



LDI利用者の使ったモジュール(調査報告)

株式会社NTTデータ
システムインテグレーション事業本部
井上裕文

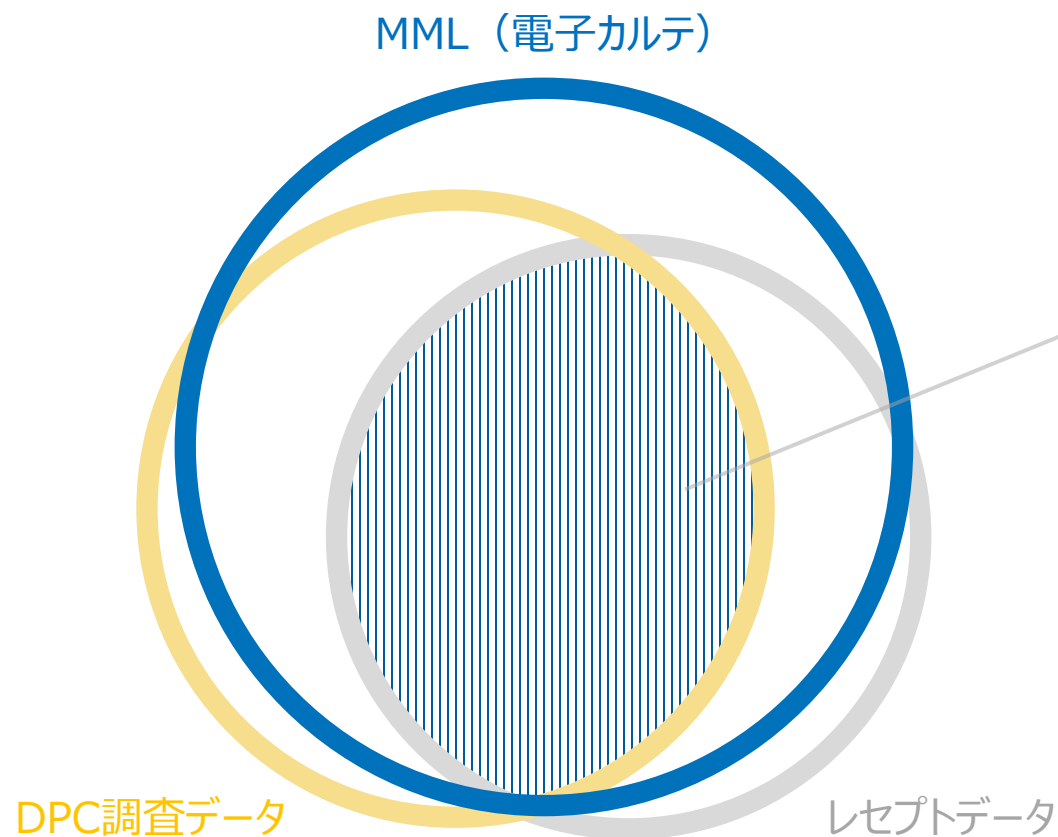
17th .May.2025
Seagaia コンベンションセンター オーチャード
Seagaia meeting 2025 in 宮崎

NTT DATA

Overview

千年カルテDB Overview

レセプト、DPC、MML（電子カルテ）が揃った患者数は190万人を超えるデータソースを活用可。
あらゆる利用ケースへの期待に対応可能なデータベース。



■ 研究に資する利用可能なデータ



施設数

24施設 + 5施設（予定）



患者数

約 **199.9** 万人



データ期間

2024年 **12**月末まで

※開始は各施設のデータ提供開始月（2016年1月～）

千年カルテDBが扱う医療情報

症例単位

- 患者情報
- 保険情報
- 診断情報

レセプト

- 診療行為
- 医薬品情報
- その他

- 入退院情報
- 診断群分類
- その他

DPC調査

【構造化データ】

- 患者情報
- 健康保険
- 診断履歴
- 検歴情報
- バイタルサイン

- 体温表
- 処方箋
- 注射記録
- その他

【非構造化データ】

- 経過記録
- 臨床サマリー(退院時サマリーを含む)
- 紹介状
- 報告書(読影レポート等)
- 医用画像(2024年10月～サービス開始)

MML(電子カルテ)

※電子カルテベンダーによって連携モジュールや取り込みに差異あり

非構造化データ(電子カルテテキスト情報)

認定事業者であるLDI、NTTデータではセキュアな環境下で、各医療機関のMMLテキスト情報を閲覧・抽出・匿名加工処理が可能

■ 非構造化データ（MMLカルテテキスト情報）のイメージ

医療記録		内容
経過記録	S（主観的情報）	XXを飲み始めてから、便秘が続いています。吐気はないです。
	O（客観的情報）	嘔気なし、排便1回/週、普通便、発熱なし、腹部平坦軟
	A（評価）	カリウム値は基準値内推移しており安定しているが、便秘症状がありx xの副作用疑いあり
	P（計画）	x xは継続処方、〇〇を追加処方し経過観察。便秘続く場合は△△を検討。
臨床サマリ（入院時・退院時）		【主訴】便秘 【既往歴】慢性腎臓病、糖尿病 【家族歴】なし #1 YYY #2 xxx… 2020年4月XXクリニックより紹介、当科でABCの診断、xxx治療を開始、2021年7月症状悪化に伴いx xの処方を開始。カリウム値は基準値の推移を経過、ただし副作用症状の発現疑いあり、 今後は…

※SOAPとは

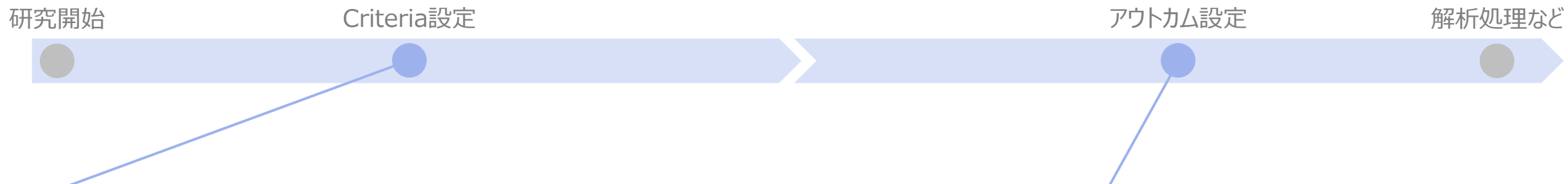
医療記録（経過記録）を記録する際の基本的な記載方針。

S(主観的情報)、O(客観的情報)、A(評価)、P(計画)の4項目に分類して記載するのが理想的とされる

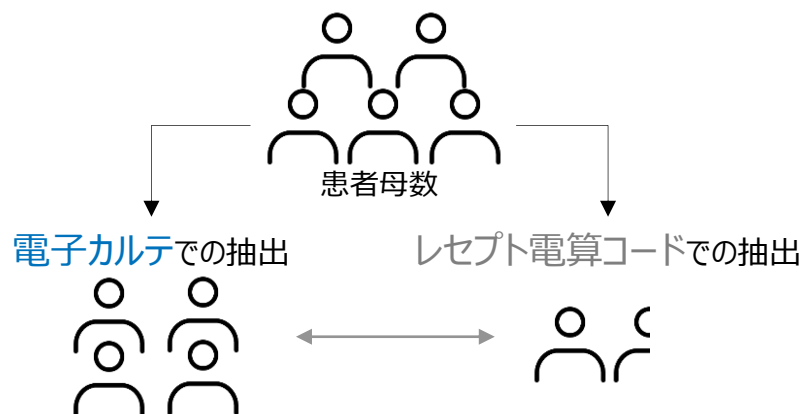
[日本内科学会雑誌第106巻第12号 \(jst.go.jp\)](https://www.jst.go.jp/)

千年カルテの特徴 | レセプト・DPCデータ由来のデータベースとの違い

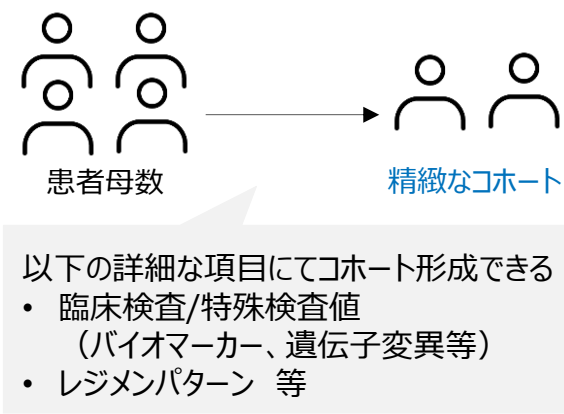
DB研究のCriteriaおよびアウトカム設定フェーズにおいて、以下 3 つの観点で違いがあります。



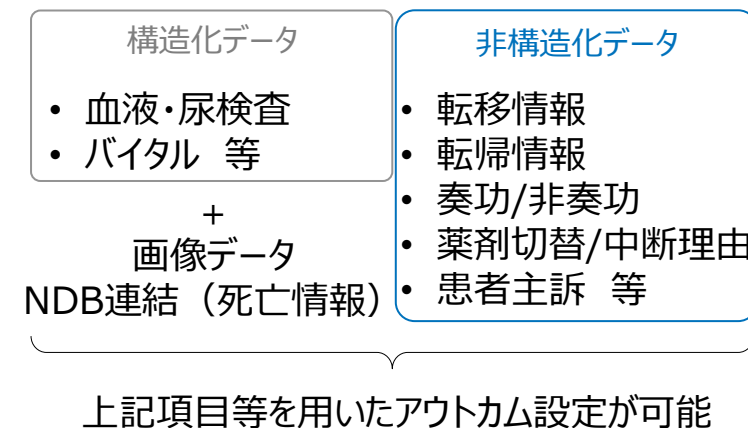
①母集団からより正しく病名・サブタイプを把握できる



②精緻なコホート形成ができる



③多様なアウトカムが設定できる



既存研究におけるリミテーションと解決の方向性

既存のリアルワールドデータベース研究において、主なリミテーションを以下の通りと理解、千年カルテでは既存研究のリミテーションへの解決の方向性を提案可能。

凡例： 構造化データ
非構造化データ

従来の研究・調査

レセプトデータ・DPC調査データの
インプット（診療行為）情報が主体

- ✓ 病名コード(ICD-10やレセプト病名)では正確な病名が定義できない。
- ✓ 患者特性(検査値・バイタル・遺伝子変異情報)などの治療に関連する臨床情報が欠落。
- ✓ 治療切替、治療中断などの理由に関する情報は含まれていない。
そのため、病勢進行や有害事象による治療変更理由を明確に分析できない

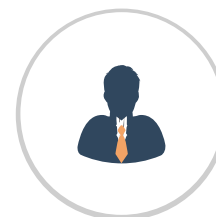
千年カルテ活用による解決の方向性

レセプトデータ・DPC調査データに加え、電子カルテデータの
アウトカム（診療行為の結果）情報を活用



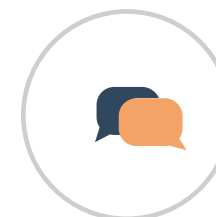
客観的情報
(検査結果)

- 血液・尿検査値
- バイタル 等
- 遺伝子・染色体異常 等



医師による評価

- 詳細な病因、病型
- 薬剤切替、抵抗性理由
- 有害事象、副作用発現



患者主訴・
コミュニケーション

- 治療における困りごと
- 有効性に関する患者評価

- ✓ より適切な患者を組み込み、患者層別化
- ✓ 医師判断や患者主訴を元に、より詳細な治療実態の把握

提供アウトプット例(解析用データセット)

提供方法の1つである「解析用データセット形式」では、患者個人単位の匿名加工処理を実施されたデータセットをご提供し、統計解析・Publicationに活用が可能

<アウトプットイメージ 前立腺がん>

DPC・レセプトデータから取得できる項目例											電子カルテ(構造化データ)から取得できる項目例					電子カルテ(テキスト情報)から取得できる項目例				
識別ID	年齢	主疾病	主疾病 確定診断日	骨転移 確定診断日	薬剤名	入外 区分	処方日	処方 診療科	投与 量	投与 回数	直前PSA 検査日	直前 PSA値	直後PSA 検査日	直後 PSA値	...	痛み	肝機能障害 有無 +日付	医師評価 +日付	...	ECOG PS +日付
Abe52edd8	70代	去勢抵抗性前立腺癌	20201201	20201225	薬剤A	外来	20210101	泌尿器科	3	7	20201225	65	20210107	65		1		奏効		0
abe52edd8	70代	去勢抵抗性前立腺癌	20201201	20201225	薬剤A	外来	20210108	泌尿器科	2	7	20210107	65	20210108	50			1	進行		1
abe52edd8	70代	去勢抵抗性前立腺癌	20201201	20201225	薬剤A	外来	20210115	泌尿器科	1	14	20210108	50	20210116	48		1	1	安定		2
abe52edd8	70代	去勢抵抗性前立腺癌	20201201	20201225	薬剤B	外来	20210115	泌尿器科	1	14	20210108	50	20210116	48		1		安定		
abe52edd8	70代	去勢抵抗性前立腺癌	20201201	20201225	薬剤B	外来	20210201	泌尿器科	2	14	20210130	60	20210215	30						
f615db623	60代	前立腺癌	20201010		薬剤C	入院	20210101	腫瘍内科	1	1							1	進行		
f615db623	60代	前立腺癌	20201010		薬剤C	入院	20210102	腫瘍内科	1	1										
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

非構造化データは
構造化処理を実施してご提供

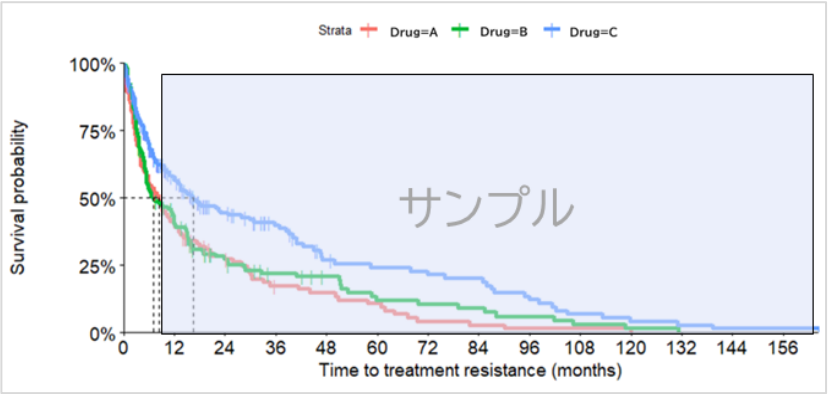
提供アウトプット例(統計情報)

提供方法の1つである「統計情報形式」では、利活用者様のご要望内容に応じて様々な形式でアウトプットを作成しご提供が可能

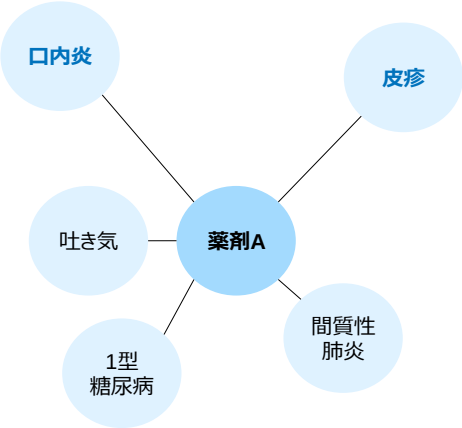
統計情報※ご提供アウトプット例

	全体 n=xxx	A群* n=xxx	B群* n=xxx
性別,n(%)			
男性	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
女性	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
BMI			
中央値 [四分位範囲]	xx.x [xx.x-xx.x]	xx.x [xx.x-xx.x]	xx.x [xx.x-xx.x]
平均値 (標準偏差)	xx.x (x.x)	xx.x (x.x)	xx.x (x.x)
不明,n(%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
喫煙歴,n(%)			
あり	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
なし	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
不明	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)

集計表形式



グラフ形式



テキストマイニング

※学会発表、論文化を目的とした統計解析は現時点では承っておりません。
CRO等の企業への委託が必要となります。

02

利用状況について

千年カルテ利活用承認実績

2024年3月時点、利活用実績47件のうち、ほぼ全ての研究においてMMLデータ（検査結果 or/and テキスト情報）を活用した研究を実施。

利用目的等審査委員会 承認実績

2024年3月13日時点

No.	承認日	課題名	活用データ項目				
			電子カルテ	DPC調査	レセプト	医用画像	活用者区分
1	2020年10月20日	乳癌のサブタイプ別、治療実態を探るための千年カルテデータのF					
2	2020年10月20日	がん患者の臨床アウトカムにおけるEHRデータベースを用いた評価					
3	2021年1月15日	自己免疫疾患領域における寛解指標のフィージビリティ確認					
4	2021年3月5日	検査値等を用いたウイルス性肝炎患者研究のフィージビリティス					
5	2021年5月26日	検査項目の多施設実用手法開発を目的とした研究					
6	2021年7月15日	非構造化データの評価方法確立を目的とした研究					
7	2021年7月15日	希少疾病領域における症状把握を目的としたフィージビリティ検					
8	2021年7月15日	乳がんデータ項目に関するフィージビリティ調査					
9	2021年8月31日	匿名加工医療情報のAI研究への利活用可能性の検討					
10	2021年9月28日	心不全データベース研究のためのFeasibility調査					
11	2021年10月28日	感染症に対するTreatment flow 及び 関連医療費の推計					
12	2021年10月28日	がん患者の臨床アウトカムにおけるEHRデータベースを用いた評価					
13	2021年11月30日	自然言語解析ー					
14	2021年11月30日	肺がん・乳がん患者の治療実態把握及び病気の進展に関する					
15	2021年11月30日	電子カルテのテキストを活用したRECIST評価の辞書作成					
16	2022年2月18日	希少疾病の罹患リスク予測モデル構築					
17	2022年3月8日	電子カルテ情報を活用した、臨床試験の新規手法論開発					

利用目的等審査委員会 承認実績

2024年3月13日時点

No.	承認日	課題名	活用データ項目				
			電子カルテ	DPC調査	レセプト	医用画像	活用者区分
17	2022年3月8日	電子カルテのテキスト情報を用いた癌患者の治療実態に関する検討					民間企業
18	2022年5月17日	アウトカムバリデーションでのフィージビリティ研究					
19	2022年7月13日	千年カルテ二次利用データベースを利用した心不全患者の病態に対する因子探索					
20	2022年8月22日	電子カルテ情報を用いた有効性等に関する新規エビデンス創出の検討					
21	2023年3月13日	先天性代謝異常症患者の治療実態の把握					
22	2023年3月13日	消化管領域における治療実態調査					
23	2023年3月13日	感染症におけるTreatment flow 及び関連医療費の推計					
24	2023年4月12日	がん患者の臨床アウトカムの薬剤群間比較におけるEHRデータベースを用いた評価方法の					
25	2023年4月12日	後ろ向き研究					
26	2023年4月12日	医学研究における匿名加工情報利用の最適化の検討					
27	2023年5月10日	検査値の患者属性別統計					
28	2023年8月24日	アウトカムバリデーションスタディの外挿可能性、代表性検討					
29	2023年9月27日	電子カルテを活用した病態進行を予測するAI開発					
30	2023年10月10日	血液がんにおける治療実態研究					
31	2023年10月10日	（電子カルテ情報を活用した患者背景設定と臨床アウトカム評価）					
32	2023年10月10日	アレルギー免疫療法の投与継続理由及び中止理由の調査					
33	2023年10月10日	CKD・心不全関連疾患治療薬における治療継続期間毎のインサイト抽出					
34	2023年11月27日	リウマチ疾患、および抗リウマチ製剤による有害事象の検証					

利用目的等審査委員会 承認実績

2024年3月13日時点

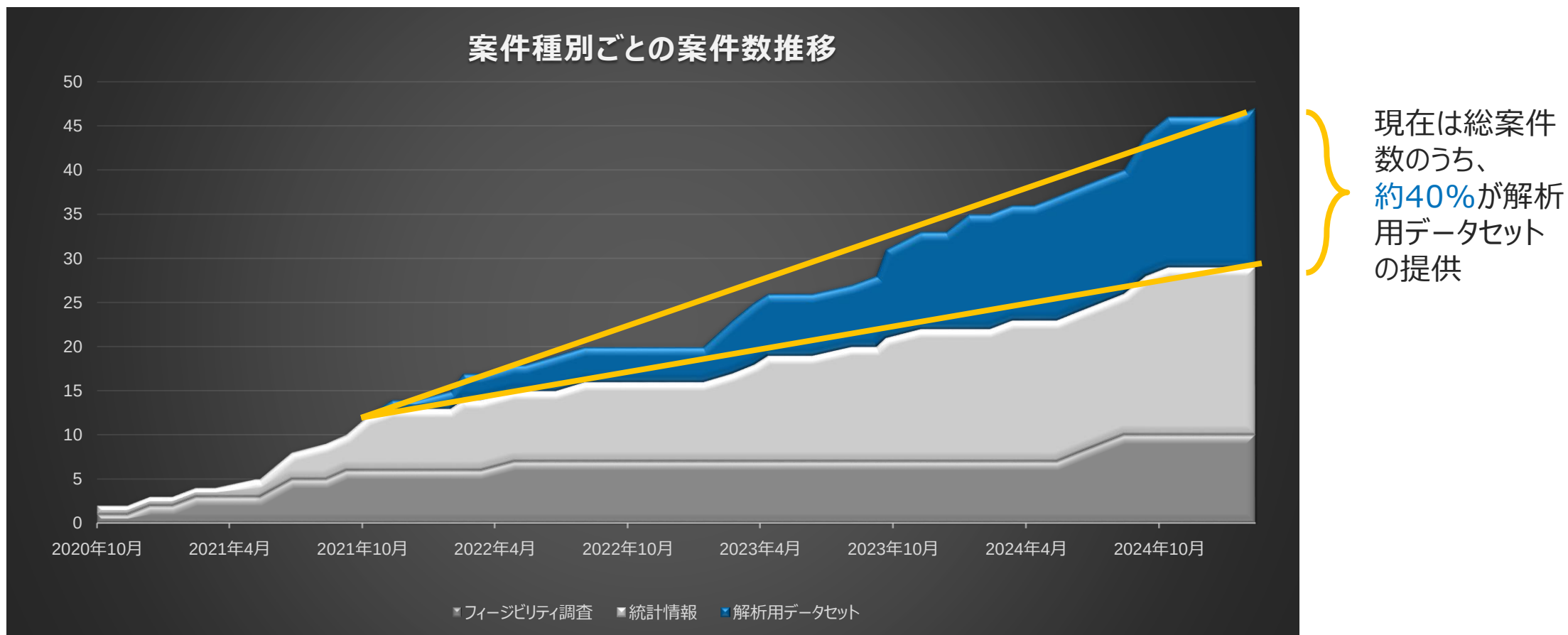
No.	承認日	課題名	活用データ項目				
			電子カルテ	DPC調査	レセプト	医用画像	活用者区分
33	2023年11月27日	高齢心不全患者における診療ガイドラインに基づく標準的治療と再入院の関連	●	●	●	-	アカデミア
34	2024年2月14日	高齢心不全患者における心臓リハビリテーションの効果と再入院の関連	●	●	●	-	民間企業
35	2024年2月14日	2型糖尿病に対する薬剤の治療効果および治療継続期間の評価	●	●	●	-	アカデミア
36	2024年4月24日	自己免疫性疾患における治療実態調査	●	●	●	-	民間企業
37	2024年6月26日	多数医療施設における頻出クリニカルパスの分析	●	●	●	-	アカデミア
38	2024年7月17日	進行・再発子宮内膜癌における治療実態調査	●	●	●	-	民間企業
39	2024年8月21日	血液がんにおける治療実態把握に向けたFeasibility調査	●	●	●	-	民間企業
40	2024年9月18日	血液がん領域におけるデータベース研究に関するフィージビリティ調査	●	●	●	-	民間企業
41	2024年10月16日	骨粗鬆症性骨折の疫学調査	●	●	●	●	民間企業
42	2024年10月16日	自己免疫性疾患領域におけるPHRおよびdBM候補の探索	●	●	●	-	民間企業
43	2024年10月16日	がん患者に対する臨床的負担におけるEHRデータベースを用いた後ろ向き研究	●	●	●	-	民間企業
44	2024年10月16日	日本における食道扁平上皮癌患者の治療実態について	●	●	●	-	民間企業
45	2024年11月12日	電子カルテ情報を用いた内分泌代謝疾患の治療実態と臨床症状に関する調査	●	●	●	-	民間企業
46	2024年11月12日	医用画像の匿名加工医療情報の生成AIへの利用可能性検討	-	-	-	●	アカデミア
47	2025年3月5日	レセプト情報、電子カルテ情報を用いた救急関連モデル構築、医療経済分析	●	●	●	-	民間企業

詳細はホームページに掲載しております。
利用目的等審査委員会の公表 | ライフデータイニシアティブ / NTTデータ | 千年カルテ (Idi.or.jp)

千年カルテ利活用承認実績 | 詳細

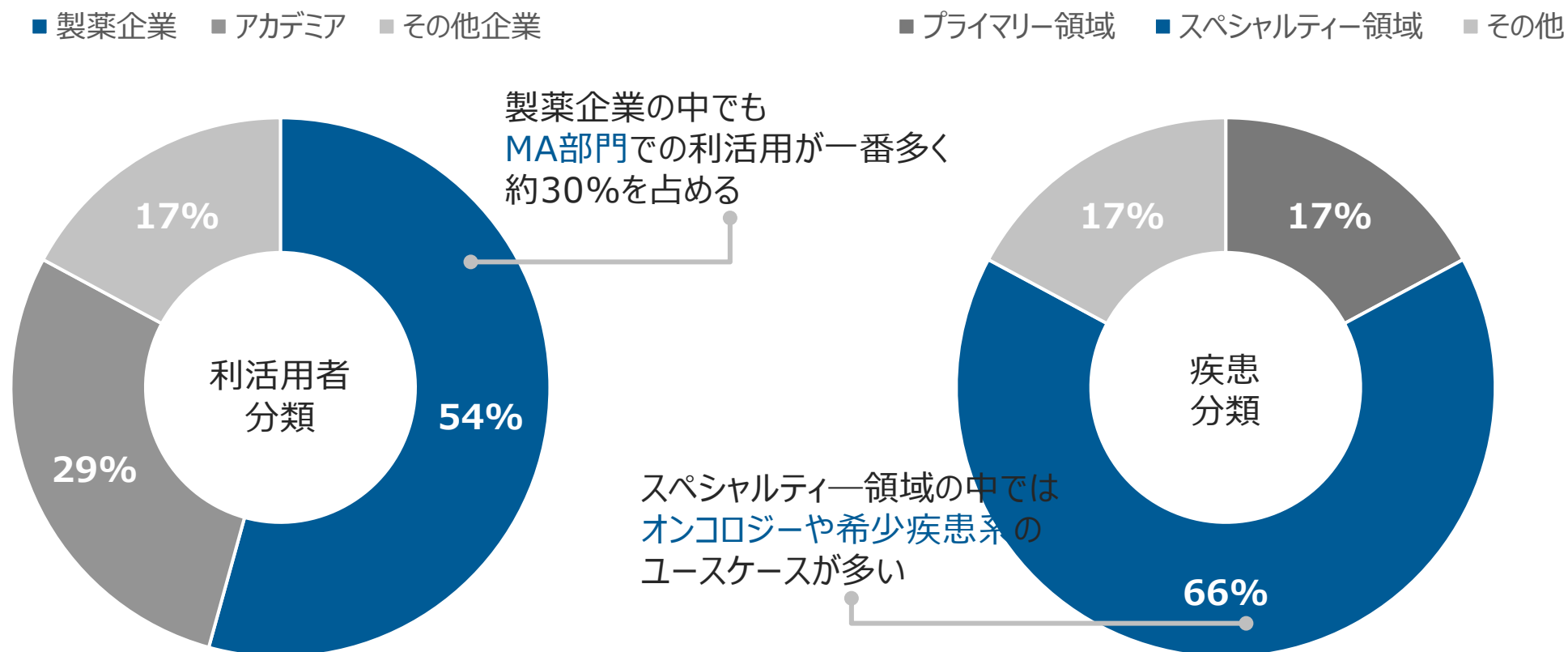
サービス開始直後は、フィージビリティ調査や統計情報での提供案件が中心。

徐々に、検査情報やテキスト情報の活用可能性についての理解が浸透した結果、匿名加工医療情報データの提供が現在の利活用における中心となっている。



千年カルテ利活用承認実績 | 詳細

利活用者分類としては製薬企業様に多く利用いただいております、疾患分類としては、オンコロジー領域・Specialty領域での利活用が進んでいる。最近では、心不全や糖尿病などプライマリー領域での利活用も増加傾向。

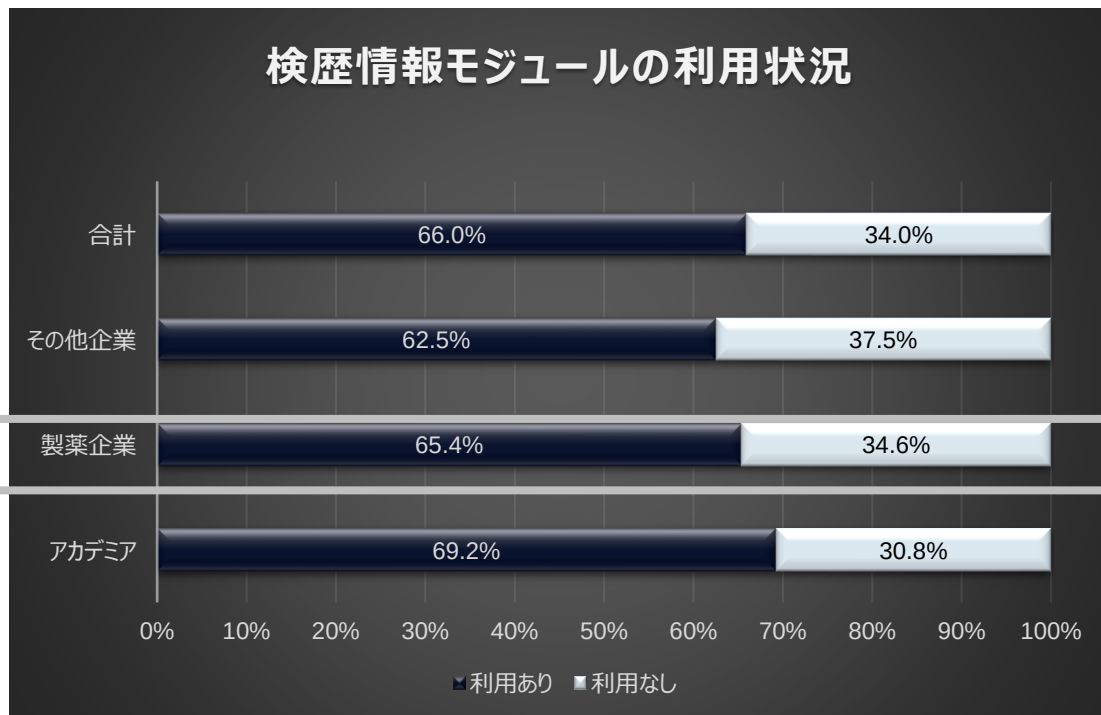


千年カルテ利活用承認実績 | モジュールの利用状況

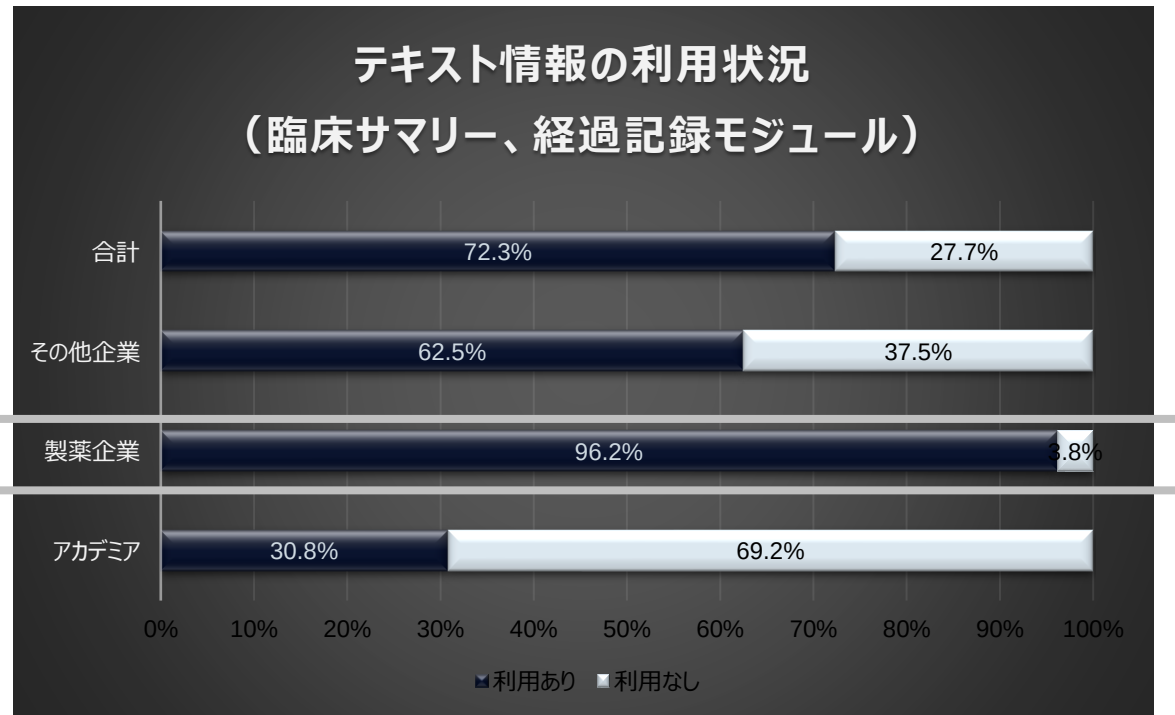
利活用者分類ごとに検歴情報と、テキスト情報（臨床サマリー、経過記録モジュール）の利用状況を分析。
特に製薬企業様の研究においてはテキスト情報を利用することが千年カルテを利用するうえでの目的となっている。

■ 利活用者分類

検歴情報モジュールの利用状況



テキスト情報の利用状況
(臨床サマリー、経過記録モジュール)



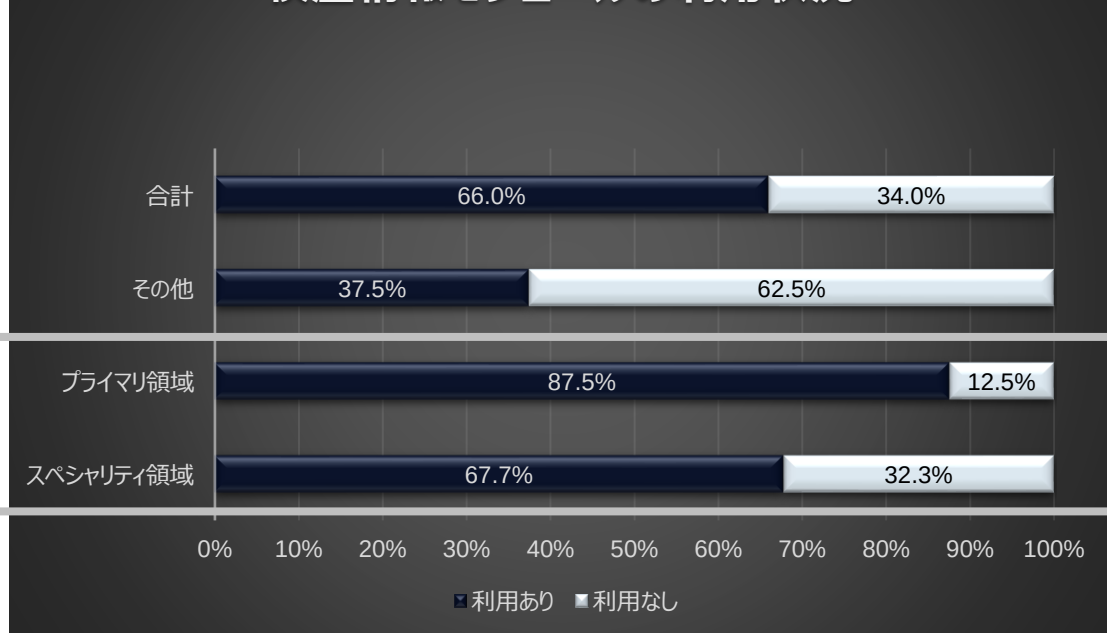
製薬企業の案件においては、ほぼすべての研究
においてテキスト情報を利用

千年カルテ利活用承認実績 | モジュールの利用状況

疾患分類ごとに検歴情報と、テキスト情報（臨床サマリー、経過記録モジュール）の利用状況を分析。
疾患に関する研究においては領域に関わらず、検歴情報、テキスト情報ともに利用されている。

■ 疾患分類

検歴情報モジュールの利用状況



テキスト情報の利用状況
(臨床サマリー、経過記録モジュール)



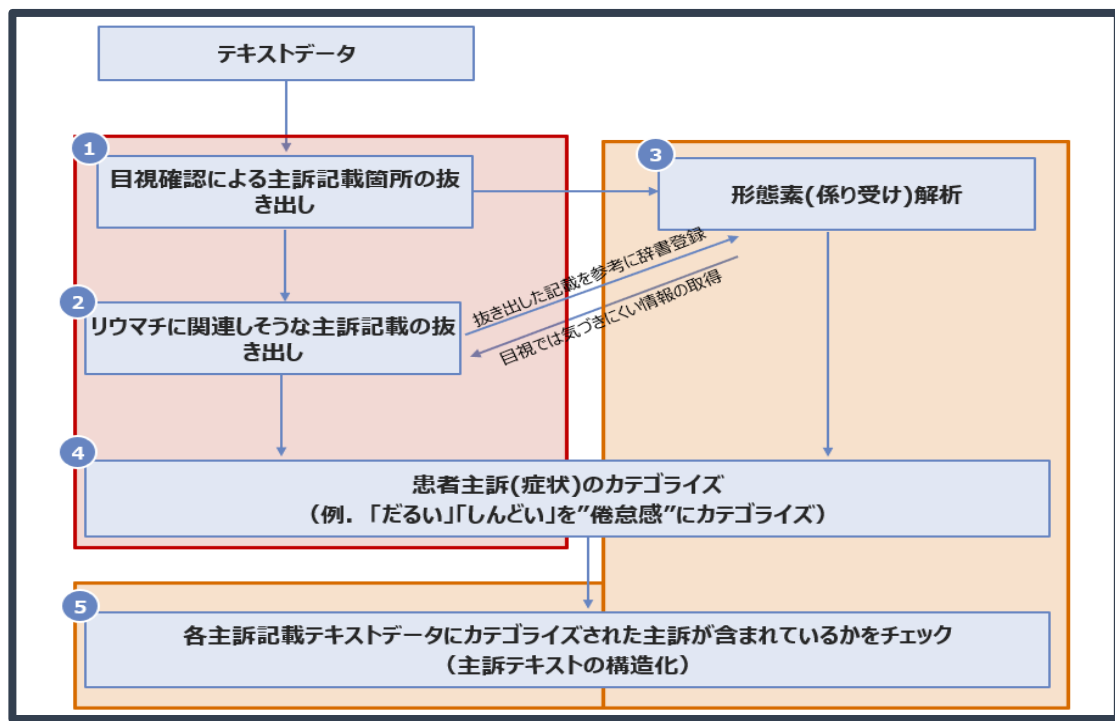
領域に関わらず、疾患に関する研究においては
検歴情報、テキスト情報を利用

テキスト情報データ | 抽出アプローチ

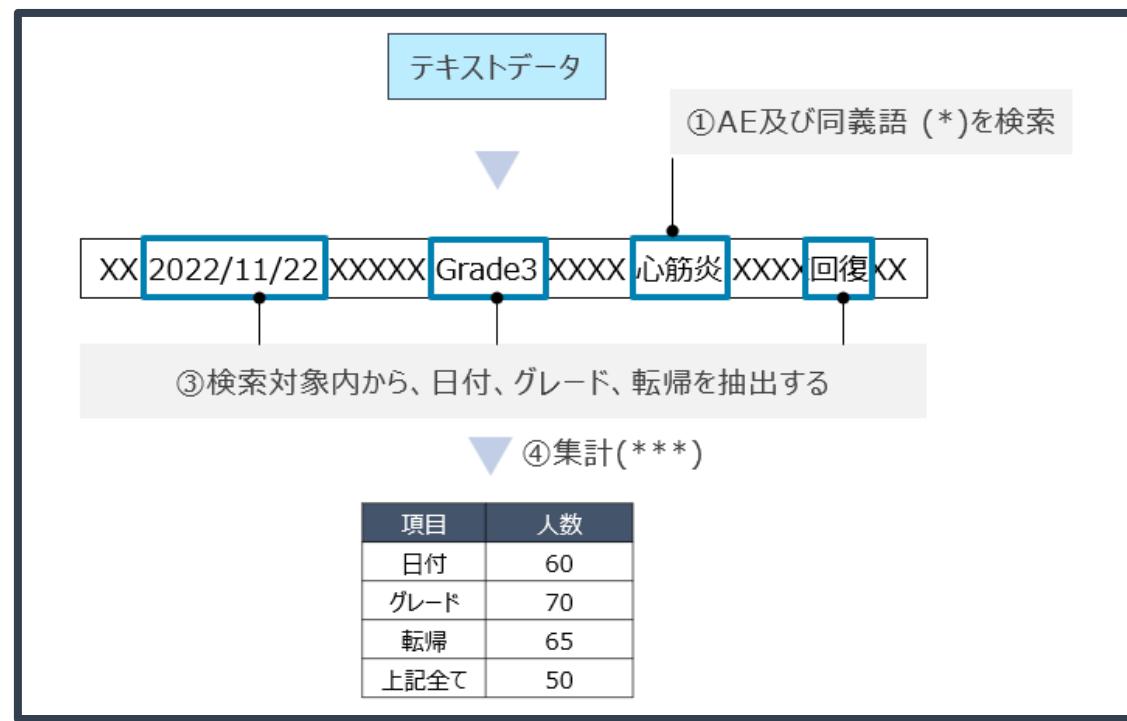
MMLにおけるテキストの記述量は膨大であり、必要な情報を効率よく取得する手法が重要となる。

当社ではテキストマイニングをベースとしたアプローチ、ルールベースで機械的に情報を取得するアプローチなどを準備しており、案件に応じ使い分け、効果的かつ迅速に情報を抽出している

テキストマイニングをベースとしたアプローチ



ルールベースの機械抽出をベースとしたアプローチ



テキスト情報データ | 抽出対象データ

MMLの経過記録や臨床サマリーなどの自由記述には、レセプト等の他DBには登録されない臨床上重要な情報が豊富に含まれている。これらを高解像度RWDとして製薬企業などへ提供し、このデータをもとに現場の実態を捉えた臨床研究や意思決定を推進している。下記は、これまでに当社が抽出してきた情報の一例である

疾患分類	データ分類	データ項目
oncology	遺伝子変異	p53, EGFR, ROS1, PD-L1, MMR, MSI
oncology	染色体変異	del(17p), t(4;14), t(14;16)
oncology	検査値	MRD, M蛋白
oncology	リスクスコア	ISS/IPSS/グリソンスコア/癌のステージ
oncology	効果指標	RECIST (CR, PR, PD, SD) ,寛解, 完全寛解
oncology	患者状態	PS, がんの初発/再発
oncology	薬剤	レジメン
循環器	検査値	LVEF (左室駆出率) , 6分間歩行距離, mPAP (平均肺動脈圧)
代謝性疾患	検査値	肝硬度,肝脂肪量
疾患横断的	副作用・有害事象	副作用 (心毒性・免疫チェックポイント阻害剤によるirAE)
疾患横断的	理由	薬剤の切り替え理由・救急搬送の理由

利活用ユースケース例 | 母集団定義

取得可能な情報

1

詳細な病名情報

希少疾患病名、がんサブタイプ病名 等

2

飲酒・喫煙情報

キーワード例)
飲酒有/無・量、喫煙有/無・本数 等

利活用ケース

Criteria設定

詳細な病名の把握および患者組み入れ可能

- ICD-10コードを有するものの、
複数疾患が包含されている可能性がある場合
- ICD-10コードが存在せず、
電子カルテに記載されている医師の判断を元に病名を特定しなければならない場合

患者背景情報

- 飲酒・喫煙情報等、**レセプトデータでは取得できない**
詳細な患者背景情報を含めた治療実態調査が可能

※DPC調査データと組み合わせることで、
入院患者＋外来患者を含めて取得

アウトプット例（イメージ）

特定疾患コード付与患者
※ICD-10コードより

詳細病名付与患者

- 下記電子カルテデータより、
詳細病名の探索が可能
- 電子カルテ病名
- フリーテキスト
(経過記録・サマリ 等)

	全体 n=xxx	A群* n=xxx	B群* n=xxx
性別,n(%)			
男性	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
女性	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
BMI			
中央値 [四分位範囲]	xx.x [xx.x-xx.x]	xx.x [xx.x-xx.x]	xx.x [xx.x-xx.x]
平均値 (標準偏差)	xx.x (x.x)	xx.x (x.x)	xx.x (x.x)
不明,n(%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
喫煙歴,n(%)			
あり	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
なし	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
不明	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)

例) 特定疾患の喫煙歴：DPC+電子カルテにより入院・外来
合わせ、7割ほど取得実績

利活用ユースケース例 | アウトカム情報活用

取得可能な情報

1

様々な主訴情報

キーワード例) 「痛み」・「かゆみ」
「出血」「血便・便潜血」等

2

AE・副作用情報

キーワード例) 【万病辞書等より定義】
痺れ、自律神経、口内炎、下痢・・・等

利活用ケース

アウトカム活用

- 検査値等の客観的アウトカム情報では掴めない**患者さん主訴情報の把握で薬剤切替・中止の理由を把握可能**
- 薬剤服用時から出血等が発現する迄の期間を見ることが可能
(**Time to Eventの把握**) ※右図サンプル

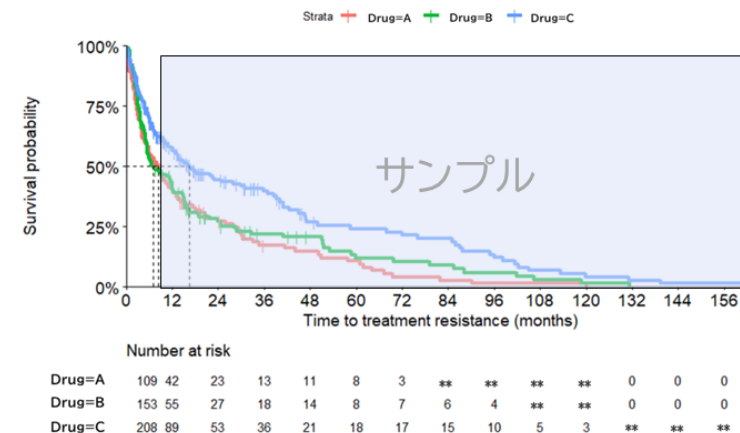
Criteria設定

- レセプトに載らない小さな副作用を電子カルテテキストから検出し、同定した患者におけるICI製剤の治療実態調査

アウトカム活用

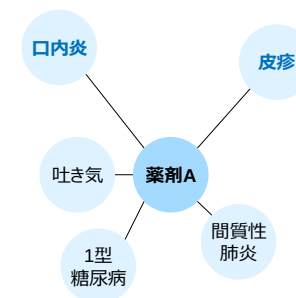
- 万病辞書等における**副作用の上位ワードを包括的に調査することによるシグナル検出**

アウトプット例 (イメージ)



	皮膚発現患者	割合(%)
全体	xxx	xxx.0%
性別		
男性	xxx	xxx.0%
女性	xxx	xxx.0%
年代		
40代未満	xxx	xxx.0%
40代	xxx	xxx.0%
50代	xxx	xxx.0%
60代	xxx	xxx.0%
70代	xxx	xxx.0%
80代	xxx	xxx.0%
90代以上	xxx	xxx.0%
不明	xxx	xxx.0%
合併症		
合併症①	xxx	xxx.0%
合併症②	xxx	xxx.0%
合併症③	xxx	xxx.0%
併用薬		
薬剤①	xxx	xxx.0%
薬剤②	xxx	xxx.0%
薬剤③	xxx	xxx.0%
検査値		
検査値①	xxx	xxx.0%
検査値②	xxx	xxx.0%
検査値③	xxx	xxx.0%
バイタル		
バイタル①	xxx	xxx.0%
バイタル②	xxx	xxx.0%
バイタル③	xxx	xxx.0%

副作用が発現した患者の
背景テーブルを活用した調査



共起ネットワークを活用した
シグナル検出



Life Data Initiative

NTT DATA