

一般口演

一般口演9

医療データ分析4（ビッグデータ・ゲノム情報・がん登録等）

2018年11月23日(金) 16:00 ～ 17:30 I会場 (福岡サンパレスH平安（中継未広）)

[2-I-3-1] 過去の臨床研究を用いた汎用的な変数抽出モデルの開発

○豊島 雄太¹, 安藤 文彦¹, 荒木 賢二¹, 山崎 友義¹, 串間 宗夫¹, 小川 泰右¹, 松尾 亮輔¹, 下川 忠弘² (1.宮崎大学医学部附属病院医療情報部, 2.京都民医連中央病院)

近年、次世代医療基盤法に代表される法整備や、ナショナルデータベースの開放などにより、本邦における医療データ2次活用の推進体制が整ってきている。一方で、研究者のデータ分析を支援するデータサイエンティストの不足が問題視されている。本研究では、研究に用いる汎用的な変数の抽出モデルを作成し、データベース研究に不慣れな研究者やデータサイエンティストが簡便かつ迅速に研究用データの抽出を行うための仕組みを構築する。既存の臨床試験から複数の臨床研究概要を選択し、分析プロトコルを整理する。対象とする臨床研究は、宮崎大学附属病院において2017年に処方額上位となった薬剤を対象に、薬効別に JAPIC臨床試験概要の検索を行い抽出した。選択した臨床試験それぞれのプロトコルにおいて必要となる項目や指定する抽出方法、条件を洗い出し、多くのプロトコルにおいて共通する項目と固有に必要となる項目に分け体系的に整理した。多くの研究で共通する項目、指定方法について汎用的変数として設定し、抽出のためのモデルを構築した。11薬効、18薬品を対象とした臨床研究概要について確認を行ったところ、分析のため必要となる情報として123属性が確認され、大項目として19種類に区分された。このうち年齢や妊娠の有無、併用薬の有無等については多くの研究で必要とされる共通項目であった。各共通項目を汎用的変数として抽出モデルを設定した。複数の研究において共通的に必要となる項目が多くあり、事前に汎用項目の抽出モデルを設定することは研究を迅速に進める上で有効である。今後、汎用的な変数抽出モデルへ組み込む項目を増やし、データベース研究を迅速に行う体制を整備していくことが重要である。一方で、研究で必要となる項目は多岐にわたり全項目をカバーすることは出来なかった。研究者を支援するデータサイエンティスト育成も並行して進めていくことも重要と考える。

過去の臨床研究を元にした汎用的な変数モデルの開発

豊島 雄太^{*1}、安藤 文彦^{*1}、荒木 賢二^{*1}、山崎 友義^{*1}、串間 宗夫^{*1}、小川 泰右^{*1}、松尾 亮輔^{*1}、下川 忠弘^{*2}^{*1} 宮崎大学医学部附属病院 医療情報部、^{*2} 京都民医連中央病院Development of general-purpose variable model
based on past clinical researchYuta Teshima ^{*1}, Fumihiko Ando^{*1}, Kenji Araki^{*1}, Tomoyoshi Yamasaki^{*1}, Muneco Kushima^{*1}, Taisuke Ogawa^{*1},
Ryosuke Matsuo^{*1}, Tadahiro Shimokawa^{*2}^{*1} Department of Medical Informatics, University of Miyazaki Hospital, ^{*2} Kyoto Miniren Chuo Hospital

In recent years, due to legislation represented by the Next Generation Medical Infrastructure Act and release of the National Database, a secondary use of medical data is being promoted in Japan. On the other hand, the lack of data scientists that support researchers' data analysis is becoming problem. In this research, by creating an extraction model of general variables used for research, we will construct a mechanism to easily and quickly extract research data for researchers and data scientists unfamiliar with database research.

We have selected multiple clinical research outlines from existing clinical trials for 11 medicinal products 18 drugs. These became the top prescription amount in Miyazaki University Hospital in 2017 and we have organized the analysis protocol. We have extracted subject clinical studies by searching the JAPIC clinical trial outline and by PubMed search.

We confirmed 992 attributes as information required for analysis and classified it into 21 types as major items. Of these, age, disease name, drug prescription were common items required for many studies. We set an extraction model with common items as general purpose variables.

Since there are many items that are commonly required in multiple studies, setting extraction models for general-purpose items in advance is effective in advancing research quickly. In the future, it is important to increase the number of items to be incorporated into a general-purpose variable extraction model and to develop a system to conduct database research promptly.

Keywords: Medical Informatics Computing, Database management system, Statistical Computings

1. 緒論

近年、次世代医療基盤法の施行等による政府の後押しもあり、医療情報の活用の機運が高まっている。特に今回は、電子カルテデータや、医事請求データなど、研究のためにデザインされたデータではなく、日常診療で用いられるデータそのものの活用が注目されており、「Real World Data」、及び、「Big Data」という2つのWordに代表されるように、データの項目としての質とデータ量の両面での変革がある。

上記を受け、研究者のデータの活用ニーズが高まってきているが、一方でデータの活用できる項目が増え、取り扱うデータ量が増大することで、データの分析作業は複雑化し、研究者一人で結果まで到達することが難しくなっている。

そこで研究者と協働し目的とする結果を導き出すデータサイエンティストの重要性が高まっているが、一方でその不足が問題となっている。本研究では、データサイエンティストの不足を補うため、様々な研究において共通的に求められる分析について、簡易に分析結果を得るための定型化したモデルを作成する。

2. 目的

医療情報DBは、蓄積された医療情報を活用した後ろ向き研究や横断研究を行うためには、様々な軸を設定し、情報を加工、分析することが必要となる。前もって蓄積された医療情報の項目は多岐にわたり、細かい分析軸の設定が必要であり、様々な研究を行う度にそのつど設定していたのでは、データサイエンティストの負担も大きい。そこで、様々な研究において共通的に必要となる分析軸について整理を行い、汎

用的に活用できる変数モデルの作成を行う。モデルを活用することで、共通的に必要となる分析軸について簡易に結果を得られるようになり、データサイエンティスト、及び研究者の分析の負担軽減となることを目指す。

3. 方法

3.1 データ準備

既存の臨床試験から複数の臨床研究概要を選択し、分析プロトコルを整理する。対象とする臨床研究は、宮崎大学附属病院において2017年に処方額上位となった薬剤群から11薬効、18薬品を対象に、薬効別にJapicCTIの検索およびPubMedによる論文の検索を行い抽出した。対象として、合計58本の臨床研究を設定した。

No	著者名	表題	誌名	発行年 ほか
1	Lang JE, et al.	Lansoprazole Is Associated with Worsening Asthma Control in Children with the CYP2C19 Poor Metabolizer Phenotype	Ann Am Thorac Soc	2015 Jun;12(6):878-85.
2	Nishida T, et al.	Comparative study of esomeprazole and lansoprazole in triple therapy for eradication of Helicobacter pylori in Japan	World J Gastroenterol	2014 Apr 21;20(15):4362-9.
3	Takebayashi K, et al.	Effect of combination therapy with alogliptin and lansoprazole on glycemic control in patients with type 2 diabetes	Endocr J	2014;61(10):1031-9.

No	著者名	表題	誌名	発行年 ほか
4	Abe M,et al.	Efficacy of thrombomodulin for acute exacerbation of idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia: a nonrandomized prospective study	Drug Des Devel Ther	2015 Oct 23;9:5755-62.
5	Saito H,et al.	Efficacy and safety of recombinant human soluble thrombomodulin (ART-123) in disseminated intravascular coagulation: results of a phase III, randomized, double-blind clinical trial	J Thromb Haemost	2007 Jan;5(1):31-41
6	Hagiwara A, et al.	Can recombinant human thrombomodulin increase survival among patients with severe septic-induced disseminated intravascular coagulation: a single-centre, open-label, randomised controlled trial.	BMJ Open	2016 Dec 30;6(12)

種類に区分された。

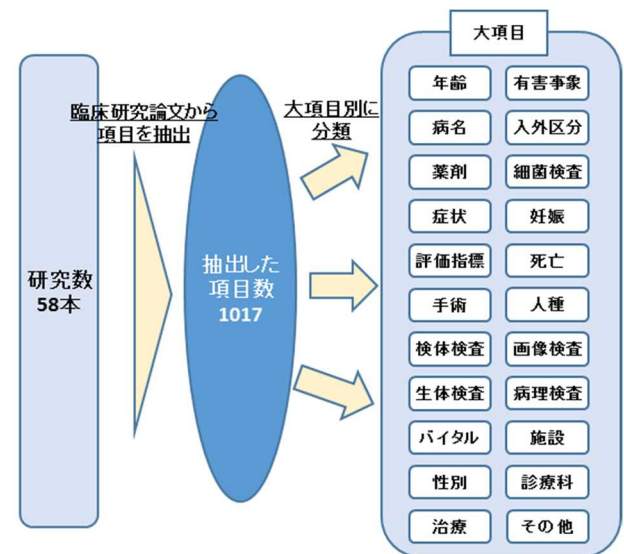


図1 設定イメージ

表1 論文一覧(PubMed からの抜粋例)

No	JapicCTI-No.	表題	試験実施者
1	JapicCTI-101061	AG-1749製造販売後臨床試験-非びらん性胃食道逆流症に対する臨床試験-	武田薬品工業株式会社
2	JapicCTI-050059	造血器悪性腫瘍あるいは感染症が直接誘因となり発症した汎発性血管内血液凝固症(DIC)患者を対象としたART-123の第3相試験	旭化成ファーマ株式会社
3	JapicCTI-132150	M801801の尋常性乾癬に対する第III相臨床試験	マルホ株式会社
4	JapicCTI-050174	YP-18の敗血症・心内膜炎を対象にした臨床試験	富山化学工業株式会社
5	JapicCTI-080577	BEATRICE: triple negativeの乳癌に対するR435(ペバシブマブ)術後補助化学療法の国際多施設共同非盲検2群間比較第3相試験	中外製薬株式会社
6	JapicCTI-060314	帯状疱疹後神経痛患者を対象としたPREGABALIN(CI-1008)の有効性および安全性を評価する、13週間の多施設共同プラセボ対照無作為化二重盲検比較試験	ファイザー株式会社

表2 論文一覧(Japic からの抜粋例)

3.2 項目整理

対象とした臨床試験それぞれのプロトコルにおいて必要となる項目や指定する抽出方法、条件を洗い出し整理を行った。

項目の整理により、今回対象とした研究では、分析のため必要となる情報として1017項目が確認され、大項目として22

3.3 体系化

整理した大項目ごとに、今回対象とした研究での項目としての採用状況を算出している。採用状況の高い大項目を臨床研究において共通的に必要となる項目として定義し、分析を行う上での条件を変数として洗い出した上で、体系化を行った。

4. 結果

4.1 項目の整理結果

今回対象とした臨床研究58本について、分析に必要な項目は1017項目が確認され、大項目として22種類に区分している。

整理した大項目ごとに、対象とした臨床研究での項目としての採用状況を調べると、大項目ベースで3種類、項目数で385項目が全58研究において共通的に必要とされていた。また、大項目ベースで19種類、項目数で942項目が複数の論文で必要とされていた。

今回、50%以上の研究で共通的に必要とされていた大項目5種類、項目数527項目について、汎用モデルの設定対象として設定した。

大項目	項目数	全体における割合	研究数	全体における割合
年齢	58	5.7%	58	100.0%
病名	207	20.4%	58	100.0%
薬剤	120	11.8%	58	100.0%
症状	74	7.3%	38	65.5%
評価指標	68	6.7%	30	51.7%
手術	28	2.8%	28	48.3%
検体検査	151	14.8%	27	46.6%
生体検査	36	3.5%	26	44.8%

バイタル	50	4.9%	20	34.5%
性別	18	1.8%	18	31.0%
治療	31	3.0%	18	31.0%
有害事象	17	1.7%	17	29.3%
入外区分	14	1.4%	14	24.1%
細菌検査	24	2.4%	11	19.0%
妊娠	13	1.3%	11	19.0%
死亡	14	1.4%	9	15.5%
人種	8	0.8%	8	13.8%
画像検査	9	0.9%	5	8.6%
病理検査	2	0.2%	2	3.4%
施設	1	0.1%	1	1.7%
診療科	2	0.2%	1	1.7%
その他	72	7.1%	—	—
合計	1017	100.0%	58	100.0%

表2 大項目別項目数・研究数

4.2 汎用モデルの設定

汎用モデルの設定では、変数として主要確認項目を設定し、加えてそれぞれの分析における条件を規定するための条件項目を設定した。例えば、年齢のモデルでは、主要確認項目として生年月日を設定し、条件項目として測定時を設定する。測定時は、更に固定と可変の分岐設定があり、固定では年齢を測定すべき日時が変数として設定される。例えば、研究のベースライン開始日時などが固定値として入力される。一方、可変では症例毎に測定すべき日時がイベントに連動して決定される。例えば、疾患の発症日や、治療の開始日などが該当し、それぞれの日時の決定では、病名や薬剤のモデルを用いることになる。

今回表2に示した大項目毎にモデルの作成を試みた。

大項目	主要確認項目	条件項目					
年齢	・生年月日	固定・可変・発症日・薬剤処方日					
病名	・疾患名	・起病時の有無・既往歴の有無	・通院加療の有無・滞在国		・主疾患かどうか・検査結果/判定・重症度	・手術可否・前治療の有無・出血の有無	
薬剤	・薬剤名	・処方日・頻度・処方順位・投与期間	・診療科		・用量・併用薬	・禁忌・前治療の有無・前治療の最終治療からの経過期間	
症状	未設定						
評価指標	未設定						
手術	・手術名	・実施日からの経過日数	・実施時の年齢	・過去に実施した手術			
検体検査	・検査名	・基準日からの日数	・年齢	・検査結果・バイタル結果・併発疾患		・薬剤使用有無・対象の治療有無	
生体検査	・検査名	・実施日・基準日からの日数・前治療からの実施期間	・出身地	・検査名		・設定された検査結果以上(以下)	
バイタル	・バイタル内容	・基準日からの経過日数		・バイタル結果		・薬剤使用有無・併用薬	

大項目	主要確認項目	条件項目					
性別	・性別						
治療	未設定						
有害事象	・有害事象	・発生日					
入外区分	・入外区分	・基準日からの経過日数・入院日・退院日・入院期間	・診療科		・入院理由	・緊急入院・再入院	
細菌検査	・検査内容	・基準日からの日数	・年齢	・検査結果			
妊娠	・妊娠	・時期				・出産の結果	
病理検査	・検査名	・実施日					
死亡	・転記	・基準期間内の死亡有無			・検査値が指定値以上(以下)		
人種	・人種・国籍						
画像検査	未設定						
施設	未設定						
診療科	・診療科						・主診療

表3 モデルの設定状況

今回の試みにおいて大項目レベルでは、変数の設定が多岐にわたり、個別設定が必要となった項目については未設定となった。それぞれの研究が求める分析に対応できる汎用モデルが17個作成できた。一方で、5個については、現時点で研究が求める分析に対応できず、汎用的なモデル作成ができなかった。

5. 考察

今回、対象とした臨床研究は薬剤をキーに検索を行い、内容もランダム化による前向き研究が多かった。そのため、項目の整理を行う上で、偏りが生じている可能性がある。例えば、薬剤をキーとしない研究で求められる項目や、比較する群間を規定する場合に用いる項目などが不足している可能性がある。モデルの適応範囲をより広げるには、今後、確認対象とする研究に、薬剤をキーとしない横断研究や後ろ向きコホート研究なども取り込む、より多くの研究を確認することが必要と考える。

モデルの作成では、すべての項目を汎用モデルとして設定できなかった。モデルの作成が容易な項目と、汎用的に作成することが難しい項目があった。特に病態や重症度を測るスコアなどの評価指標については、スコアの記述そのものを求める物もあれば、複数のデータを複合的に用い判定が求められる物もあり、同じ評価指標でも求め方が多岐にわたり、汎用モデルを設定することができていない。今後、確認する研究を増やすことで、多くの研究で共通的に求められる評価指標が特定されていくのであれば、大項目としてではなく、それぞれの指標毎にモデルを作成することも想定される。

今回、作成したモデルでは、研究者が分析プロトコルを設定する手助けや、データサイエンティストが研究者の要望を分析プロトコルに落とし込む際のツールとしての活用が想定されるが、それ以降の作業は残ってしまう。今後は、モデルがより汎用的に用いることができるよう適応範囲を広げるとともに、分析に用いるDBにあわせカスタマイズし、SQL等のコマンドを予め準備していくことが必要と考える。そうすることで、それぞれの変数を、研究者、もしくは支援するデータサイエンティストが指定するだけで、簡単な分析を完結することができるようになる。今後、大規模な研究用のデータベースが立ち上がることが想定されるが、それら今後稼動するであろうDBに臨床研究の基板として汎用モデルと、それを具体化するコマンド等をセットで組み込んでいくことが重要と考える。

ただ、論文で用いるすべての項目をモデルとして設定できてはならず、研究者の要望をすべてまかなうには、今後もデ

ータサイエンティストの支援が必要となる。

6. 結論

複数の研究において共通的に必要となる項目が多くあり、事前に汎用項目の抽出モデルを設定することは研究を迅速に進める上で有効と考える。今後、汎用的な変数抽出モデルへ組み込む項目を増やし、データベース研究を迅速に行う体制を整備していくことが重要である。一方で、研究で必要となる項目は多岐にわたり全項目をカバーすることは出来なかった。研究者を支援するデータサイエンティスト育成も並行して進めていくことも重要と考える。

参考文献

- 1) 一般財団法人日本医薬情報センター.医薬品情報データベース
iyakuSearch 臨床試験情報 JAPIC Clinical Trials Information,
[<http://www.clinicaltrials.jp/user/ctrSearch.jsp>].
- 2) National Center for Biotechnology Information, Pubmed.gov,
[<https://www.ncbi.nlm.nih.gov/pubmed/>].