

データ活用社会創成シンポジウム ポスターセッション20

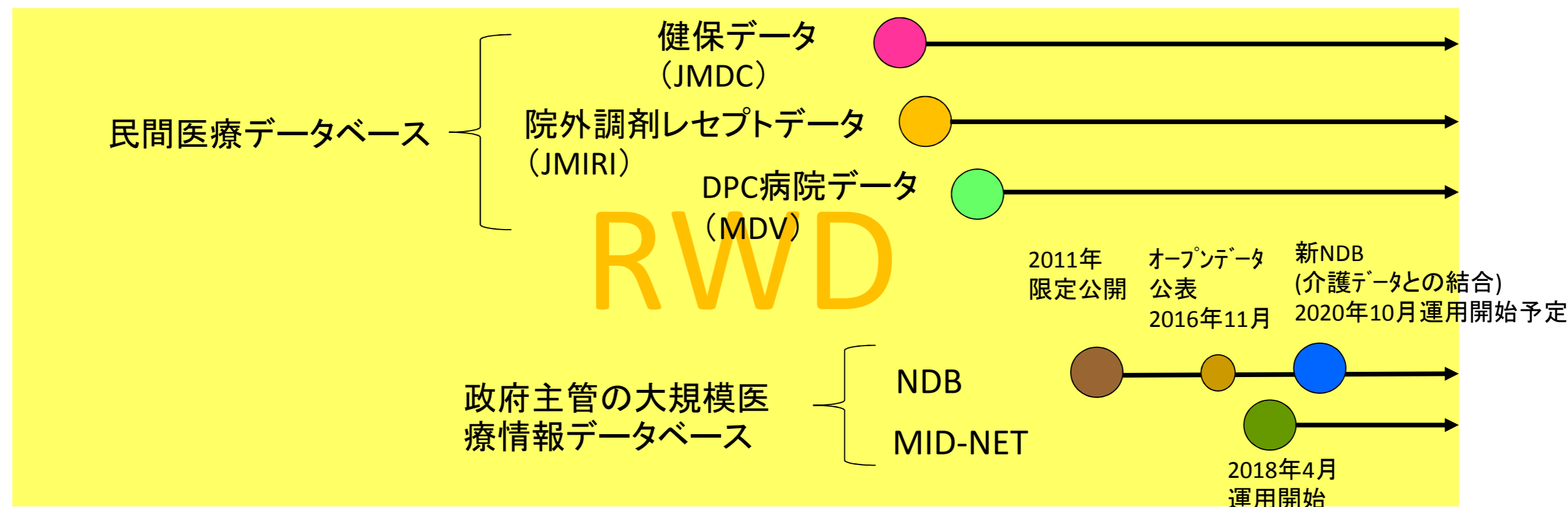
東京大学大学院 清水央子¹⁾、加藤絵津子¹⁾、岩崎宏介¹⁾²⁾

1) 薬学系研究科ITヘルスケア社会連携講座、2) ミリマン・インク

医療ビッグデータを社会インフラとするための挑戦



● 1965年 IMS: 「医薬品市場統計」(JPM)



【日本の医療情報データ概要】

本邦では国民皆保険制度の下、レセプトデータ*を中心に各種の医療情報データが蓄積されてきており、疫学や医療経済などの研究をはじめ、臨床研究や医薬品の開発や販売への応用可能性が飛躍的に高まっている。しかしこれらのデータベースを情報の「粒度」と「悉皆性(カバー率)」の観点から整理してみると(図1)個々のデータベースでは悉皆性、情報粒度のいずれかが限定的であり、医療の世界にはビッグデータは未だ存在しておらず、各種の“スモール”データが散在している状態であることがわかる。医療データに関してはグランドデザインがないままに、結合がしにくいスモールデータが蓄積され散在している状態であり、諸外国と比べると周回遅れになってしまっているのが実情である。

図1 各種RWD相関図(概念図)

そのような中、医療費の適正化を図る目的で全国のレセプトデータを集積した「NDB(レセプト情報・特定健診等情報データベース)」が構築され、現在は政策担当、アカデミアなどが一定の条件の下で申請することにより、医療経済だけでなく、臨床・疫学研究や医薬品開発、市販後の安全性管理などの諸目的での活用が推進されている。

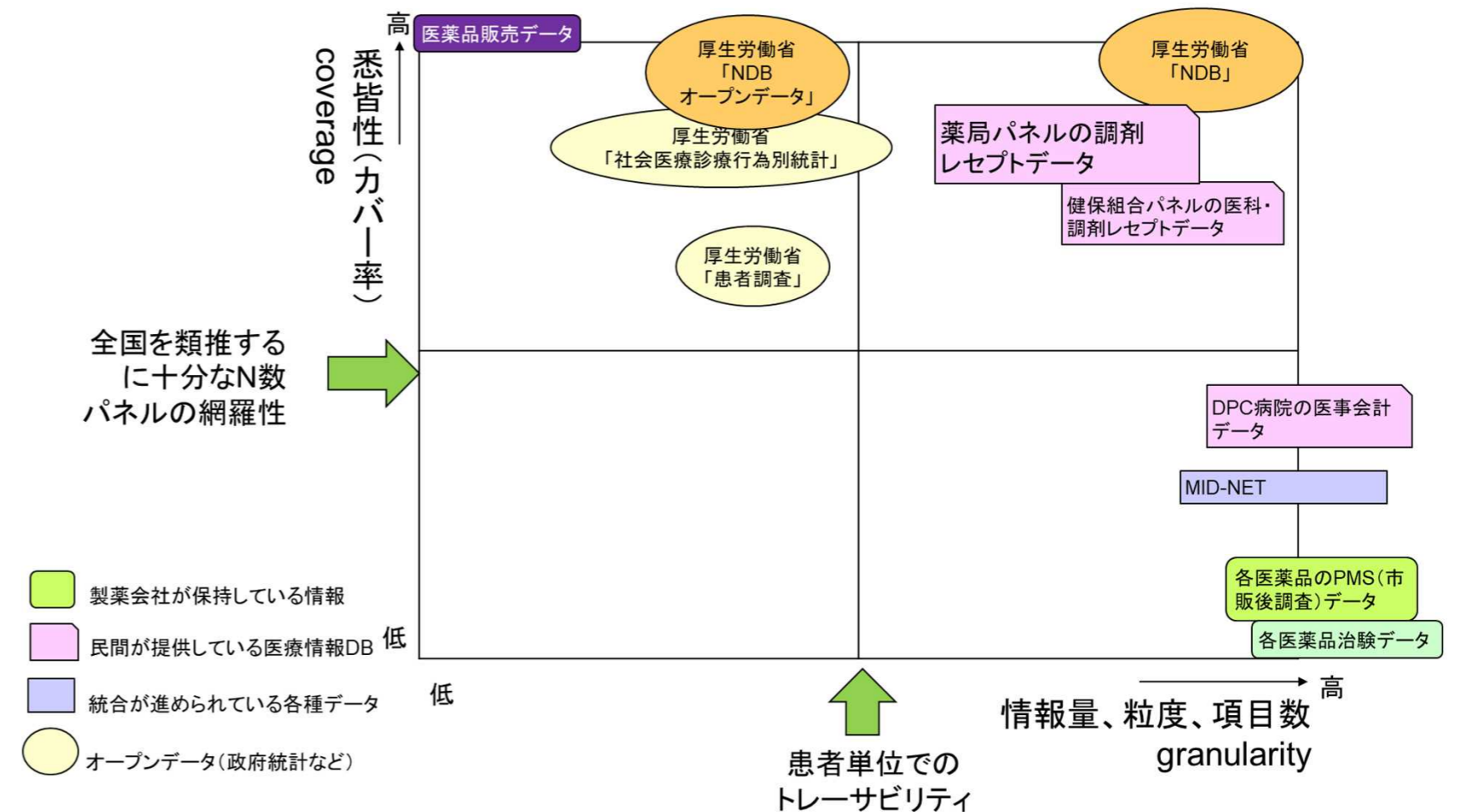


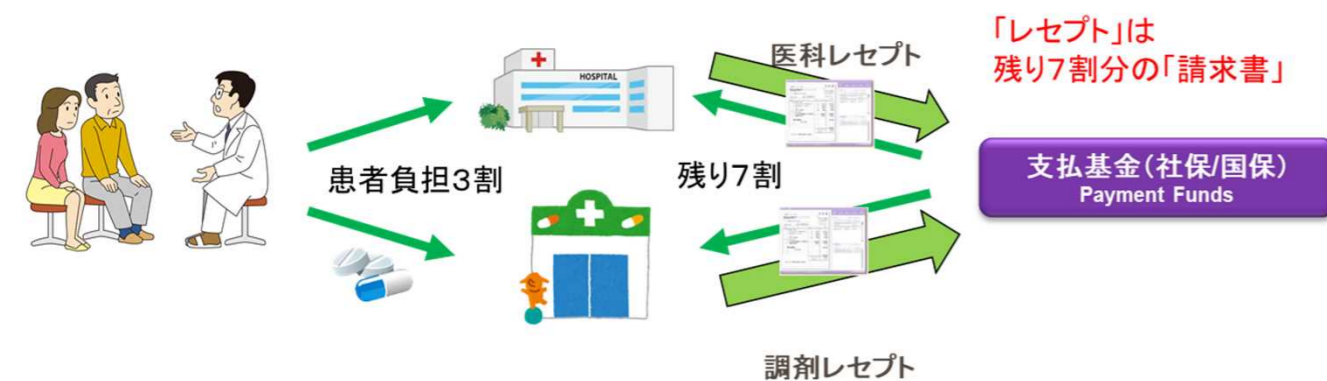
図2 レセプトデータベース比較

	レセプト・データベース	NDB	NDBオープンデータ	JMDC* *日本医療データセンター
悉皆性		全国95%のレセプト	同左	300万人程度、ただし65歳以上は少ない(社会保険組合)
データ期間		2009.4-	第1回:2014.4-2015.3 第2回:2015.4-2016.3 第3回:2016.4-2017.3 第4回:2017.4-2018.3	2005.1-
粒度		○(ただし患者名、施設名は匿名化) レセプト記載内容、ただし一部は抽出不可	△(特定項目の集計値のみ)	◎(ただし患者名、施設名は匿名化) レセプト記載内容
解析上の長所・短所				
	患者ID単位での解析、トレーサビリティ	△ ID1とID2による突合作業が困難	× 患者ID単位での解析は不能	◎ 既に突合した上で集積され患者IDで管理されているため容易。ただし65歳以上の高齢者のデータは少ない
	患者の長期にわたる診療履歴	◎データ集積期間内において、ID突合ができた場合においては、異なる施設間、保険組合異動前後での治療履歴がトレース可能		○データ集積期間内において、同一保険組合に在籍している限りにおいてトレース可能
	解析環境	×(特別抽出の場合) 高性能のハードウェアを有する専用の解析室があることが要件	◎1つ1つのテーブルは小さなEXCELファイルなので扱いは簡単	○ウェブツールで一定の解析が可能。生データ解析については相応のハードウェア、解析ツールや能力が求められる
	アクセシビリティ(誰が利用できるのか?)	△ 現時点では民間には解放されていない	◎ 無料(ウェブよりダウンロード可能)	○ 商業用なので有料、高価
	データ鮮度	△現時点では利用申請後約1年後に、申請されたデータセットの取得が可能となる	×例:2019.8月に2017.4-2018.3の集計値が発表される	◎約3か月前までのデータが解析可能
	その他	マスター整備がされていないので、最初にクレンジングが必要	全国の集計値なので、規模感をつかむのに最適	解析ツールやマスターが充実しており、解析しやすい。

【背景：レセプト情報】

NDB(レセプト情報・特定健診等情報データベース)は電子化されたレセプトデータ*を2009年度より集積したものであり、社会基盤の一部としての期待も高い。しかしながら、データベース構造、その運用体制、制度などから、現時点での活用は限定的となっている。本研究ではNDBの課題について考察し、社会基盤としてのNDBをはじめとする医療データの活用が進むための方策について検討を行う

*「レセプト(診療報酬明細書)」とは医療費の請求明細のことで、保険医療機関、保険薬局が保険者に医療費(患者負担通常3割を除く残り7割部分)の請求書であり、レセプトデータベースとは、レセプトを保険者から匿名加工した上で集積し、解析可能な形でデータベース化したものである。



【考察】

NDBについては、より活用を推進するために、2020年10月の民間公開を目標に「新NDB」の構築が進められている

*「新NDB」: 介護データも統合し、ある程度の経験のある解析者は直接データセットの抽出ができる仕組みを目指している

一方「NDBオープンデータ」(NDB集計値)は2016年以来毎年1回、この8月には第4回の公開がされた。NDBからの集計値のため全国の規模感がつかめることと併せて、全国を均一に捕捉できているため、地域による治療の差異の把握も可能となる。しかしながら、データ公開までに約1年半も要するのはデータの鮮度という観点から課題が残る。今後は公開までの期間を半年程度にまで短縮するとともに、月次での集計値の公開など、より粒度の高い集計も期待したい。

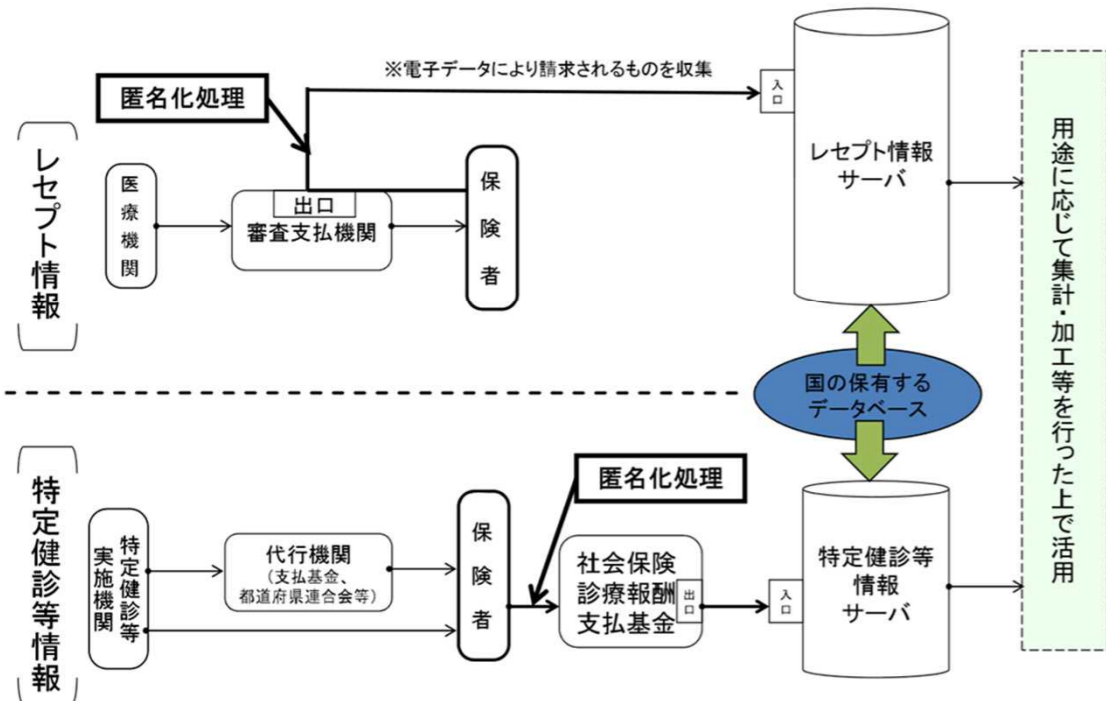
当面は民間のデータベースの優位性は変わらず、今後もさまざまな医療情報データベースの構築が進むと思われるが、NDB単独での解析だけでなく、複数のデータベースを結合した解析を行うことが情報の価値を高めるためには必須であり、そのためには**マスターデータの整備**はますます重要となる。

今後NDBベースに必要なマスター構築を行うことにより、他の医療データさらには異分野のデータと統合した解析が可能となり、社会基盤としてのデータ活用が高まることが期待される

NDB解析におけるチャレンジ

NDB

レセプト情報・特定健診等情報(レセプト情報等)の収集経路



https://www.kantei.go.jp/jp/singi/shakaihoshoukaikaku/wg_dai1/siryou4-2.pdf

NDB利用申請の方法

- ◆ **サンプリングデータセット**: ある1か月の「医科」・「DPC」・「調剤」レセプトのデータを一定の割合で抽出したデータセットでの提供
- ◆ **集計表情報**: 依頼した項目での集計値
- ◆ **特別抽出**: 依頼した条件に合う個票単位のデータセットでの提供

「NDB特別抽出」における解析のハードル

- **解析者ごとに高い仕様の解析環境が求められる**
 - NDB専用の解析室の設置
 - 解析用ハードウェア性能
 - データセットの大きさによって高容量のHDDも必要
 - 解析ツール
 - ローカル設定、HWは盗難防止のため鎖でつなぐ等
 - 専用室の入退出者の厳格な管理
- **解析ごとにデータクレンジングが必要となる**
 - 患者IDの突合、「死亡」フラグ付与等
 - マスターデータの整備
 - ✓ 対象施設のフラグ付
 - ✓ 解析対象の傷病、医薬品、診療行為の特定

NDB解析に必要な「マスターデータ」

主要マスターデータ	一般公開されているマスターの種類
傷病名マスター (約26,000レセ電処理用傷病名)	「診療報酬情報提供サービス(厚労省保険局)」1) 「社会保険診療報酬支払基金」1) 2) 「医療情報システム開発センター[MEDIS]」2) 3)
医薬品マスター (約2万の医療用医薬品)	
診療報酬マスター (医科診療行為: 約7,500)	
施設マスター (全国約18万医療施設)	全国版のマスターは存在しない 研究者各自が各地方厚生局のサイトより47都道府 県分の医療機関リストをダウンロードし統合

共通課題: いずれもある時点でのマスターであり、過去に遡った累積情報を収載していない

データ解析という“二次利用”を想定したマスターデータが整備されていない

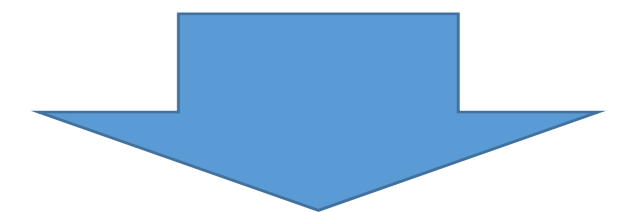
- 1) 別サイトで運用されているがソースは同じ
- 2) 「傷病名」においては共同で更新
- 3) 「診療報酬」においては、手術・検査など一部の項目のみ

「医薬品マスター」の課題を整理

1. ある時点でのマスターしか存在せず、過去に遡って医薬品を収載したマスターが存在しない
⇒分析対象期間に存在していた薬剤の抽出が困難
2. 名称変更に伴い医薬品コードが変更になってしまった場合、同一薬剤を同定することが容易でない
3. レセプト申請・審査、物流管理などの観点で必要な項目しか収載していないため情報が十分でない
 - 「成分名」の情報が含まれていないため、成分単位で薬剤を括った解析が困難
 - 複数の剤型や規格が存在する薬剤における数量分析の場合等、成分量に換算した分析ニーズは多いが「成分量」に関する項目はない

他のデータベースからの情報と組み合わせる際にも、マスターが重要となる

現状、各研究者が各自の分析領域、リサーチクエスチョンに応じて、都度手作業で独自マスターを作成しているのが実態



☞ 解析のための
共通マスター
構築の必要性

医薬品マスター課題#3

【事例】アリピプラゾール“内用液”の後発品浸透率の計算

レセプトコード	販売名	規格単位	薬価単位で の成分量	成分単位
先発	620009105 エビリファイ内用液0.1%	0.1% 1 mL	1	mg
後発	622539401 アリピプラゾール内用液3mg分包「タカタ」	0.1% 3 mL 1包	3	mg
	622539501 アリピプラゾール内用液6mg分包「タカタ」	0.1% 6 mL 1包	6	mg
	622539601 アリピプラゾール内用液12mg分包「タカタ」	0.1% 12 mL 1包	12	mg
	622541601 アリピプラゾール内用液3mg分包「トーワ」	0.1% 3 mL 1包	3	mg
	622541701 アリピプラゾール内用液6mg分包「トーワ」	0.1% 6 mL 1包	6	mg
	622541801 アリピプラゾール内用液12mg分包「トーワ」	0.1% 12 mL 1包	12	mg
		・ ・ ・	6メーカーより後発品3mg分包、6mg分包、12mg分包が販売されている	

上記先発品、後発品ともに容量3mL、6mL、12mL包装のものが100包ずつ処方された場合の後発品浸透率『計算式：後発品の数量／(先発品の数量＋後発品の数量)』は？

本来知りたいのは、同じ単位(成分量)に換算した浸透率

50%

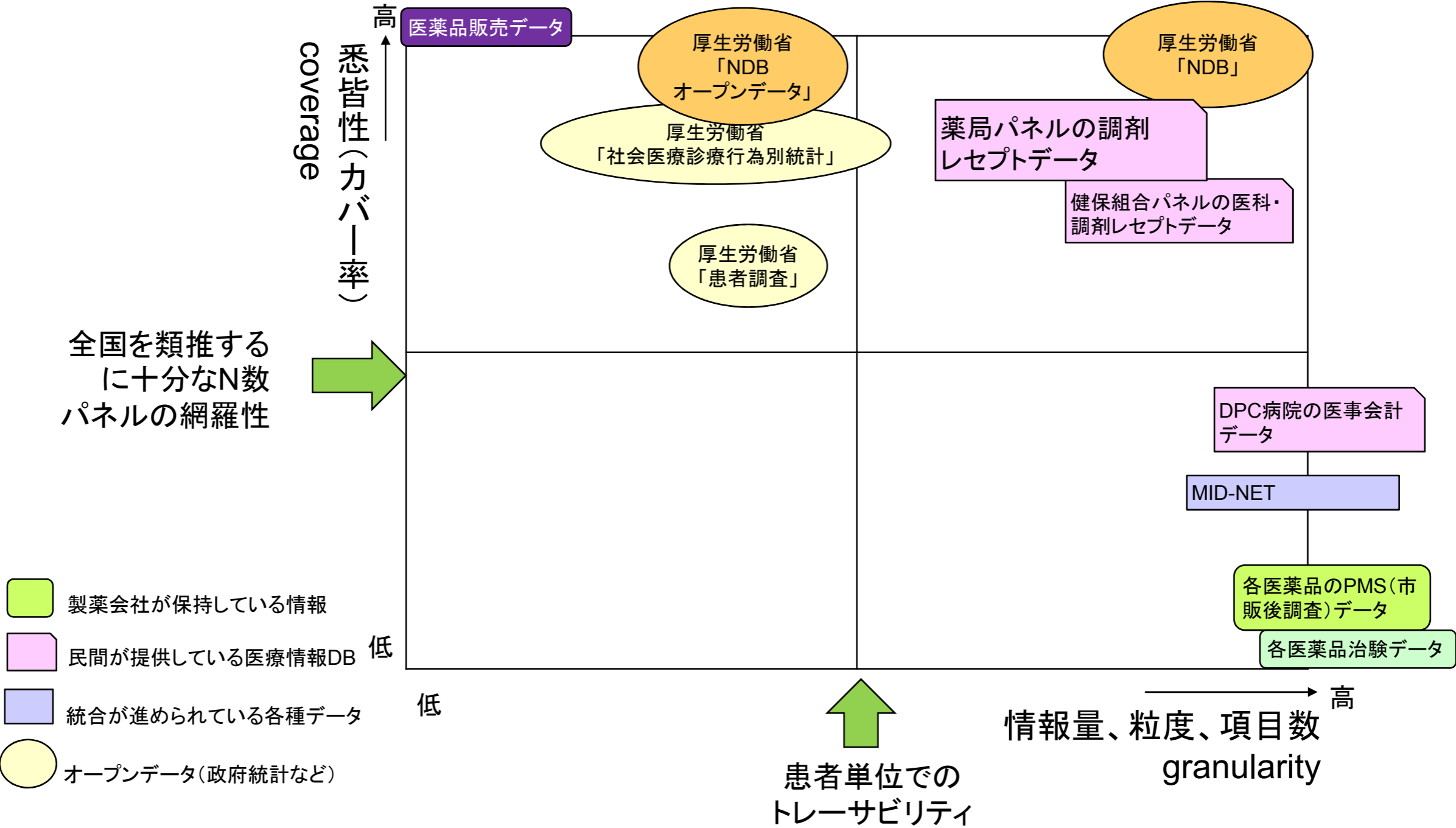


しかし、レセプト上の処方数量データ(薬価単位)を元に計算すると

$$\frac{\text{後発品}300}{\text{先発品}(300+600+1200) + \text{後発品}300}$$

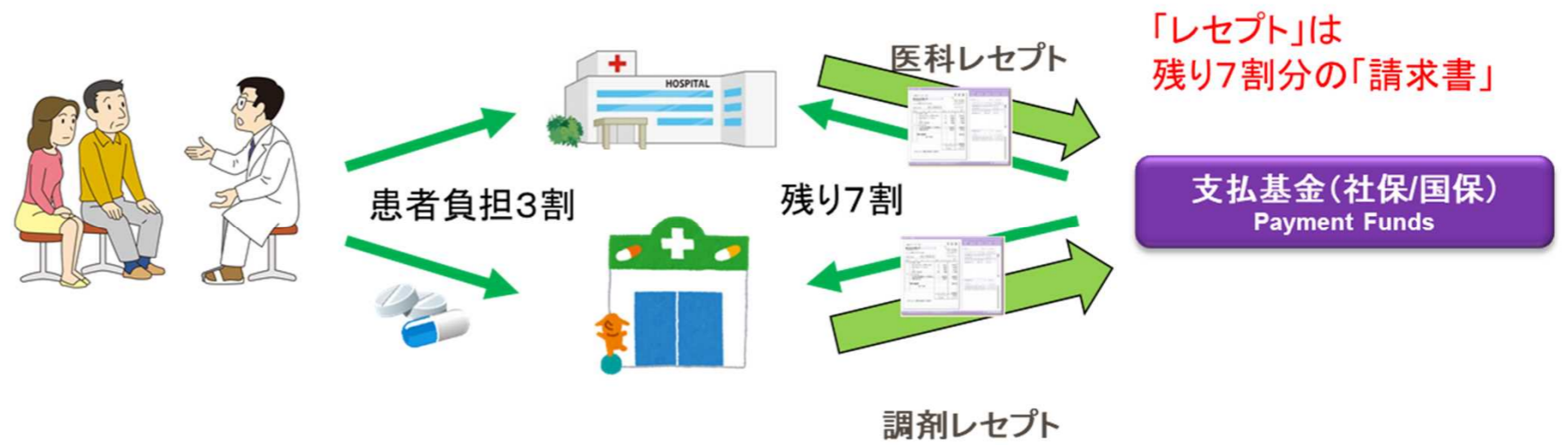
で12.5%となってしまう...

医療ビッグデータ相関図



	レセプト・データベース	NDB	NDBオープンデータ	JMDC* *日本医療データセンター
悉皆性		全国95%のレセプト	同左	300万人程度、ただし65歳以上は少ない(社会保険組合)
データ期間		2009.4-	第1回:2014.4-2015.3 第2回:2015.4-2016.3 第3回:2016.4-2017.3 第4回:2017.4-2018.3	2005.1-
粒度		○(ただし患者名、施設名は匿名化) レセプト記載内容、ただし一部は抽出不可	△(特定項目の集計値のみ)	◎(ただし患者名、施設名は匿名化) レセプト記載内容
解析上の長所・短所				
	患者ID単位での解析、トレーサビリティ	△ ID1とID2による突合作業が困難	× 患者ID単位での解析は不能	◎ 既に突合した上で集積され患者IDで管理されているため容易。ただし65歳以上の高齢者のデータは少ない
	患者の長期にわたる診療履歴	◎データ集積期間内において、ID突合ができた場合においては、異なる施設間、保険組合異動前後での治療履歴がトレース可能		○データ集積期間内において、同一保険組合に在籍している限りにおいてトレース可能
	解析環境	×(特別抽出の場合) 高性能のハードウェアを有する専用の解析室があることが要件	◎1つ1つのテーブルは小さなEXCELファイルなので扱いは簡単	○ウェブツールで一定の解析が可能。生データ解析については相応のハードウェア、解析ツールや能力が求められる
	アクセシビリティ(誰が利用できるのか?)	△ 現時点では民間には解放されていない	◎ 無料(ウェブよりダウンロード可能)	○ 商業用なので有料、高価
	データ鮮度	△現時点では利用申請後約1年後に、申請されたデータセットの取得が可能となる	×例:2019.8月に2017.4-2018.3の集計値が発表される	◎約3か月前までのデータが解析可能
	その他	マスター整備がされていないので、最初にクレンジングが必要	全国の集計値なので、規模感をつかむのに最適	解析ツールやマスターが充実しており、解析しやすい。

【レセプトデータベース】



レセプトデータでは、一人の患者が別の疾患で複数の医療機関にかかっている、あるいは一つの疾患で大病院で手術を受けた後地元の診療所でフォローアップを行っているなど、一人の患者の治療履歴が長期にわたって追うことができるなどの長所の反面、レセプト病名(保険適用するために実際とは異なる病名を記載するケース)、治療内容はわかるが治療のアウトカム(結果)は把握できないなどの短所も指摘されている。

レセプトデータベースの代表としてJMDC(日本医療データセンター)が数十～数百万人のデータを蓄積して2005年頃より商業用データベースとして販売している他、3年前よりNDBの1年度ごとの集計値を「NDBオープンデータ」として公開している。同じレセプト情報を集積したもののでもそれぞれに長所・短所があるが、カルテやIoTからのバイタル情報のデータと併せて、医療の実態や経済を把握するのに重要な情報源として期待されている