

ポスター

ポスター6

医療データ解析・地域医療

2017年11月22日(水) 08:45 ～ 09:35 L会場（ポスター会場2）（12F ホワイエ）

[3-L-1-PP6-1] 病態・疾患特定のためのアウトカム定義（phenotyping）調査研究

伊豆倉 理江子¹, 山下 貴範¹, 野尻 千夏², 高田 敦史¹, 野原 康伸¹, 中島 直樹¹（1.九州大学病院 メディカル・インフォメーションセンター, 2.株式会社 ケア・フォー）

【はじめに】 Real World Dataである電子カルテ情報やレセプト情報、健診情報などを用いて、病名や病態を抽出する「Phenotyping」が試みられ、米国では Human Phenotype Ontologyや Phenotype Knowledge baseに集約されている。我が国では、特に薬剤疫学の分野において医薬品曝露による副作用の発現「アウトカム」を大量の症例から特定するための Phenotyping（＝「アウトカム定義」）活動が、AMEDや MIHARI Project、医療情報データベース基盤整備事業（MID-NET）などで進んでいる。特に MID-NETでは、平成30年度から製薬会社や研究機関などの利用が開始されるため、その整備は急務である。今回、AMED中島直樹班において、MID-NETでの利活用推進を主目的とし、各研究のアウトカム定義を調査・整理したので報告する。

【方法】 MIHARIや MID-NET事業、AMED及び医療情報学連合大会の演題項目の中で phenotypingに関する研究から、一定の疾患や病態に対するアウトカム定義を収集・整理した。

【結果と考察】 収集したアウトカム定義は計27個で、うち疾患の発現症例を抽出する定義が20件、特定の医薬品の副作用発現症例を抽出する定義が7件で、病態別では代謝・内分泌系が最も多かった。今回の整理により、急性腎不全や無顆粒球症のような検査値の異常や変動が（ほぼ）直接傷病名となっている疾患では検査値単独の定義で陽性的中率（PPV）が100%近くなる一方で、PPVが極めて低い定義の存在も判明し、これは疾患特性あるいは診断方法に依存することがわかった。今後は AMED宇山佳明班などでこれらのアウトカム定義の検討に機械学習を用いるなど工夫し、さらに多くの疾患・病態に対するアウトカム定義を整備して、精緻化していく必要がある。これらは、MID-NET利用者の抽出スクリプトの基礎的資料となり、抽出データの信頼性確保にも重要であると同時に、他 DB事業での phenotypingの重要な資料となる。

病態・疾患特定のためのアウトカム定義 (phenotyping) 調査研究

伊豆倉 理江子^{*1} 山下 貴範^{*1} 野尻 千夏^{*2} 高田 敦史^{*1} 野原 康伸^{*1} 中島 直樹^{*1}

^{*1} 九州大学病院 メディカル・インフォメーションセンター ^{*2} 株式会社ケア・フォー

The review of outcome definition (phenotyping) to detect diseases

Rieko Izukura^{*1}, Yamashita Yamashita^{*1}, Chinatsu Nojiri^{*2}, Atsushi Takada^{*1}, Yasunobu Nohara^{*1}, Naoki Nakashima^{*1}

^{*1} Kyushu University Hospital Medical Information Center, ^{*2} Care four, Inc.

Nowadays, “phenotyping”, the procedure for patient identification with specific diseases or pathological status by using EHR, administrative and claims data and information from physical examination, has been in widespread use. In US, various phenotyping algorithms are sorted out in “Human Phenotype Ontology” or “Phenotype Knowledge base”. In Japan, AMED, MIHARI Project and MID-NET Project have created the various outcome definitions throughout phenotyping. As the use of MID-NET system in public starts in FY 2018, it is extremely important to immediately sort out these outcome definitions. Therefore, we collected and reviewed the outcome definitions studied in these Project and the literatures of JAMI. Twenty-seven outcome definitions were found in this study. The largest number of disease or syndrome was those of “metabolism or endocrine system”. It was likely that the disease with the complicated pathological status or specific features reduces PPV unlike the disease named as an abnormal laboratory result. We need the further collection and fixing of the outcome definitions. Also we are going to improve these definitions by applying machine learning methods. Our efforts must be helpful for new MID-NET users and the other DB projects.

Keywords: phenotyping, outcome definition, outcome, adverse effects

1. 結論

「phenotyping」とは、研究の対象となる特定の病名や病態をもつ集団を自動的に検知・抽出する手法であり 1)2)、米国では、Real World Data である電子カルテ情報やレセプト情報、健診情報などを用いて、phenotyping 手法により生成した疾患アルゴリズムを Human Phenotype Ontology や Phenotype Knowledge base に集約している。わが国においても、特に薬剤疫学の分野において、医薬品曝露による副作用の発現「アウトカム」を大量の症例から抽出するための Phenotyping (「アウトカム定義」) 活動が、日本医療研究開発機構 (AMED) 事業や MIHARI プロジェクト、医療情報データベース基盤整備事業 (MID-NET) を中心に進んでいる。しかしながら、これらのアウトカム定義を統合し整理するまでには至っていない。特に MID-NET においては、平成 30 年度から製薬会社や研究機関などの第三者利用が開始されるため、その整備は急務である。

そこで、AMED 中島直樹班では、MID-NET を用いたアウトカム定義の策定における基盤整備の一環として、MID-NET 利用を目的とした既存のアウトカム定義を調査・整理したので報告する。

2. 目的

Phenotyping に関する研究事業から、一定の疾患や病態に対するアウトカム定義を収集・整理することを目的とした。

3. 方法

MID-NET 利用のためのアウトカム定義の生成を目的とした研究事業及び医療情報学会の演題より「アウトカム定義」を収集し、世界保健機関 (WHO) が作成する国際疾病分類 (ICD) 別に整理した。

アウトカム定義の生成においては、抽出ルールに従って対象症例を抽出した後、対象疾患の専門医が電子カルテを確認することにより真のアウトカムを発現した症例かどうかを判定する作業が必要であり、この作業を実施したアウトカムに

おいては定義の妥当性の指標として陽性的中度 (PPV) が算出されていた。

4. 結果と考察

4.1 アウトカムの属性

アウトカム定義数は、AMED 川上純一班で 7 個³⁾、田嶋尚子班 1 個⁴⁾、MIHARI Project で 4 個⁵⁾、医療情報データベース基盤整備事業 (MID-NET) 13 個⁶⁾、医療情報学会論文 2 個⁷⁾、計 27 個であった。

研究の対象疾患を国際疾病分類 10 版 (ICD10) 別に分類すると、「内分泌、栄養及び代謝疾患」が全体の 37% と最も多く、ついで「尿路器系の疾患」、「循環器系の疾患」の 11% であった (図 1)。

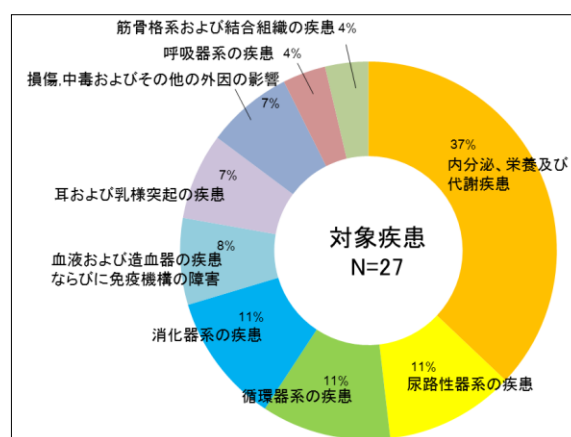


図 1 アウトカム定義の対象疾患の分類

分類別の内訳をみると、「内分泌、栄養及び代謝疾患」においては、高血糖や糖尿病のアウトカム定義の生成数が最も多く、糖尿病においては、I 型、II 型といった病態別や薬剤

性といった原因別のアウトカム定義があった(表 1)。

また、27 個のアウトカム定義の中には薬剤性肝機能障害など、医薬品曝露とアウトカムを関連づけたアウトカム定義も見受けられた。

表 1 アウトカム定義を生成した疾患の内訳

国際疾病分類名	病名	個数
内分泌・栄養および代謝疾患	高血糖	2
	糖尿病	1
	I 型糖尿病	1
	II 型糖尿病	1
	薬剤性糖尿病	1
	甲状腺機能亢進症	1
	脂質異常症	1
	低カルシウム血症	1
泌尿器系の疾患	急性腎不全	3
	薬剤性腎障害	1
循環器系の疾患	急性心筋梗塞	1
	高血圧	1
	頭蓋内出血	1
消化器系の疾患	消化管出血	2
	薬剤性肝機能障害	1
血液および造血器の疾患ならびに免疫機構の障害	無顆粒球症	2
損傷・中毒およびその他の外因の影響	アナフィラキシー	2
耳および乳様突起の疾患	聴覚障害	1
	難聴	1
呼吸器系の疾患	間質性肺炎	1
筋骨格系および結合組織の疾患	横紋筋融解症	1

4.2 抽出ルールによるアウトカム定義の妥当性の違い

表 2 において本研究で収集したアウトカム定義のうち、PPV の記載があるもののみを抜粋示した。急性腎不全においては、傷病名単一の条件とすると PPV が 65%程度であったが、血清クレアチニン値単一の条件にすると PPV が 100%近くまで上昇していた。傷病名で PPV が上がらない背景には、抗腫瘍薬であるシスプラチンの投薬に伴う有害事象を想定した傷病名の事前傷病名登録の実態が明らかとなっており⁸⁾、同様の研究で急性腎不全のアウトカム定義においても、傷病名と血清クレアチニン値を and 条件で組み合わせると PPV が 53.7%に留まっていた。このように検査値の異常やその変動が確たる診断要素になる疾患においては、検査値単一のアウトカム定義で真のアウトカム発現症例を高い精度で抽出できる可能性が高いことがわかった。無顆粒球症や低カルシウム血症においても検査値の異常がほぼ傷病名となっている疾患であることから、急性腎不全と同様に検査値単一のアウトカム定義が有効であると推察されるが、今回収集した両疾患のアウトカム定義結果からは明らかにできなかった。

また、病態が単純で聴力検査といった疾患特有の検査で診断される聴覚障害では、傷病名単一の条件のみで PPV が 81.6%と高かった。

一方で、アナフィラキシーにおいては、傷病名単一の条件で PPV が 20%以下と極めて低く、これは疾患特性あるいは診断方法に依存すると考えられた。他にも間質性肺炎や急性心筋梗塞など、病態が複雑な疾患においては傷病名だけでなく他のデータ種目との組み合わせを注意して検討していく必要性が示唆された。

糖尿病をアウトカムとした定義は 4 つあり、いずれも傷病名と糖尿病治療薬及び血糖値の 3 つのデータ種目を組み合わせた定義であった。I 型糖尿病においては、除外条件を詳細にすることで、症例数はかなり絞り込まれるが PPV が 81.8%に上昇していた。

本研究で収集した 27 個のアウトカム定義のうち PPV の算

出があったアウトカムは 15 個で、PPV の算出のないアウトカムにおいては、既存のアウトカム定義の妥当性の検証を行うとともに、機械学習を用いるなど工夫し、さらに多くの疾患・病態に対するアウトカム定義を整備して、精緻化していく必要がある。

5. 結論

本研究では収集した 27 個のアウトカム定義には、大きく医薬品曝露とアウトカムを関連付けた定義と疾患の特定に特化した定義の 2 種類があった。

また、急性腎不全や聴覚障害のような疾患特異的な検査で診断される疾患では、それぞれ検査値や傷病名単一の条件のみで PPV が 高くなる一方で、アナフィラキシーなど PPV が極めて低い定義の存在も判明し、これは疾患特性あるいは診断方法に依存する可能性が示唆された。

MID-NET システムでは複数の条件を組み合わせたアウトカム定義の作成が可能であり、特に初めて MID-NET を用いる利用者においては、対象の疾患や病態を抽出するためのデータ種別の選定やアルゴリズムの構成に困惑する可能性が高いため、ベースとなるアウトカム定義集の存在は大変貴重な基礎資料となり得る。

アウトカム定義は現在も収集中であり、今後も引き続き整理を行い、最終的にライブラリ化を目指して検討を重ねていく予定である。

6. 謝辞

本研究は AMED 事業「MID-NET を用いた医薬品等のベネフィット・リスク評価のための薬剤疫学研究等の実践的な分析手法及び教育に関する研究(番号:17mk0101064h0002)(中島直樹班)」によって行った。

また、川上純一班、田嶋尚子班、及び医薬品医療機器総合機構(PMDA)をはじめ、MID-NET 協力医療機関の関係者に対して、アウトカム定義の情報を提供いただいたことに、深く感謝申し上げます。

7. 参考文献

- 1) 香川璃奈, 河添悦昌, 井田有亮, 篠原恵美子, 今井健, 大江和彦. 特定患者集団の抽出(Phenotyping)手法確立に向けた技術的課題に関する考察, 人工知能学会 2016.
- 2) Jie X, Luke VR, Pamela LS, Guoqian J, Richard CK, Huan Mo, et al. Review and evaluation of electronic health records-driven phenotype algorithm authoring tool for clinical and translational research. J. Am. Med. Inform. Assoc. 2015.
- 3) 川上純一班「医薬品等の市販後安全対策のための医療情報データベースの利活用方法に関する薬剤疫学研究(16mk0101017h0003)」年次報告書. 医薬品等規制調和・評価研究事業, 2015-2016.
- 4) 田嶋 尚子班「1 型糖尿病の実態調査、客観的診断基準、日常生活・社会生活に着目した重症度評価の作成に関する研究(H28-循環器等一般-006)」年次報告書. 厚生労働科学研究費補助金(循環器疾患・糖尿病等生活習慣病対策政策研究事業), 2016.
- 5) 病院システムデータを利用したバリデーションスタディ報告書. MIHARI PROJECT 医薬品医療機器総合機構. 2014.
- 6) 医療情報データベース分析手法高度化のためのデータ検証(バリデーション)事業 年次報告書. 医療情報データベース整備基盤事業(MID-NET 事業), 2014-2016.
- 7) 香川璃奈, 河添悦昌, 篠原恵美子, 今井健, 大江和彦. 高血圧の phenotyping 手法の開発および他疾患との比較検討. 医療情報学 2016; 36: 770-773
- 8) 伊豆倉理江子 山下貴範 野尻千夏 野原康伸 安德恭彰 中島直樹. 医療情報データベース基盤事業の本格稼働に向けたデータ検証. 医療情報学 2014; 34: 710-713

表 2 アウトカム定義一覧（一部抜粋）

対象疾患	アウトカム定義			解析	引用	
	アウトカムを特定するデータ種目	条件式	主条件			
内分泌・栄養および代謝疾患						
糖尿病	傷病名 注射・処方情報 検査情報(結果値)	a and b and c	a: 病名がある b: 治療薬がある c: 初回発現月の前6か月～後1か月に、以下が2回以上存在する(ただし血糖値を必ず含む) ・空腹時血糖値: 126mg/dL以上 ・随時血糖値: 200mg/dL以上 ・HbA1c: 6.5%以上	-	PPV: 45.4 (a,bで抽出しcを判定基準にした場合)	MIHARI
I 型糖尿病	傷病名 注射・処方情報 レセプト診療行為	・主条件 (A集団): (a and (b or c or d or e)) or (b and c and d) ・除外条件 (B集団): 7 or 4 or 5 ・除外の例外条件 (C集団): I or O ファイナル条件式: A集団 - (B集団 not C集団)	a: 「1 型糖尿病」を含む病名(確定)がある b: インスリン処方がある c: 血中CPR検査で陽性 (0.6ng/mL未満)がある d: 「ケトアシドーシス」を含む病名の登録が期間中にある e: 「臓移植」を含む病名がある	(除外条件) 7: A集団の登録病名に、別表に示した「除外対象疾患リスト」に該当がある病名を抽出する。病名登録日を比較し(「1 型糖尿病」「除外対象疾患」が複数登録されている場合は、各々の最終日付を採用する。)、 「1 型糖尿病」が先行して、2 型糖尿病やその他の「除外対象疾患」が、同日を含めて後に開始された場合を除外対象とする。 4: 除外対象薬剤を「SU 剤」「グリニド剤」「DPP4阻害剤」とし、A 集団の対象期間中の処方データを抽出する。患者ごとの最終処方日を求め、最終処方日より前に「1 型糖尿病」病名が登録されている場合を除外対象とする 5: 「1 型糖尿病」病名が死亡以外で転帰している場合、除外対象とする。 (除外の例外条件) I: 血中CPR陽性が期間中に1度でもあれば、除外の取消対象とする O: 「臓移植」または「緩徐進行1 型糖尿病」があれば、除外の取消対象とする	PPV: 81.8 田嶋尚子班	
II 型糖尿病	傷病名 注射・処方情報	a and (b or c)	a: 「ICDコード上3桁がE12(栄養障害に関する糖尿病)、E13(二次性糖尿病)、E10(1 型糖尿病)の病名登録がない、または、抗GAD抗体の最大値が1.5U/mL未満、または、抗IA2抗体の最大値が0.4U/mL未満 b: 経口血糖降下薬の処方がある c: 2 型糖尿病の病名登録がある、かつ、異常検査値を少なくとも一度満たす	-	Accuracy: 87.54 Precision: 93.94 Recall: 46.97	医療情報学会(香川 et al, 2017)
薬剤性糖尿病	検査情報(結果値)	a and (b or c) × 2回 not (7 or 4 or 5 or I)	a: プレドニゾロンの新規処方がある b: 新規処方後に、空腹時血糖値 ≥ 126mg/dL がある c: 新規処方後に、HbA1c ≥ 6.5% がある	7: 新規処方前に、糖尿病治療薬の処方がある 4: 新規処方前の病名に、インスリン非依存性糖尿病(E11)、高血糖(R379)がある。疑いも含める。 5: 新規処方前に、空腹時血糖値 ≥ 110mg/dL がある I: 新規処方前に、HbA1c ≥ 6.5% がある	PPV: 52.4	川上純一班
甲状腺機能亢進症	傷病名 注射・処方情報 検査情報(結果値)	a and b and c	a: 病名がある b: 治療薬がある c: 初回発現月の前1か月～後1か月に、以下の面方を満たす ・遊離T4: 1.8ng/dL以上 or 遊離T3: 4.0ng/dL以上 ・TSH: 0.1 μU/mL以下	-	PPV: 66.7 (a,bで抽出しcを判定基準にした場合)	MIHARI
脂質異常症	傷病名 注射・処方情報 検査情報(結果値)	a and b and c	a: 病名がある b: 治療薬がある c: 初回発現月の前1か月～後1か月に、以下のいずれかを満たす ・LDLコレステロール: 140mg/dL以上 ・HDLコレステロール: 40mg/dL未満 ・トリグリセリド: 150mg/dL以上	-	PPV: 63.8 (a,bで抽出しcを判定基準にした場合)	MIHARI
低カルシウム血症	傷病名 注射・処方情報 検査情報(結果値)	a not (7 or 4 or 5 or I or O or 力)	a: 初回曝露日の翌日以降90日以内に血中カルシウム濃度 8.5mg/dL以下がある(アルブミン値補正済)	7: 初回曝露日より前6か月間に低カルシウム血症の既往がある(検査値、傷病名) 4: 初回曝露日より前6か月間にカルシウム製剤の処方がある 5: 初回曝露日より前に副中状腺機能低下症(ICD10E20)がある I: 初回曝露日より前に慢性腎不全(ICD10N18)がある O: 初回曝露日より前6か月間にビタミンD欠乏症(ICD10E55.9)がある 4: 初回曝露日より前6か月間にeGFR<30がある	PPV: 72.2	川上純一班

表 2 アウトカム定義一覧（一部抜粋）

対象疾患	アウトカムを特定するデータ種目	条件式	主条件	アウトカム定義 除外条件	解析	引用
尿路性器系の疾患						
急性腎不全	傷病名	a and b not 7	a: 対象期間内に急性腎不全の病名(疑いあり以外)がある b: aの病名の開始日の前後30日以内に入院日がある。	ア: aの病名の開始日の前日より以前に腎移植がある	PPV:65%	MID-NETバリ デーション
急性腎不全	検査情報(結果 値) レセプト診療行為	(a and b and c) not 7	a: 対象期間内に血清クレアチニン(Cr)値の異常値※がある b: aの血清クレアチニン値の異常値の日を含む前14日以内に血清Cr値の正常値※がある c: aの血清クレアチニン値の異常値の日の前後30日以内に入 院日がある ※血清Cr値の異常値: 男性: 2.0mg/dL以上、女性: 1.5mg/dL以 上 ※血清Cr値の正常値: 男性: 1.1mg/dL以下、女性: 0.7mg/dL以 下	ア: aの血清クレアチニン値の異常値の日の前日より以前に腎移植がある	PPV: 99.5%	MID-NETバリ デーション
急性腎不全	傷病名 検査情報(結果値)	a and b	a: 病名がある b: 初回発現月の前1か月～後1か月に、以下のいずれかを満た す ・血清クレアチニン値: 過去3か月以内の前値+0.3mg/dL 以上上昇 ・血清クレアチニン値: 過去3か月以内の前値の150% 以上上昇	-	PPV: 53.7 (診療録をゴールド スタンダードとした 場合)	MIHARI
循環器系の疾患						
高血圧	傷病名 注射・処方情報	a and (b or c)	a: 「二次性高血圧」「妊娠高血圧」「白衣高血圧」「高血圧性」ま たは、二次性高血圧を示す病名(例: 腎血管性高血圧)が登録 病名に含まれていない b: 高血圧の登録病名がある、ただし条件aに合致した登録病名 は除く c: ARB/ACE阻害薬の処方がある	-	Accuracy:71.56 Precision: 75.50 Recall: 75.50	医療情報学 会(香川 et al, 2016)
血液および造血器の疾患ならびに免疫機構の障害						
無顆粒球症	検査情報(結果値)	a and b and c and d and e (※)	a: 初回曝露日の翌日以降90日以内にANC<500/ μ Lがある 好中球(顆粒球)数: ANC b: 最終投与日+21日以内にANC<500/ μ Lがある c: ANC<500/ μ L時に、Hb $\geq$ 9g/dLかつPLT $\geq$ 10万/ μ Lがある d: ANC<500/ μ L後30日以内に、被疑薬の再投与無し e: ANC<500/ μ L後30日以内に、ANC $\geq$ 500/ μ Lがある ※主条件cは類似疾患を除外するための条件であるが主条件 に含めている。	-	PPV: 100	川上純一班
損傷,中毒およびその他の外因の影響						
アナフィラキ シー	傷病名	a not 7	a: アナフィラキシーの病名(疑いあり以外)がある	ア: 初回曝露日の-14日から+14日において除外病名(化学薬品、金属、生物等 による中毒を示す病名)がある	SS-MIX2病名の PPV:17 レセプト 病名のPPV:14	MID-NETバリ デーション
耳および乳様突起の疾患						
聴覚障害 (抗腫瘍薬)	傷病名	a	a: 被疑薬処方日の翌日から3か月以内に、聴覚障害の病名登 録がある。	-	PPV: 81.6	川上純一班