

匿名医療保険等関連情報データベース（NDB）の第三者提供

よくあるご質問（FAQ）

2024年4月作成

⑫医療・介護データ等解析基盤（HIC）に関するご質問

（HIC上で利用可能なデータ）

1	Q	HICで利用可能なデータ形式は何があり、媒体提供とどう違うのか。
	A	2024年4月現在、HICでは①研究者の申出内容に従って任意のデータを抽出する「特別抽出」、②プリセットデータであるトライアルデータセット、の利用が可能です。ただし、HICのスペックの関係で取り扱えるデータ量の上限がございますので、特別抽出の規模によってはこれまで通り媒体でお渡しすることになる場合があります。集計表、サンプリングデータセットは従来通り媒体提供のみ行っています。また、従来通り、特別抽出の媒体提供を希望する申出も可能です。

2	Q	トライアルデータセットとサンプリングデータセットの違いは何か。
	A	トライアルデータセットは、サンプリングデータセットと同様に、各年1月、4月、7月、10月の医科・DPC・調剤レセプトから外来1%、入院10%程度を抽出し、機微な項目を除いたプリセットのデータセットです。ただし、トライアルデータセットは、2024年秋に簡易な審査による迅速な提供を可能とする予定のため、情報の安全性を高める観点から、低頻度レコード（サンプリング後のデータで算定回数が10回未満の傷病名、医薬品、診療行為）を含むレセプトについてレセプトごと削除する処理を行っています。2024年4月現在、トライアルデータセットはHIC上でのみ利用可能で、サンプリングデータセットは媒体提供でのみ利用可能です。

3	Q	現在NDBデータが提供され既に利用開始している場合（例、媒体提供済又はオンサイトリサーチセンターを利用中）、2024年秋に迅速提供が開始されるプリセットデータの利用に切り替えられるか。
	A	必要となる安全管理措置が異なることや、提供環境数に限りがあること、再度手数料が生じること等から、現在媒体提供しているお申出については、現在の利用形態で研究を完遂いただきますようお願いいたします。

4	Q	HIC上での他公的DBと連結は可能なのか。連結子はどうか。
	A	介護DBとNDB特別抽出の連結案件に限り、HIC上での提供が可能です。他の公的DBについては、2024年4月現在、HICによる連携はしておりませんので、これまで通り、各DBからのデータを媒体提供いたします。DB間を連結するIDにつきましては、現在と同様にID4又はID5が利用可能です。

（HICの環境）

5	Q	HICの解析環境のスペックはどの程度か。選択できるのか。
	A	2024年4月現在、特別抽出でご利用可能なHICの仮想サーバとして、以下の2種類を用意しています。 ① vCPU 8, メモリ 32GB, ストレージ 1000GB ② vCPU 16, メモリ 64GB, ストレージ 3000GB トライアルデータセットをご利用の場合は、①のみ選択可能です。

6	Q	HIC 上で利用可能な OS とデータベースソフトは何か。
	A	2024 年 4 月現在、HIC 環境で利用可能な OS は Linux(Ubuntu)と Windows です。データベースソフトについては、ミドルウェアは PostgreSQL が利用可能です。解析用 BI ツールは、トライアルデータセットの解析環境では R/RStudio のみ、特別抽出等の解析環境では SPSS、Stata、R/RStudio が利用可能です。BI ツールの SPSS/Stata については、どちらか一方のみ利用可能であり、ライセンス取得のため利用可能となるまで時間を要する可能性があります。

7	Q	HIC 環境に解析用の中間データベースを作成することはできるか。
	A	中間データベースの作成は可能ですが、HIC 環境にて一度に取り扱えるデータ量は仮想ストレージに依存しますので、3,000GB あるいは 1,000GB のうち選択した容量となります。

8	Q	HIC 環境において、解析用 SQL 以外のスクリプト (Python 等) を実行可能か。
	A	Python については導入されているため、スクリプトの持込手続きを経ることで実行可能ですが、スクリプト作成用の統合開発環境等の用意はございません。その他の実行環境の整備が必要となる言語についても、HIC 環境へのソフトウェアやミドルウェアの追加導入はできません。

9	Q	HIC 環境から中間生成物のデータの持ち出しやダウンロードはできるのか。
	A	NDB データ、中間生成物については持ち出しできません。HIC 環境から最終生成物の持ち出しを希望される場合、公表物確認を窓口に依頼してください。公表物確認を経た成果物は、持ち出し可能です。(※2024 年 4 月現在、ダウンロード機能は整備中のため、媒体格納して窓口より発送いたします。) なお、公表物の基準を満たさない中間生成物・最終生成物の持ち出しは認めておりません。なお、「成果物」「中間生成物」「最終生成物」の定義については、ガイドライン第 2 版の 3 ページをご参照願います。

10	Q	2 つ以上の申出に取扱者として登録した場合、HIC のアカウントは別々に払い出されるのか。
	A	その通りです。

(HIC の利用料金)

11	Q	HIC の利用料金はどの程度か。
	A	2024 年 4 月時点で利用できる環境の手数料の目安は以下の通りです。 <トライアルデータセットの解析環境> - 3 ヶ月間利用 60 万円~90 万円 - 6 ヶ月間利用 90 万円~120 万円 <特別抽出等の解析環境 (6 ヶ月間利用を想定) > - 初回申出/抽出条件の変更を伴う延長申出 150 万円~280 万円 - 延長申出 (HIC 利用期間のみ延長) 100 万円~200 万円 HIC (クラウド) 上の解析環境をご提供するため、自研究室でのサーバ等機器購入、維持費用は不要となります一方で、これまで通りの窓口事務手続きの費用に加え、HIC (クラウド) 利用に係る費用が利用期間に応じ別途発生いたします。2024 年秋の HIC 拡充に合わせて、再度利用料金の見直しが行われる予定です。

(2024 年秋以降に予定している提供体制)

12	Q	2024 年秋に HIC 上で迅速に利用可能となるデータセットは何か。特別抽出データは迅速提供されるか。
	A	2024 年秋から、プリセットのデータセットであるトライアルデータセットや通年パネルデータセットに係る探索的利用環境については原則 7 日で利用可能となることを目指しています。一方、研究内容に応じて抽出や加工が必要な特別抽出データについては、引き続き抽出条件等を提出していただくためのお時間が必要であり、7 日で利用可能となる予定はありません。

13	Q	トライアルデータセットと通年パネルデータセットはどんなデータセットか。これらの共通点と異なる点は何か。
	A	どちらも全レセプトからサンプリングを行ったプリセットデータで、2024 年秋からは HIC 上の探索的利用環境において迅速に利用開始できるようになる予定です。 トライアルデータセットの内容については Q2 をご参照ください。 通年パネルデータセットは、1～数年分のレセプトからサンプリングされた ID に紐づく医科・DPC・歯科・調剤レセプトと特定健診情報・保健指導情報をまとめたデータセットで、ID を用いることでコホートを追跡する縦断的な分析が可能です。また、機微なコードの一部（地域に係る情報、機関を表す匿名化コード等）が一定の粒度で格納されています。当初は ID 5 が利用可能な 2022 年度以降のデータセットのみが提供可能となる予定ですが、順次拡大を検討しています。

14	Q	現在 NDB データの利用が既に承諾されている状況でデータが未提供の場合（例：媒体提供に係る抽出条件調整～提供待ちの状況）、2024 年秋に迅速提供が開始されるプリセットデータの利用に切り替えられるか。又は、特別抽出したデータを媒体提供では無く HIC 上で利用することは可能か。
	A	迅速提供が可能となる探索的利用環境のデータセットについては、利用するデータの内容や申出書類が大きく異なるため、別途新規申出をお願いいたします。その際、既存の特別抽出のお申出は手続中止としても抽出継続としても構いません。

15	Q	探索的利用環境の利用を経ずに HIC 用 NDB や特別抽出に係る利用申出を行うことは可能か。その場合は迅速に利用可能なのか。
	A	まず探索的利用環境に格納されたプリセットデータを用いることで、NDB データの構造を把握したうえで、詳細な研究計画を立案してスムーズに大規模な NDB データを用いた研究を実施されることを推奨しております。一方で、探索的解析環境の利用を経ずに、より多くのデータを用いる HIC 用 NDB（現在のオンサイトリサーチセンターのように、全量データを参照できる利用方法）や特別抽出に係る利用申出を行うことも可能です。それらの申出については、現在と同様に抽出条件等の詳細な研究計画を提出していただく必要があり、7 日で利用可能となる予定はありません。