛に通知が届く仕組みが備わっている。しかし、掲示板とメッセージ機能の使い分けについて明確な記載がなく、利用者が混同する可能性があった。

質問に対する回答は、アドバイザープラットフォームの構成機関（支払基金、第三者提供窓口事業者、厚生労働省、科研班のメンバー）がそれぞれ得意とする分野に応じて行った。回答は、HIC ポータル利用者が、質問内容について分野を選択することで、ある程度割り振りされるが、実際の運用では分類が難しいケースも想定された。2024 年度において、本研究に関与する以外の者からの質問はなかったが、想定される質問内容は下記のような内容が考えられた。

- 別添 8 の項目についての説明が複数の資料に分散しており、どれを参照すればよいか判断が難しい。
- 長期入院、外泊、同日の入退院や転院、短期間での再入院など、入退院データの処理方法について知りたい。
- 医療費総額、自己負担額、公費負担額等の計算方法や判断基準について知りたい。
- DPC レセプトにおける SI/IY/TO レコードおよび CD レコードに重複があると聞いているが、どのように対応すべきか。

## C-3 HIC 上ポータル上に掲載する NDB 基礎資料の整備に関する課題

HIC ポータル上には、NDB の利活用を促進することを目的として、研究者等が参照可能な基礎資料を配置する予定である。これに伴い、既存の基礎資料については体系的に整理したうえで掲載を行うとともに、新たに必要性の高い基礎集計に関する情報の追加についても検討した。既存および新規に整備を検討している基礎集計

項目は、下記の内容の資料が考えられた。

- (1) データの粗集計に関する項目
  - レセプト等の格納件数
  - 医科レセプト等における ID5 の格納割合
  - 郵便番号および市区町村コードの格納割合
  - 限度額適用認定証適用区分の格納割合
- (2) データの突合割合に関する項目
  - 入院外レセプトと調剤レセプトとの突合割合
  - 各 ID と過去 1 年分のデータとの突合割合
  - 死亡情報との連結に係る検証結果
  - 他データベースとの ID 連結テストの結果

#### C-4 海外における通年パネルデータセットに類似したデータセットの提供例

##### C-4-1 台湾

- データベースの概要

台湾の医療保険制度は、1995 年に「全民健康保険」として単一保険制度が導入された。診療報酬制度を採用、レセプトの請求はオンライン化され、ほぼ 100%の普及率を誇る。<sup>1</sup>そのため、請求データに基づく研究用データベースである National Health Insurance Research Database (NHIRD)をはじめ、国民健康聞き取り調査など、約 70 種類の医療関連データベースがあり、衛生福利部 (Ministry of Health and Welfare: MOHW) が Health and Welfare Data (HWD) センターでこれらを統括・管理している。これらのデータベースは、個人識別番号を用いて相互に連結可能な設計となっている。<sup>1</sup>

- 通年パネルデータセットに類似したサンプリングデータ

台湾の個票データの解析は、サンプリングデータを含め、HWD センターや国立台湾大学、高雄医学大学等の指定施設に設置されたオンサイト環境に限定されている。研究者向けには、約 200 万人規模のサンプリングデータが提供されており、これらは年齢、性別、地域によって層別ランダム抽出され、1997 年から時系列データとして 2000 年、2005 年、2010 年に作成されている。

研究者が依頼した個票データの利用には施設内倫理審査委員会 (IRB) 承認が必要であるが、サンプリングデータの利用は IRB の承認が不要である。サンプリングデータセットが初めて作成されたのは 2000 年であり、当時は、100 万人規模および疾患特異別データセットを研究者向けに提供した。2018 年時点で、提供されているサンプリングデータセットは、2000 年から 2016 年の期間での傷病名、処方、検査、治療といった情報が含まれている。2005 年および 2010 年に作成されたデータセットは、2005 年から 2016 年まで、2010 年から 2016 年までの期間で追跡可能なデータセットも利用可能である。これらのサンプリングデータセットは、レセプト情報に加え、死因・

死亡情報、癌レジストリ、主要な疾患の情報や病院の情報も含まれる。含まれる患者数は下記の通りである。

起点となる年	全患者数	サンプリング数	割合	追跡期間
2000年	21,400,826	2,000,118	9.35%	2000–2016
2005年	22,314,647	2,000,120	8.96%	2000–2016, 2005–2016
2010年	23,074,487	2,000,126	8.67%	2000–2016, 2010–2016

※2018年時点

その他、2002年から2015年までの追跡可能な糖尿病疾患データセット等の13種類の疾患別データベースも提供している。これらのデータセットは、がん登録データや死因データと連結することも可能である。<sup>2</sup>

#### 【参考文献】

1. Hsieh, C. Y., Su, C. C., Shao, S. C., Sung, S. F., Lin, S. J., Kao Yang, Y. H., Lai, E. C.. Taiwan's National Health Insurance Research Database: past and future. *Clinical epidemiology*, 2019;11:349–358. <https://doi-org.utokyo.idm.oclc.org/10.2147/CLEP.S196293>
2. Lin LY, Warren-Gash C, Smeeth L, Chen PC. Data resource profile: the National Health Insurance Research Database (NHIRD). *Epidemiol Health*. 2018;40:e2018062. doi:10.4178/epih.e2018062

### C-4-2 韓国

#### • データベースの概要

韓国の医療保険制度は1963年に開始され、1999年に公的医療保険制度が整備されて2000年に単一保険者体制が導入された。保険者はNational Health Insurance Corporation (NHIC) と呼ばれ、公的医療保険であるNational Health Insurance Service (NHIS) を提供している。また、医療評価や医療費審査を担う独立組織としてHealth Insurance Review Agency (HIRA) が設立され、審査結果をNHISや医療機関に通知する役割を負っている。これにより、韓国には公的な医療データベースとしてNHIDとHIRAの2種類の医療データベースが存在している。<sup>1</sup>

#### • 通年パネルデータセットに類似したサンプルデータ

NHISでは、請求データ、健診データ、介護保険データ、出生・死亡・居住地情報、所得関連情報、障害登録情報など、さまざまな医療関連情報が収集されている。これらを基に、2006年を基準に選定された約4800万人の対象者が抽出され、2002年から2019年にかけて追跡されたデータのうち、2%に相当する約100万人分の性

別、5歳刻みの年齢階層、地域等で層別化ののち、ランダムサンプリングされたデータセットが作成されている。<sup>2,3</sup>

令和6年度に行ったNHICとのヒアリングでは、OPEN COHORTの概念で、2006年以降に生まれた新生児は新規対象者として追加される形でコホートを設計し、今後、対象者規模の拡大及び標本抽出基準の変更、セキュリティ問題の発生などによって対象者の変更が行われる可能性があるが、別途対象者を変更することなく、既存の対象者の最新情報を追加する方式で更新を進める予定とのことであった。<sup>3</sup>

NHISのデータへのアクセスおよび分析は指定された施設内でのみ可能だが、サンプルデータに関してはリモートアクセスが許可されており、使用期限は6ヶ月となっている。HIRAにもサンプルデータが存在するが、NHISのサンプルデータとは異なり、追跡可能なデータではない。

#### 【参考文献】

1. Park, J. S., & Lee, C. H. Clinical Study Using Healthcare Claims Database. *Journal of rheumatic diseases*, 2021;28(3):119–125. <https://doi-org.utokyo.idm.oclc.org/10.4078/jrd.2021.28.3.119>
2. Kim, M. K., Han, K., & Lee, S. H. Current Trends of Big Data Research Using the Korean National Health Information Database. *Diabetes & metabolism journal*, 2022;46(4):552–563. <https://doi-org.utokyo.idm.oclc.org/10.4093/dmj.2022.0193>
3. 令和5年度厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）NDBの迅速提供に向けたスキーム再構築に資する研究 報告書別紙3 韓国におけるデータ利活用に関するヒアリング調査 [https://mhlw-grants.niph.go.jp/system/files/report\\_pdf/202306007A-sonota03.pdf](https://mhlw-grants.niph.go.jp/system/files/report_pdf/202306007A-sonota03.pdf)

### C-4-3 英国

#### • データベースの概要

英国は、イングランド、北アイルランド、スコットランド、ウェールズの4か国から成る連合王国（UK: United Kingdom）であり、保健サービスの名称は、北アイルランド以外の3か国でNHS（National Health Service）、北アイルランドではHSC（Health and Social Care）と呼ばれている。<sup>1</sup>

研究目的で利用可能なデータベースは、NHSが運営する一般診療所（General practitioner: GP）の一次診療データが基本となっており、1993年にGeneral Practice Research Database（GPRD）が開始され、2012年にはClinical Practice Research Datalink（CPRD）として拡大された。CPRDでは、月次でデータ提供に同意した一般診療所から匿名化された電子健康記録データを定期的に収集しており、参加する診療所に登録された全患者のデータが含まれている。ただし、患者がデータ提供を希望しない場合、自ら依頼して提供を無効にすることができる。<sup>2</sup>

- 通年パネルデータセットに類似したサンプルデータセット

CPRD は、1987 年から、患者の基本情報 (ID、性別、出生年月、配偶者の有無、家族 ID 等)、診断、症状、処方、紹介状、予防接種、検査等の情報を収集している。平均フォローアップ期間は 5.1 年である。患者レベルのデータは第三者機関を介して入院データ、死亡・死因データ、がんや虚血性心疾患などの疾患レジストリ等とリンク可能である。<sup>2</sup>

CPRD には、GP システムの違いにより Aurum と GOLD の 2 種類があり、CPRD Aurum は GOLD より後の 2017 年に導入された。<sup>3</sup> 2021 年にリリースされた CPRD Aurum サンプルデータセットは、UK 全体の 0.02% に相当する 39,388 名の患者のデータを含み、転院、死亡した患者を除く 13,858 名の患者が追跡可能である。1995 年 1 月 1 日から収集されており、記録されている 13,858 名の平均追跡期間は 11.13 年である。<sup>4</sup> 同様に、2023 年にリリースされた CPRD GOLD サンプルデータセットは、0.13% に相当する 103,093 名の患者データが含まれ、91,780 名の患者が追跡可能である。このデータセットも、1995 年 1 月 1 日から収集され、91,780 名の平均追跡期間は 20.61 年である。<sup>5</sup>

CPRD Aurum および GOLD のサンプルデータセットの利用には、組織向けの 1 年間の使用ライセンス (Multi-study annual license: MSL) の契約があれば、追加の契約や利用料金は不要である。<sup>6</sup>

#### 【参考文献】

1. 地域における医療・介護の連携強化に関する調査研究 諸外国における健康・医療・介護分野のデータベースの現状調査報告書 平成 31 年 3 月
2. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827–36.
3. Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, Myles P. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol*. 2019;48(6):1740-1740g.
4. Medicines & Healthcare products Regulatory Agency. Release Notes: CPRD Aurum Sample Dataset October 2021 Available from URL: <https://www.cprd.com/sites/default/files/2022-02/CPRD%20Aurum%20Sample%20Dataset%20Release%20Notes.pdf> Accessed May 2025.
5. Medicines & Healthcare products Regulatory Agency. Release Notes: CPRD GOLD Sample Dataset April 2023 Available from URL: <https://www.cprd.com/sites/default/files/2023-05/Release%20notes%20CPRD%20GOLD%20Sample%20Dataset.pdf>

Accessed May 2025.

6. Medicines & Healthcare products Regulatory Agency. Pricing. Available from URL:  
<https://www.cprd.com/pricing> Accessed May 2025.

#### C-4-4 フィンランド

- データベースの概要

フィンランドは、長年にわたり体系的に医療データが収集・管理されている。主なデータソースとして Digital and Population Data Services Agency (DVV)、Cancer Registry、Finnish Centre for Pensions (ETK)、Kanta Services、Kela、Finnish Institute for Health and Welfare (THL)、Statistics Finland 等があり、利活用されている。<sup>1</sup>医療データの二次利用については、詳細なデータにもとづいた臨床研究を加速させる事で、様々な研究開発、教育、健康及び社会サービスの改善につなげていくことを目的として「医療保健情報の二次利用に関する法律 (Act on the secondary Use of Health and Social Data)」<sup>2,3</sup>が 2019 年に成立している。フィンランドには社会福祉、医療データの二次利用許可機関として Findata があり、研究者は Findata を通して、二次利用に関する許可申請を行い、審査を受けることができる。

- 通年データパネルセットに類似したサンプルデータセット

Findata は事前に処理された既製データセットを提供している。約 534 万人をベースパーソン (2010 年 1 月 1 日にフィンランドに居住していた者) とし、その家族を含め対象グループの規模は約 716 万 6000 人である。また、利用可能な期間は 60 年以上 (1953 年-2021 年) である。既製のデータセット利用では、抽出データセットに比較して、コスト見積もりや管理者からの抽出料金を必要としないため、より迅速に利用できるとされている。<sup>4</sup>既製のデータの使用許可料金は 300 ユーロである (2025 年 5 月時点)。<sup>4</sup>

#### 【参考文献】

1. FINDATA. Data. Available from URL:  
<https://findata.fi/en/data/#what-data-are-available-via-findata>  
Accessed May 2025.
2. Act on the secondary Use of Health and Social Data
3. 木村映善,大寺祥佑,佐々木香織,黒田知宏 フィンランドにおける医療分野レジスタとデータ提供の状況 日本統計学会誌 2020 年第 50 巻第 1 号 47-80
4. FINDATA. Ready-made datasets. Available from URL:  
<https://findata.fi/en/data/ready-made-datasets/#Available-datasets>  
Accessed May 2025.

## D. 考察

本研究にて、新たに導入された HIC ポータル上での申請およびメッセージ機能等を用いた質問を行った。現行の所属機関の決裁等手続きを鑑みると、申請書類の作成は、従来通り文書で作成する必要があった。また、HIC ポータルでの申請やデータ等について、質問できる掲示板およびメッセージ機能は、マニュアル等で両者の使い分けについて周知が必要と感じられた。

掲示板やメッセージ機能を通じて寄せられた質問には、アドバイザリープラットフォームの構成機関がそれぞれの担当分野に応じて回答することとなっているが、担当の割り振りについては難渋する例も考えられた。質問の件数が増加し、内容が多様化した時点で、より適切かつ効率的な対応ができるよう、質問内容の割り振り機能について再検討する必要があると感じられた。

通年パネルデータセットの年度の拡張においては、複数年の全データをサンプリングしてデータセットを作成するか、既存の通年パネルデータセットを元に、年度ごとに情報を追加する方法があった。今後、年度をまたぐ追跡性のある通年パネルデータセットを作成する際は、コスト、労力、利便性や代表性等を鑑みて、検討する必要があると感じられた。

## E. 結論

本研究にて、新たに導入された HIC ポータル上での運用および通年パネルデータセットの拡充についての知見を得られた。これらの知見は、提供体制の利便性の向上と NDB 利活用拡大に関連した参考事項となることが期待された。

## 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