

2次利用から見た 千年カルテデータの特長

一般社団法人 ライフデータイニシアティブ 理事
NPO日本医療ネットワーク協会 理事長
宮崎大学医学部附属病院 病院IR部 教授
荒木 賢二

Section1

RWD(Real World Data)

の評価

Section2

千年カルテデータの現 状と未来

Section3

文書情報の価値



リアルワールドデータの価値を測るには

価値が高い → 多くの情報が正確に収集されている

価値の観点	価値が高いとは	千年カルテの状況
施設数	●多いほうが良い	●AMED事業参加 106施設 ↷後述
施設の種別	●様々な種別の施設がある方が良い 病院、診療所、介護、検査、薬局、健診、等	●病院、診療所、検査、薬局、健診 ●介護は今後の課題
情報の種別	●様々な種別の情報がある方が良い DPC/レセ、検査結果、処方/注射、病名、オーダー情報、文書、部門レポート、観察項目、等	●文書を含む多くの種別をカバー
欠損	●欠損はない方が良い ● ミッシングデータ分析 が必要 ↷後述	●欠損なし：検査結果、処方/注射 ●欠損多い：部門レポート
抽出時期	●リアルタイムが良い 年次、月次、日次、リアルタイム	●日次で収集
標準化	●施設間で情報が名寄せできる方が良い	●MMLで標準化 ●検査コード辞書の作成 ↷後述



ミッシングデータ分析

ミッシングデータ（欠損データ）とは、本来あるべきだが欠損している情報
通常の臨床研究では、重要な情報が欠損するとNが一つ減る

情報の変換ステップ	ミッシングの原因（前ステップからの消失）	ミッシングの具体例
本来あるべき診療		
実際に行われた診療	A.様々な状況により、定型的な診療を行わなかった B.施設によって手順が異なる C.医療従事者がうっかりして忘れていた	A.患者が小児なので術前検査を減らした B.施設によるクリニカルパスの違い C.白血球は測定したが好中球は忘れた
電子カルテの情報	A.電子カルテではなく、紙に書いた B.電子カルテではなく、部門システムに入力した C.電子カルテに記録し忘れた	A.病棟の血ガスデータ（結果は紙出力） B.病理システムの病理レポート
MMLの情報	A.紙や部門システムに情報があり、抽出できず B.電子カルテにあるが、MMLでは対象外としている C.電子カルテにあるが、MMLに変換できず D.電子カルテにあるが、病院毎の運用差が大きく対応しない	A.病理システムの組織学的診断 B.処方、注射以外のオーダー情報 C.家系図 D.2号紙、手術記録
2次利用で使われる情報	A.運用差が大きく悉皆性がないため活用できず B.画像や自由記載のため、活用できず C.匿名加工処理における制限で個票としては活用できず D.2次利用としては不要と判断した情報	A.多くの観察項目の情報 B.各種文書 C.希少疾患 D.レセプトの注記

Section1

RWD(Real World Data)

の評価

Section2

千年カルテデータの現 状と未来

Section3

文書情報の価値



LDI保有データ統計 疾患別

4病院、2020年12月までの累積ユニーク患者数

レセプトデータからの統計

入院および外来 疑い病名を除く 辞退患者をのぞく

新生物<腫瘍> (C00-D48)

疾患名	患者数
前立腺癌	4,245
乳癌	3,127
子宮筋腫	2,619
甲状腺腫瘍	2,419
肺癌	2,136
卵巣腫瘍	2,066
甲状腺癌	1,759
直腸癌	1,629
悪性リンパ腫	1,616
乳房上外側部乳癌	1,550

循環器系の疾患 (I00-I99)

疾患名	患者数
高血圧症	36,129
心不全	14,568
狭心症	9,217
うっ血性心不全	7,122
心房細動	4,850
慢性心不全	3,942
脳梗塞	3,877
本態性高血圧症	2,686
深部静脈血栓症	2,563
不整脈	2,448

消化器系の疾患 (K00-K93)

疾患名	患者数
便秘症	35,406
胃潰瘍	21,316
難治性逆流性食道炎	15,264
逆流性食道炎	14,530
慢性胃炎	10,769
胃炎	8,947
維持療法の必要な難治性逆流性食道炎	5,535
大腸ポリープ	4,993
肝機能障害	4,275
口内炎	4,076



DPCデータとMMLの比較

比較項目	DPCデータ	MMLデータ
オーダー情報	●請求する全てのオーダー実施情報を網羅	●オーダー実施情報は処方/注射
アウトカム情報	●なし	●あり
検査結果	●なし	●あり ○検査項目コードの施設間名寄せも可能
観察項目	●なし	●あり（血圧、体温、呼吸数等のバイタル）
文書情報	●なし	●あり
退院時サマリ	●なし ○文書ではないが、様式1には、有用な情報あり	●あり ○構造化はされていないが、有用な情報が多数記載されている
手術記録	●なし	●あり ○術式だけでなく術中の合併症なども記載
抽出時期	●月次もしくは3ヶ月おき	●日次



検査コードの名寄せ

検体検査コードの名寄せ用の類義語辞書の作成

宮大独自コード	レセ電コード	名称	名称_類義語	単位
2100	160017410	T P	TOTAL PROTEINS_TP_総蛋白_ソウタンパク	G/DL
2200	160018910	A l b	ALB_ALBUMIN_アルブミン_アルブミン	G/DL
2300	160019010	U N	BUN_UN_尿素窒素_ニョウソクツツ	MG/DL
2400	160019310	U A	UA_URIC ACID_尿酸_ニョウサン	MG/DL
2500	160019210	C R E	CRE_CREATININE_クレアチン_クレアチン	MG/DL
2700	160017010	T B	BILIRUBIN_BILIRUBINS, TOTAL_TB_総ビリルビン_ソウビビリルビン	MG/DL
2800	160017110	D B	BILIRUBIN, CONJUGATED_DB_直接ビリルビン_直接ビリルビン_チヨクセツビリルビン	MG/DL
2900	160137850	総胆汁酸	TBA_[総胆汁酸]_総胆汁酸_胆汁酸_ソウタンジユウサン	
2903	160137850	総胆汁酸	総胆汁_総胆汁酸_TBA_総胆汁酸_胆汁酸_ソウタンジユウサン	MMOL/L
3000	160019410	G L U	GLU_GLUCCOSE_グルコース_グルコース	MG/DL
3100	160022410	T C	TC_コレステロール_ソウコレステロール_テトラサイクリン	MG/DL
3200	160020910	T G	TG_TRIGLYCERIDE_中性脂肪_チュウセイシヨウ_トリグリセリド	MG/DL
3400	160023110	遊離脂肪酸	NEFA_NON-ESTERIFIED FATTY ACIDS_遊離脂肪_遊離脂肪酸_ユウリシヨウ_ウサン	MEQ/L_MEQ/L
3500	160023410	H D L - C	HDL-C_HDL-CHOLESTEROL_HDL_コレステロール_HDLコレ_HDLコレステロール	MG/DL
4000	160021110	N a	NA_SODIUM_ナトリウム_ナトリウム	MMOL/L
4100	160021410	K	K_POTASSIUM_カリウム	MMOL/L
4200	160021110	C l	CHLORIDE_CL_クロール	MMOL/L
4300	160021510	C a	CA_CALCIIUM_カルシウム_カルシウム	MG/DL

Section1

RWD(Real World Data)

の評価

Section2

千年カルテデータの現 状と未来

Section3

文書情報の価値



人工知能による文書からの情報抽出

患者情報を抽出する手法

ルールベース（正規表現マッチ）

機械学習（系列ラベリング）

概要	<ul style="list-style-type: none"> ● 人手でルールを書いて処理する 	<ul style="list-style-type: none"> ● 教師データを用意し、そこから正解を導くルールを導出する
特徴	<ul style="list-style-type: none"> ● 処理がわかりやすい ● 新たなルールに弱い（メンテが必要） 	<ul style="list-style-type: none"> ● 総合的に判断するため、表記ゆれや新たなルールにある程度対応できる
処理イメージ	<p>【電子カルテテキスト（例）】 HER2(+)</p> <p>【ルール（例）】 ([Hh][Ee][Rr][2])[¥t]*(¥(陽性 ¥(?¥+¥)?¥))</p> <ul style="list-style-type: none"> ✓ “Her2” や” HER2” はOK ✓ その後ろにスペース・タブがいくつあってもOK ✓ “陽性” もしくは” (+)” や” +” もOK ✓ ¥1と¥2で” HER2” と” (+)” のペアを取得 	<p>【電子カルテテキスト（例）】 HER2(+)</p> <p>【ルール適用結果（例）】</p> <p>HER2 …検査項目開始 (…検査結果開始 + …検査結果途中) …検査結果途中</p> <ul style="list-style-type: none"> ✓ 上から見て” HER2” と” (+)” のペアを取得

ご清聴ありがとうございました