



リアルワールドデータ(RWD)研究のための データベース選定ガイド

YUJ 横浜市立大学

滋賀医科大学
SHIGA UNIVERSITY OF MEDICAL SCIENCE

北里大学
KITASATO UNIVERSITY

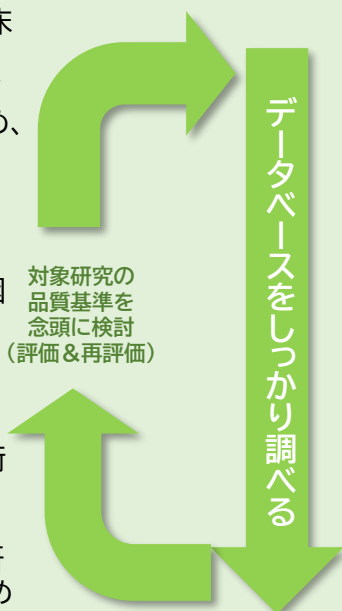
本ガイドは、リアルワールドデータを用いた研究(RWD研究)を実施する予定の研究者に、データベース選定プロセスの概観を理解いただくためのガイダンス資料となります。RWD研究を行うのが初めての方でも理解しやすいように、RWD研究の基本的な考え方や主要なデータベース事例にも触れています。

はじめに RWD研究の特徴を確認しましょう

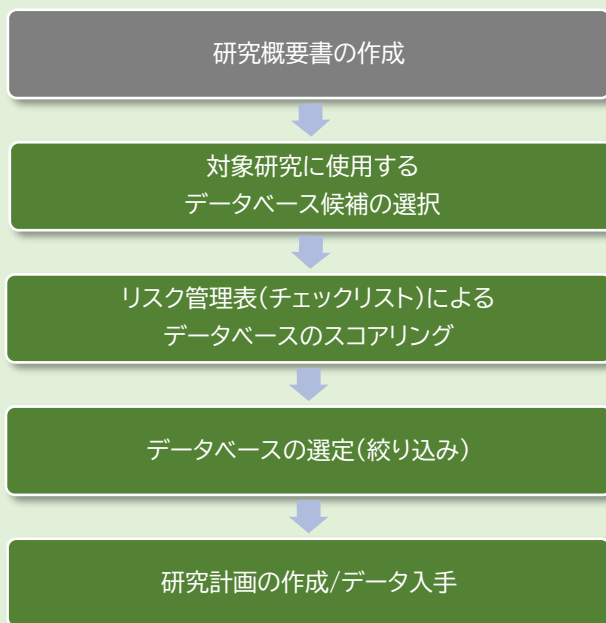
RWD研究は通常の臨床研究とは異なり、個々の研究を目的としてデータが収集されていないため、制限があります。

- ① 研究者が元データを再確認することや追加での情報収集が困難である
- ② データベース毎にデータ定義が異なる
- ③ データの生成背景の理解や前処理の技術が必要である

データベースの選定と研究計画立案を両輪で進める必要があります。



対象研究の
品質基準を
念頭に検討
(評価&再評価)



事前準備① 対象研究の大枠を確認しましょう

対象研究の内容や目標(品質方針)によって選ぶべきデータベースが異なってきます。データベース選定に関係する内容について、まず、研究チーム内で確認するようにしましょう。

研究の大枠 : 記載例

【研究課題名】

指定難病Aの医療連携課題の解決を目標とした疫学調査

【目的】

増加する指定難病Aの適正な医療配分を目指し、リアルワールドデータを用いた疫学調査を実施する

【本研究が最終的に目指すもの】

学会発表、論文作成

指定難病Aのリアルワールドにおける診療を記述し、各医療機関の強みを生かした地域ネットワークを構築するための基礎資料を提供

地域ネットワークモデルの作成

地域難病対策課へのネットワークモデルの提言

明らかとなった基礎データを用いて、指定難病Aの診療を基幹病院と診療所、地域中核病院で連携して実施するために紹介・逆紹介の実臨床へ応用可能な判断基準を作成

【研究概要】

対象者: 日本で指定難病Aと確定診断された患者(全年齢)

研究デザイン: 観察研究(横断研究)

評価項目: 専門機関病院へ通院する患者数
医療機関へ通院する患者数・・・ etc

チェックすべきポイント

☑品質方針

RBA^{※1}を実装する前に、本研究が目標とするゴール(学会発表、英語論文作成、ガイドライン改訂、薬事承認など)を明確にしましょう。

☑評価項目

データベース内の項目を用いて、クリニカルクエスション(CQ)を解決できるリサーチクエスション(RQ)に変換しましょう。

※1 Risk Based Approach

事前準備②

RWD研究でのデータベース選定の重要性を知りましょう

RWD研究において、データベース選定は研究実施に非常に大きな影響を与えます。事前に、主なポイントを確認しましょう。

※詳細は、「研修動画(入門編)」をご覧ください。

全編はこちらから



リアルワールドデータ(RWD)研究のための
QbD(Quality by Design)
RBA(Risk Based Approach)



RWD研究の目的と適正実施の必要性

リアルワールドデータ研究の目的は、疾病や疫学の理解、診断や治療の評価、健康被害リスクの早期発見、医療政策の評価や対策などが挙げられます。最近では臨床開発や承認申請に使用する事例も増加しています。このように、リアルワールドデータ研究においても品質目標に応じたリスクの評価や管理が必要となってきています。




- 疾病の理解
- 疫学の理解
- 診断・治療の評価や健康被害リスクの早期発見
- 医療政策の評価・対策

臨床開発・承認申請に使用する事例も増加

RWD研究においてもリスク評価・管理が必要

RWD研究で使用されるデータベース

臨床医学系のリアルワールドデータの例には、大きく3種類(保険請求データ、電子カルテデータ、患者レジストリ)に分けられます。保険請求データの中にDPC、特定健診、レセプトデータベースがあり、これらは患者さんが受けた保険診療に関する請求やDPC、レセプト、健康診断等のデータが含まれます。それぞれのデータベースに含まれるデータや申請条件などが異なるため、データ申請前にデータベースを詳細に調べる必要があります。

保険請求(DPC、特定健診、レセプトデータベース)	電子カルテ	患者レジストリ
 患者が受けた保険診療に関する請求、DPC、レセプト、健康診断等のデータを収集したデータベース	 複数の医療機関からレセプト、DPCに加えて、電子カルテデータを収集したデータベース	 特定の疾患患者の詳細データを複数の医療機関から収集し登録したデータベース
【例】 <ul style="list-style-type: none">● NDB (レセプト情報・特定健診等情報データベース)● KDB (国保データベース)● 介護DB (匿名介護情報等)	【例】 <ul style="list-style-type: none">● MID-NET● NCDA(診療情報集積基盤)	【例】 <ul style="list-style-type: none">● 全国がん登録● 指定難病患者データベース● NCD
<ul style="list-style-type: none">● レセプト傷病情報● レセプト医学管理料情報● レセプト手術情報● レセプト診療行為情報● レセプト特定器材情報● レセプト医薬品情報	<ul style="list-style-type: none">● 患者情報● 来院等情報(外来、入院、退院)● 傷病情報(診断コード、病名コード)● 処方情報(処方箋)● 注射情報(処方情報、生薬、補助剤など)● 検査情報(検体、検査種別、生薬、補助剤など)	<ul style="list-style-type: none">● 好発年齢● 疾患の症状・合併症・死亡● 主観的健康観● 医療費● 福祉サービス利用状況● 医薬品の市販後調査 など

臨床医学系RWD研究で考慮すべきポイント

臨床医学系リアルワールドデータ研究では、選定するデータベース(規模、項目、形式、質)、データベースの入手可能性(申請要件、審査の有無、費用、スケジュール)、実現可能な解析体制の構築(データクリーニング、前処理が必要か等)によって、対応可能なRQ(リサーチクエスチョン)や評価項目がデータにより制約されてしまう可能性があります。

1 選定するデータベース

- ✓ データベースの規模
- ✓ データ項目
- ✓ データ形式
- ✓ データの質 など

2 データベースの入手可能性

- ✓ 申請要件
- ✓ 審査の有無
- ✓ 費用
- ✓ スケジュール など

3 実現可能な解析体制の構築

- ✓ データ抽出
- ✓ データクリーニング
- ✓ データ前処理 など

対応可能なRQ(research question)や評価項目がデータにより制約される可能性

研究の目的や取得したいデータ内容などに従って使用するデータベース候補を絞り込んでみましょう。

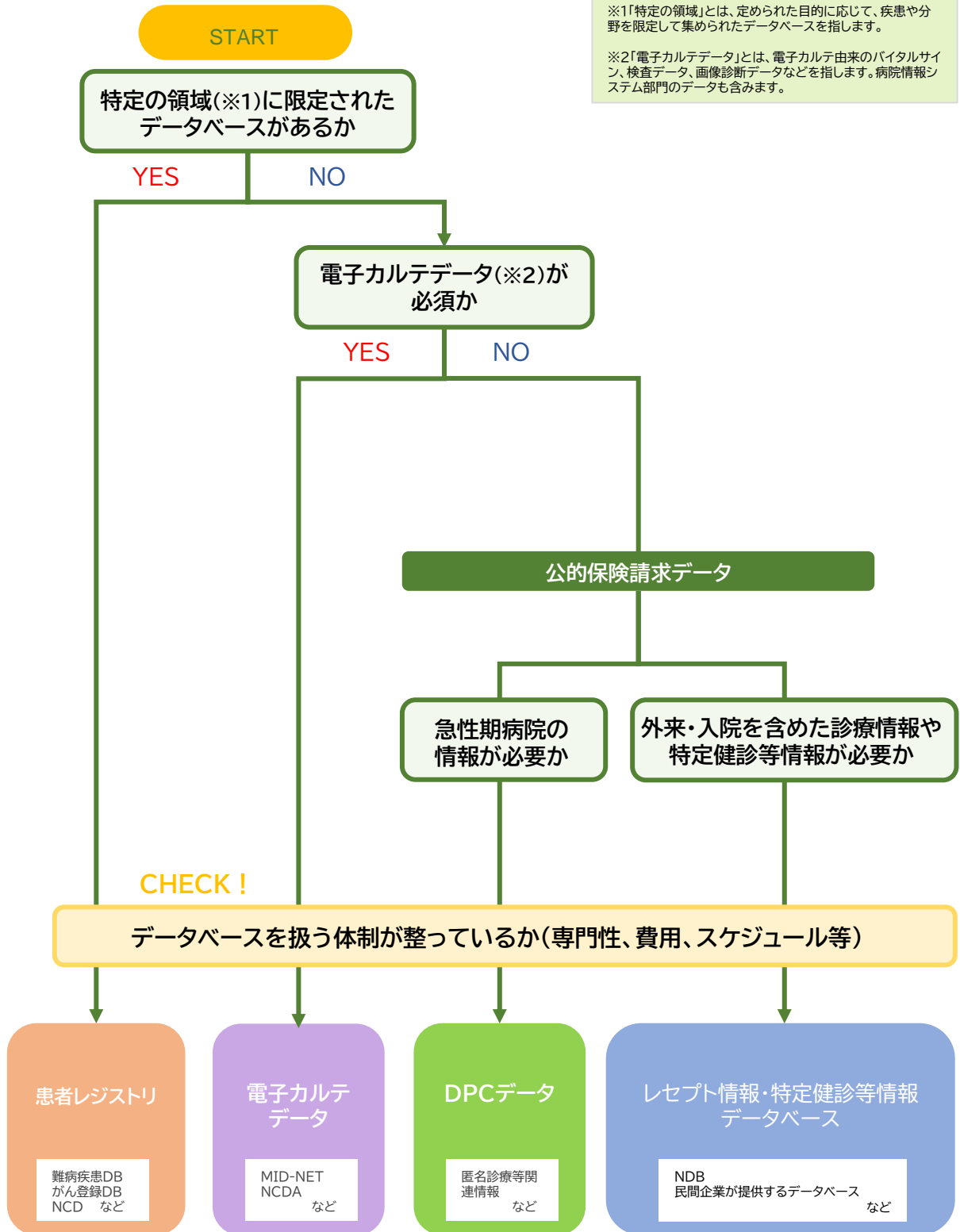
【使用上の注意】

本フローチャートは、各データベースの特徴を理解するために作成したものです。データベースによっては複数カテゴリーの情報を有しているもの等があります。また、日本国内のデータベースの使用を想定しておりますことご注意ください。

医療データベースの選定

※1「特定の領域」とは、定められた目的に応じて、疾患や分野を限定して集められたデータベースを指します。

※2「電子カルテデータ」とは、電子カルテ由来のバイタルサイン、検査データ、画像診断データなどを指します。病院情報システム部門のデータも含まれます。



診療所診療に
関するRWD
研究で使用する
データベースは？

外来診療データを
有する民間保険請求
データベース等を
利用する傾向が
あります

かかりつけ医に関する制度整備や訪問診療の推進に伴って病院だけでなく、診療所を対象としたRWD研究を実施する事例も増えてきています。

外来診療の評価のためには、病院診療に関するRWD研究とは異なり、個人単位の外来診療の利用状況の追跡が必要であることが多く、保険者ベースで集められた保険請求データを利用した分析等が行われています。民間企業が提供するデータベース(JMDC Clinical Databaseなど)を利用して分析するケースが多くなっています。

ただし、現在のデータベースではカルテ情報が含まれず病名をレセプトや投薬から定義するという点で限界があり、研究の目的や内容によって選択するデータベースが変わってきます。必ず、データベースに関する最新の情報を入手した上で研究目的やスケジュール等に合わせてデータを入手できるか確かめてから利用するようにしましょう。



データベースを選ぶときには、想定している研究チーム内でデータの品質やセキュリティを担保できる専門人員やデータの保管・解析に関する環境が揃っているかチェックするようにしましょう。

●データ解析体制のチェックポイント

- データの品質方針(論文投稿、薬事申請等)にあったガイドラインをクリアできるか
- データ抽出依頼から入手までの時間を考慮した研究期間を設定できるか
- 分析の目的に則したデータの前処理ができるか

一般的に、有償データベースはデータ分析を考慮したデータの前処理が行われていることが多く比較的容易に解析などができるものの、レセプトデータを分析する際には、日本独自の診療報酬制度の理解やデータ構造の理解も重要となります。

※RWD研究に関するガイドラインについては本ガイドp11に詳しく記載しています



データ解析の
体制整備の
基準は？

研究の品質方針と
データベース構築時の
データ品質担保の
プロセスや基準に
よって変わります

データベース選定のポイント

データベース取得費用がしばしば問題になります。また、研究スケジュール、利用者の条件、手続きの煩雑さ、データ前処理の難しさが障壁となることもあり、データベース選定の際には複数の観点から条件を確認することが必要です。

●公的データベース

データ取得費用は、無料または格安のものから民間データベースと変わらないものまで多岐に渡ります。一般的に、公的データベースは民間と比較し申請手続きが煩雑で条件を設けて利用者を制限している場合もありますので、データベース選定の際には確認が必要です。

- 費用が高額：MID-NET
- 手続きが煩雑：NDB、KDB
- 一部の研究者に利用を制限：
学会主導の患者レジストリなど

データベースの選定には総合的な判断が必要

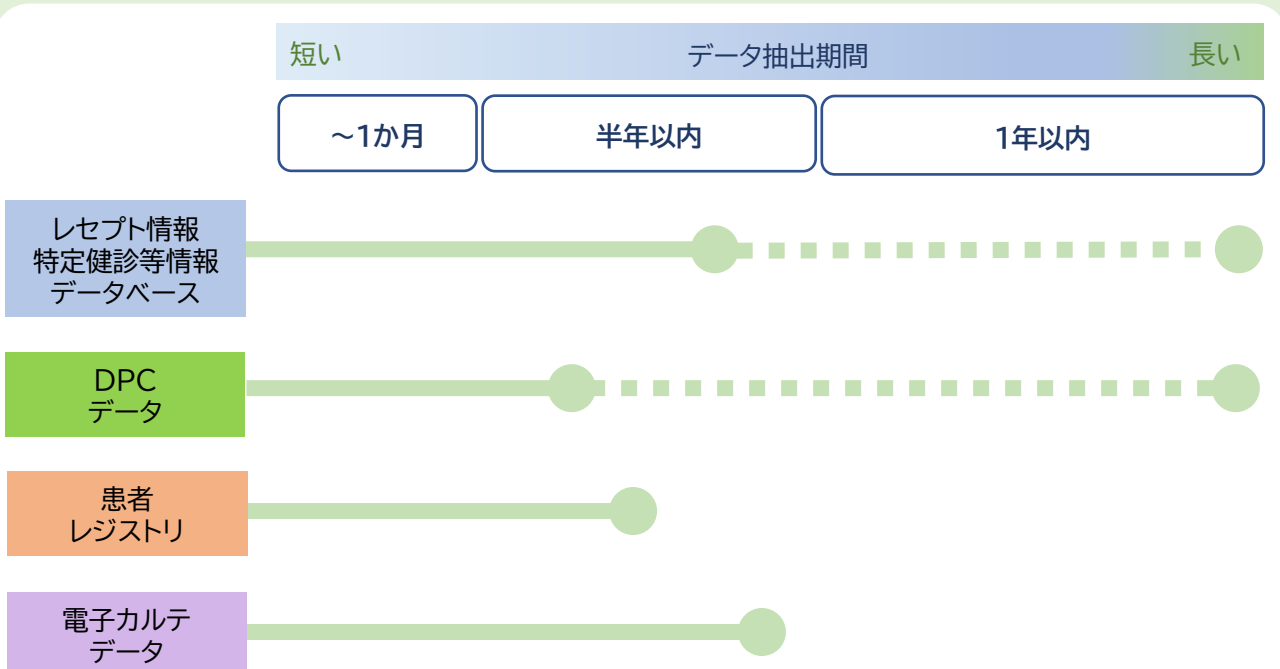


●民間データベース

データ取得に費用はかかりますが、データ取得までの期間が早く、前処理などの煩雑さも少ないという利点があります。一方、加工されて提供されている変数、あるいは、元データには含まれるものの提供されない変数もあるため、データベースに含まれる対象集団の特性なども考慮して、総合的に判断する必要があります。

(参考)データ取得までの時間目安

一般的なデータベースのデータ取得までの期間は、3-5か月程度ですが、民間保険請求データはすぐに入手可能です。入手する内容によっては半年以上かかってしまう場合もあります。研究実施期間に合わせて、適切なデータベースを選ぶことが重要となります。



患者レジストリ

- 【定義】** 特定の疾患を持つ患者の詳細なデータを、多くの医療施設から収集・登録したデータベース。罹患率や有病割合を調べたり、疾患の経過や予後を把握したりするために収集されるデータ収集後に、分析テーマが設定される場合もある
- 【特徴】** 特定の疾患の詳細な患者情報をもつ
Clinical Innovation Network: CIN(<https://cinc.ncgm.go.jp/>)から検索が可能
- 【使用例(研究)】** 例1) 手術方法Bのリスクモデルの作成
例2) 感染症パンデミックが外科的処置に与えた影響の検討

		指定難病 データベース	National Clinical Database (NCD)	がん登録
管轄		厚生労働省	一般社団法人 National Clinical Database	国立研究開発法人 国立がん研究センター
データ種別		患者レジストリ(国)	患者レジストリ(学会)	患者レジストリ(国)
総登録者数		不明	1200万人 (手術については全国の手術の95%以上)	年間100万件の登録
対象患者		各疾患の特定医療費 受給者認定者	外科系学会15学会 での登録患者	がんと診断された 全患者
データ 範囲	カルテ	-	-	-
	(DPC)	-	-	-
	レセプト	-	-	-
	その他	臨床調査個人票、 医療意見書	病名、術式等 (学会毎に異なる)	病名、進行度、治療、 診断の経緯、死亡日等
データ抽出が 完了するまでの期間		最短4か月	各学会に問い合わせ	最短2～3か月
データ抽出者が 満たすべき要件		厚労省・文部科学省等の 科研費補助を受ける者	各学会に問い合わせ	国、都道府県の関係者、 「がんに関する調査研究を 行う者(条件あり)」
事前の倫理審査の要否		要	要	要
費用		無料	各学会に問い合わせ	有料
特徴		・疾患に特化した重症度、 治療内容、検査値などが 含まれる ・軽症者は含まれない	・各疾患に特有の評価項目がある ・診療科を横断する研究には 不向きである ・運用が各学会に委ねられている ・各領域の主要医療機関の データはほぼ網羅されている	・網羅性が高い(※) ※法律に基づくレジストリであるため登録に強制力が伴う

各疾患特有の
評価項目が
登録されている

データ抽出
依頼は限定的
かつ
手間がかかる

電子カルテデータ

- 【定義】** レセプトやDPCデータに含まれる患者背景および治療の履歴データに加えて、電子カルテ由来のデータを統合したデータベース
- 【特徴】** 電子カルテにある各種データを活用することにより、研究の対象集団、暴露要因、交絡要因、アウトカムなどを正確に定義できる
 全国規模の保険データベースよりは小規模である
 病院間でデータ登録基準に差があり、国の事業等で同質性担保にかかる対応がなされている
- 【使用例(研究)】** 例1)MID-NET®を用いたワクチンCに関する安全性プロファイル等の評価
 例2)疾患D患者の予後を含めた実態を把握するための研究

		Medical Information Database Network (MID-NET)	診療情報集積基盤 (NCDA・MIA)
管轄		独立行政法人 医薬品医療機器総合機構	独立行政法人 国立病院機構
データ種別		電子カルテを含むDB	電子カルテを含むDB
総登録者数		570万人 (2021年12月時点)	260万人 (2019年度末時点)
対象患者		協力医療機関(23施設)を受診した患者	NCDA参加病院(66病院)を受診した患者
データ範囲	カルテ	●	●(NCDA)
	(DPC)	●	●(MIA)
	レセプト	●	●(MIA)
	その他		
データ抽出が完了するまでの期間		約5か月	不明
データ抽出者が満たすべき要件		厚労省や医薬品製造販売承認を目的とした利用 公的研究費による研究者	原則、国立病院機構の職員が含まれる研究チーム (競争的資金利用)
事前の倫理審査の要否		要	要
費用		有料 1000~2100万円 調査 2年以内	有料 集計値1つあたり100円 +作業手数料10万円 個票利用で別途規定
特徴		・費用がかかる ・手続きが煩雑である ・一部の作業(データ処理等)はオンサイトで対応する必要がある	・手続きが煩雑である ・一部の作業(データ処理等)はオンサイトで対応する必要がある

電子カルテ由来の診断・検査データ(バイタル、画像、血液検査等)から研究の対象集団・暴露要因・交絡要因・アウトカムを正確に定義できる

公的な大規模研究で主に利用

DPCデータ

【定義】 日本独自に開発された診断群分類DPC(Diagnosis Procedure Combination)を用いて急性期医療を担う医療機関等の診療情報を分類・収集したもの

【特徴】 診療プロセス・データに加えて、いくつかの臨床情報が含まれている

【使用例(研究)】 例1) きわめて稀で重篤な合併症Eの発生割合の検討
例2) 急性期疾患Fに対する薬剤Gの有効性の検討

		匿名診療等関連情報 (IDPCデータ)
管轄		厚生労働省
データ種別		DPCデータ
総登録者数		不明 (全体の約64%の病院)
対象患者		DPC病院の外来・入院患者
データ 範囲	カルテ	
	(DPC)	●
	レセプト	
	その他	
データ抽出が完了するまでの期間		最短3-4か月
データ抽出者が満たすべき要件		公的機関、アカデミア、民間事業者 公的研究費による研究者
事前の倫理審査の要否		要
費用		有料(免除の規定はあり)
特徴		・公的資金獲得した場合には費用の免除規定がある ・診療情報(※)が含まれる ・急性期病院のデータのみ含む

レセプトデータには含まれない
各種臨床情報も含まれている

使用条件が厳しい
(限られた者しか利用できない)

※DPCデータで得られる診療情報
【様式1】

- ①入退院情報
- ②患者背景:身長、体重 等
- ③診断名
- ④手術情報
- ⑤診療情報

レセプト情報・特定健診等情報データベース

- 【定義】 保険診療を行った医療機関や介護福祉サービスを行った事業者が、保険者に診療報酬を請求するための明細書情報(レセプト情報)、及び、特定健診・特定保健指導情報、要介護認定情報を格納したデータベース
- 【特徴】 国が管理し、悉皆性が高い
患者単位で突合し、データの縦断化が可能である
- 【使用例(研究)】 例) 周術期口腔機能管理による術後肺炎発症予防効果の検証

		匿名レセプト情報・匿名特定健診等情報(NDB)	匿名介護情報等(介護DB)
管轄		厚生労働省	厚生労働省
データ種別		公的保険請求データ(特定健診情報)	公的保険請求データ(要介護認定情報)
総登録者数		日本の総人口にほぼ近い	約140万人(2019年度)
対象患者		医療保険を使用した患者 特定健診を受診した患者	国民健康保険を使用して介護のサービスを受けた患者(都道府県・市町村及び国民健康保険組合を介して)
データ範囲	カルテ	-	-
	(DPC)	-	-
	レセプト	●	●
	その他	特定健診データ	要介護認定情報、匿名介護レセプト等情報、匿名LIFE情報のデータ
データ抽出が完了するまでの期間		最短3-4か月(データ抽出条件により1年程度の可能性あり)	最短4か月
データ抽出者が満たすべき要件		公的機関、大学その他の研究機関、民間事業者、補助金等による業務を行う個人	公的機関、アカデミア、補助金等による研究を実施する民間事業者等
事前の倫理審査の要否		要	特別抽出依頼時のみ要
費用		有料(免除の規定はあり)	有料(免除の規定はあり)
特徴		・悉皆性に富み、保険者が変わっても個人の追跡が可能である	・データ抽出方法の選択肢が3パターンある(特別抽出、集計表情報、サンプリングデータセット) ・NDBとの連結が必要な時はNDB側への申請も必要である

限られた者しか
利用できず
抽出手続きも煩雑

2022年より
国が保有するNDB・介護DB・DPC DBを連結したデータも提供開始

民間の保険請求データベース

データ利活用が一層進んでおり、民間企業が有償で提供する商用のデータベースもあります。データをアカデミアのみならず民間企業も利用可能である、早く入手できる等の特徴があります。研究目的や実施体制、スケジュール等にあわせて使い分けましょう。

		JMDC Claims Database	統合医療データベース Cross Fact
管轄		(株)JMDC	(株)インテージ リアルワールド
データ種別		民間保険請求データ	民間保険請求データ
総登録者数		約1000万人強	700万人以上
対象患者		JMDC社が契約した 医療機関or保険者(健康保険)	社会保険の加入者 院外調剤を受けた患者
データ 範囲	カルテ	●	-
	(DPC)	●	●
	レセプト	●	●(社保、院外調剤)
	その他	一部検査値	特定検診等データ 医薬品マスター
データ抽出が 完了するまでの期間		契約から2週間	契約から約10営業日
データ抽出者が 満たすべき要件		-	-
事前の倫理審査の要否		不要	不要
費用		有料	有料

他のデータベース
よりも早く簡便に
入手できる

RWD研究特有のガイドラインを確認しましょう

RWD研究に特化した研究ガイドラインがいくつか出ています。論文タイトル作成から考察まで論文作成の各段階におけるチェックリストが記載されていますので、論文を書き始める前に抑えるべきポイントを確認するようにしましょう。

■ 観察研究論文報告の質を改善するためのガイドライン(STROBE声明)
(Strengthen the Reporting of Observational Studies in Epidemiology)
<https://www.strobe-statement.org/>

■ RWDを用いた研究論文報告の質を改善するためのガイドライン(RECORD声明)
(REporting of studies Conducted using Observational Routinely collected Data)
<https://www.record-statement.org/>

■ RWDを用いた治療効果検討の再現性向上のためのプロトコルテンプレート
(HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force)
<https://onlinelibrary.wiley.com/doi/full/10.1002/pds.5507>



日本の保険医療制度の特徴を説明しましょう

RWD研究で使用される各種データベースは、日本独自の保険医療制度を基にしているものが多くあります。論文等を発表する際に保険医療制度の全体像を英語でしっかりと説明するようにしましょう。

たとえば、DPC(Diagnosis Procedure Combination)は、診療情報標準化と透明化を目的として開発された日本独自の患者分類法です。DPCを用いた支払制度は、日本の大部分の急性期医療機関で適用されているDPCデータの集積と分析によって、医療の質の評価、病院経営指標の作成や説明責任の強化、医療制度の適正化など、医療サービスの向上のために広く活用されています。

(参考文献) ※和訳及び一部抜粋

N Nakashima et al., Epidemiologic Research on Real-World Medical Data in Japan Volume1

K.Hayashida et al.,History and Profile of Diagnosis Procedure Combination(DPC):Development of a Real Data Collection System for Acute Inpatient Care in Japan.J Epidemiol.2021



【書籍】

1. 康永秀生, 田上隆, 大野幸子, 超入門! スラスラわかる リアルワールドデータで臨床研究: 金芳堂; 2020.
2. 康永秀生. 医療+ヘルスケアのためのリアルワールドデータ活用: 中央経済社; 2020.
3. Naoki Nakashima, *Epidemiologic Research on Real-World Medical Data in Japan, Volume1,2*: Springer; 2020.

【文献】

1. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sorensen HT, von Elm E, Langan SM and Committee RW, The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* 2015;12:e1001885.
2. K.Hayashida et al., History and Profile of Diagnosis Procedure Combination (DPC): Development of a Real Data Collection System for Acute Inpatient Care in Japan. *J Epidemiol*; 2021

【公開情報】

<研究ガイドライン>

- STROBE声明: 観察研究論文報告の質を改善するためのガイドライン (Strengthen the Reporting of Observational Studies in Epidemiology)
<https://www.strobe-statement.org/>
- RECORD声明: RWDを用いた研究論文報告の質を改善するためのガイドライン (REporting of studies Conducted using Observational Routinely collected Data)
<https://www.record-statement.org/>
- RWDを用いた治療効果検討の再現性向上のためのプロトコルテンプレート (HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force)
<https://onlinelibrary.wiley.com/doi/full/10.1002/pds.5507>

<薬事承認申請関連>

- 厚生労働省. 承認申請等におけるレジストリの活用に関する基本的考え方(令和3年3月23日発)
https://www.mhlw.go.jp/web/t_doc?dataId=00tc5763&dataType=1&pageNo=1
- 厚生労働省. レジストリデータを承認申請等に利用する場合の信頼性担保のための留意点(令和3年3月23日発)
<https://www.mhlw.go.jp/hourei/doc/tsuchi/T21032410020.pdf>

<データベース関連>

- 指定難病データベース (厚生労働省)
<https://www.mhlw.go.jp/stf/nanbyou teikyo.html>
- National Clinical Database:NCD (一般社団法人National Clinical Database)
<https://www.ncd.or.jp/>
- がん登録 (国立研究開発法人国立がん研究センター)
https://ganjoho.jp/med_pro/index.html
- MID-NET (独立行政法人医薬品医療機器総合機構)
<https://www.pmda.go.jp/safety/mid-net/0001.html>
- 診療情報集積基盤 (独立行政法人国立病院機構)
https://nho.hosp.go.jp/cnt1-1_000070.html
- 匿名診療等関連情報 (厚生労働省)
<https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou iryou/iryohoken/dpc/index.html>
- 【NDB】匿名レセプト情報・匿名特定健診等情報 (厚生労働省)
<https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou iryou/iryohoken/resepto/index.html>
- 匿名介護情報等 (厚生労働省)
https://www.mhlw.go.jp/stf/shingi2/0000198094_00033.html
- JMDC Claims Database (株式会社JMDC)
<https://www.jmdc.co.jp/>
- 統合医療データベース Cross Fact (株式会社インテージリアルワールド)
<https://www.intage-realworld.co.jp/service/cross fact/>