

小児慢性特定疾患データベースのリンケージと解析

研究分担者 森 臨太郎（国立成育医療研究センター政策科学研究部 部長）
研究協力者 盛一 享徳（国立成育医療研究センター小児慢性特定疾病情報室 室長）
西田 俊彦（国立成育医療研究センター政策科学研究部 研究員）

研究要旨

本研究は、各種データのリンケージ及び利活用の推進にあたり、主に小児慢性特定疾病登録データベースと他のデータベースとのリンケージに関する検討を行った。

平成 28 年度は、小児慢性特定疾病レジストリを縦断的に連結し、さらに他のデータベースと連結するための分析を行った。平成 29 年度は、相対的格差指標である Theil index を用いて、小児慢性特定疾病のデータ登録状況に地域間格差がないかを検討し、全般的に登録格差が少なく、慢性的に経過する内科的疾患ではとくに登録格差が大きくないことを示した。平成 30 年度は、小児慢性特定疾病登録データベースとレセプトデータベースとのリンケージを想定するにあたり、疾病名による関連付けが必須となることから、レセプト傷病名の機械的な類型化の可能性について検証を行い、自然言語解析の技術が解法の一つとなり得ることを示した。

A．研究目的

成育医療分野のあらゆる疾患の発症機序を解明し、長期予後改善に資するためにはさまざまな情報(データ)を広く用いて分析することがより効率的かつ効果的である。

本研究では、異なるデータセットの連携とさらにその連携を国際的なものにしていくための基盤を構築することを目的とし、主に小児慢性特定疾病登録データベースと他のデータベースとの連携に関して検討を行った。

B．研究方法

小児慢性特定疾病対策は、都道府県・指定市・中核市ごとに運用されている施策であり、運用主体の自治体を実施主体とよぶ。医療費助成の財源は国と実施主体が 1/2 ずつ支出することで運用されている。本研究は、登録状況が実施主体間に差異があるかを経済学で用いられている相対的格差指標である Theil index を用

いて検証を行った。利用データは電子化が完了している 2011 年度から 2013 年度までの小児慢性特定疾患治療研究事業における登録データ(以下、旧小慢登録データ)を用いて行った。

神奈川県内の国民健康保険におけるレセプトデータを用いて、レセプトデータベースにおける傷病名の登録状況を把握し、同一の疾患概念のレセプト傷病名を類型化することが技術的に可能であるかの検討を行った。

(倫理面への配慮)

本研究については、国立成育医療研究センター倫理審査委員会による倫理審査(受付番号：1637、1729)による承認を受けた。

C．研究結果

1. 小慢データと他のデータベースとのリンケージに関する検討

平成 27 年 1 月以降の新しい小児慢性特定疾病対策に係る登録データは、他のデータベース、

とくに出生届とのリンケージを視野において設計されていることから、当該登録データベースでは連結のための項目として、性別、生年月日、出生都道府県、出生体重、出生週数の5項目が利用可能となっていた。さらに他の疾病登録との連結に際しては、疾病名を利用した連結も可能である。旧小慢登録データに関しては、性別、生年月日の2項目が識別子として利用可能であり、さらに受給者番号、登録疾患名を利用することによりデータベース内部での経年的な接続を行うことが可能であった。

2. 小慢データの相対的登録格差に関する検討

Theil index は他の研究にて既に発症率が推計され、小児慢性特定疾病登録状況が比較的良好であると推定されている1型糖尿病の結果が基準値として見なせると考えた。

2011-13年における小慢対象年齢全体(0-19歳)の1型糖尿病のTheil index は、平均0.358 [95%CI 0.273-0.443]であった。

対象疾患群ごとに代表的な疾病についてTheil index の評価を行ったところ、急性リンパ性白血病 0.427 [0.350-0.503]、急性骨髄性白血病 0.380 [0.347-0.414]、ネフローゼ症候群 0.350 [0.269-0.430]、IgA 腎症 0.314 [0.247-0.381]、慢性肺疾患 0.492 [0.406-0.578]、ファロー四徴症 0.563 [0.466-0.660]、単心室症 0.417 [0.366-0.468]、成長ホルモン分泌不全性低身長症 0.376 [0.349-0.404]、甲状腺機能低下症 0.270 [0.223-0.318]、若年性特発性関節炎 0.312 [0.285-0.340]、2型糖尿病 0.309 [0.197-0.421]、フェニルケトン尿症 0.380 [0.362-0.397]、血友病A 0.403 [0.396-0.410]、免疫性血小板減少性紫斑病 0.390 [0.301-0.478]、ウエスト症候群 0.361 [0.318-0.404]、レノックス・ガスター症候群 0.714 [0.602-0.826]、先天性胆道拡張症

0.261 [0.226-0.296]、胆道閉鎖症 0.440 [0.409-0.472] という結果であった。

このうち1型糖尿病とTheil index の平均値に差が統計学的に認められなかったものは、急性骨髄性白血病、ネフローゼ症候群、IgA 腎症、成長ホルモン分泌不全性低身長症、若年性特発性関節炎、2型糖尿病、フェニルケトン尿症、血友病A、免疫性血小板減少性紫斑病、ウエスト症候群であり、1型糖尿病よりもTheil index が低値であったものは、甲状腺機能低下症、先天性胆道拡張症であった。

3. レセプト傷病名の類型化に関する検討

IRSI NLP Japanese により標準病名と自由記載された傷病名との比較・類型化を行った。その結果、以下の様な結果が得られた。

全く事前の辞書準備が無い状態で、日本語としての語句の切れ目を正しく理解するとともに、欠けている語句を補完しての認識、修飾語を伴う傷病名に対し、主たる語句を抽出して認識、長音の有無を含む語句の入れ替えの認識を行うことができた。

一方で、認識に誤りがあった場合も認められた。修飾語を主たる意味をもつものと誤認し判断しているものが多かった。多くは外傷や整形外科、耳鼻科、眼科の領域で使用される体の部位を主たる意味をもつと誤認するケースであった。

D. 考察

小児慢性特定疾病の対象疾患群の中で代表的な疾病について3年間のTheil index の平均値を計算し、基準とした1型糖尿病の値との比較を行い、全般的にはTheil index の値は小さく、登録格差は概ね少ないと考えられた。

自然言語解析の技術の一つであるIRIS NLP Japanese 利用により、これまで取扱が困難で

あったレセプト傷病名について、実現可能な作業量で、疾患概念ごとの類型化が行える可能性が示された。

E . 結論

小児慢性特定疾病の登録データは概ね実施主体間の登録格差は少なく、わが国を代表していると考えられた。また自然言語解析の技術の応用によりレセプトデータとのリンケージの可能性を見出した。

【参考文献】

- 1) InterSystems IRIS NLP Japanese の概要
Version 2.0. インターシステムズ・ホワイトペーパー .
- 2) Bronselaer A, et al. Concept-Relational Text Clustering. International Journal of Intelligent Systems. 2012;27:970-93.

F . 研究発表

1 . 論文発表

- 1) Nagata C, Moriichi A, Morisaki N, Gai-Tobe R, Ishiguro A, Mori R. Inter-prefecture disparity in under-5 mortality: 115 year trend in Japan. *Pediatr Int.* 2017;59:816-20.

2 . 学会発表

- 1) 森 臨太郎. 小児慢性特定疾病・特定疾病データベース改善のために(シンポジウム 11. 移行期医療を支える制度:小慢・指定難病制度:今とこれから). 第 120 回日本小児科学会学術集会(2017 年 4 月 16 日)
- 2) 盛一享徳, 森本康子, 柏崎ゆたか, 横谷進. 「小児慢性特定疾病登録の地域差に関する検討」. 第 121 回日本小児科学会学術集会(平成 30 年 4 月 22 日、福岡)

- 3) 盛一享徳. Natural Language Processing (NLP) を利用した病名収集の試み. 第 44 回日本診療情報管理学会学術大会(2018 年 9 月 20 日~21 日、新潟)

G . 知的財産権の出願・登録状況

なし

1 . 特許取得

なし

2 . 実用新案登録

なし

3 . その他

なし