

令和2年度 厚生労働行政推進調査事業費補助金
(医薬品・医療機器等レギュラトリーサイエンス政策研究事業)
分担研究報告書

医療情報データベースの活用推進に関する研究

研究分担者 中山 雅晴・東北大学大学院医学系研究科 教授

研究要旨：

目的：効率的なアウトカム定義の作成手法の検討及び実用化可能なアウトカム定義を確立するため、急性動脈解離(大動脈以外の解離も含む)を対象とした本研究を実施した。方法：初期条件では2018/1/8～2019/12/31にDPC病名がある患者とした。診療録を確認して真のケースかどうかを判定し、陽性的中度(PPV)を算出した。更に、機械学習によりPPVに影響する要因を検討し、PPV改善を目的に改良型アウトカムを作成した。結果：初期条件により抽出された641例中200例をランダムサンプリングして調査したところ、真のケースは62例、PPV31%であった。機械学習では急性大動脈解離等の病名が強く関連した。しかし、その他のケースである慢性大動脈解離も同じICD-10コードであり、区別が困難であった。A型大動脈解離は入院当日または翌日に手術またはニカルジピン注射の実施がある例を対象とし、入院当日または翌日に退院した例も対象とした。一方、動脈瘤では塞栓術がある例が多かったため、動脈塞栓術がある例は対象外とした。疑い病名も多く存在したため、対象外とした。これらの条件により、陽性的中度は85%まで改善し、感度も保たれた。結論：初期ルールではPPVが低かった。病名による条件の設定が困難であったため、標準的な治療に基づいて抽出条件を検討した。それにより陽性的中度は顕著に改善した。

A. 研究目的

製造販売後の医薬品安全性評価は、従来、副作用報告、使用成績調査等の結果に基づくことが主であったが、医療情報データベースの整備等によりリアルワールドでの大規模データに基づく評価が可能となりつつある。

MID-NET[®]は、厚生労働省の医療情報データベース基盤整備事業により構築されたデータベースで、独立行政法人医薬品医療機器総合機構（以下「PMDA」という。）による運用管理の下、平成30年度から運用を開始している。MID-NET[®]は、診療情報明細書（レセプト）、診断群分類（DPC）及び電子カルテの情報が利用可能で、現在では行政による活用の他、臨床研究や製造販売後データベース調査にも活用されており、製造販売後の医薬品安全対策の質の向上が期待されている。

医療情報データベースに基づく医薬品の安全性等の評価では、データベースに含まれる情報から対象とする有害事象（アウトカム）を適切に特定するために、信頼できるアウトカム定義を作成する必要があるが、本邦においてアウトカム定義のバリデーションが実施された例はほとんどない。「MID-NET[®]データの特性解析及びデータ抽出条件・解析手法等に関する研究」（日本医療研究開発機構 医薬品等規制調和・評価研究事業）（以下「先行研究」という。）において、MID-NET[®]を対象にアウトカム定義の作成及びその妥当性の評価を効率的に実施するための検討を実施し、基

礎的な検討手法を確立するとともに、バリデーションされた複数のアウトカム定義が作成された。しかしながら、実用化可能なアウトカム定義を増やすためには、更なる検討手法の効率化や具体的なアウトカム定義の作成を継続的に実施する必要がある。

本研究は、先行研究の成果を踏まえて、研究の流れを見直し All possible cases の定義を決定した上で以降の検討を行う等の検討手法の改善及び実用化可能なアウトカム定義の確立を目指し、医薬品安全性評価における医療情報データベースの活用促進と、より科学的な根拠に基づく安全対策の実現に繋げることを目的とした。

B. 研究方法

安全対策上の必要性や重要性を考慮し、検討の対象とするアウトカムを複数選定した上で、各アウトカムについて、従来法又は機械学習の手法を取り入れて作成したアウトカム定義について、複数医療機関にて妥当性の評価を行った。

研究の流れは以下のとおりである。

- 1) 複数拠点で検討対象とする All possible cases の定義及びアウトカム定義の検討
- 2) 評価基準の作成
- 3) 対象アウトカムについてカルテレビューにより真の症例を特定
- 4) 機械学習及び従来法によるアウトカム定義の

作成

- 5) 各アウトカム定義について、陽性的中度（以下「PPV」という。）及び感度の算出並びに評価
- 6) 複数医療機関の PPV の比較及び医療機関間の差異の要因検討

（倫理面での配慮）

医療機関が行う作業は、文部科学省・厚生労働省・人を対象とする医学系研究に関する倫理指針（平成 26 年文部科学省・厚生労働省告示第 3 号）を遵守して行った。

C. 研究結果

本研究では、アウトカム定義として、入院を要する急性の動脈解離（解離性動脈瘤も含む）を対象とした。対象病名は ICD-10 コードに基づいて選定した(表 1)。

表 1 対象傷病名・ICD-10 コード

| ICD-10 コード | 傷病名 |
|------------|---------------------|
| I710 | 大動脈の解離 [各部位] |
| I720 | 頸動脈瘤及び解離 |
| I721 | 上肢の動脈瘤及び解離 |
| I722 | 腎動脈瘤及び解離 |
| I723 | 腸骨動脈瘤及び解離 |
| I724 | 肢の動脈瘤及び解離 |
| I728 | その他の明示された動脈の動脈瘤及び解離 |
| I729 | 部位不明の動脈瘤及び解離 |
| S250 | 大動脈損傷 |

データ抽出対象期間は、2018 年 1 月 8 日から 2019 年 12 月 31 日とした。病名開始日は、DPC の入院日とした。

表 2 判定結果

| 動脈解離 | 真のケース | 疑われるケース | その他のケース | 判定対象ポテンシャルケース | |
|------|---------------|-----------------|---------------|---------------|----|
| | 件数 | 件数 | 件数 | 件数 | 件数 |
| | 62 | 1 | 137 | 200 | |
| 動脈解離 | 真のケース | 疑われるケース+その他のケース | 判定対象ポテンシャルケース | 陽性的中度 | |
| | 件数 | 件数 | 件数 | 割合 (%) | |
| | 62 | 138 | 200 | 31.0 | |
| 動脈解離 | 真のケース+疑われるケース | その他のケース | 判定対象ポテンシャルケース | 陽性的中度 | |
| | 件数 | 件数 | 件数 | 割合 (%) | |
| | 63 | 137 | 200 | 31.5 | |

All possible cases の抽出を目的とした初期ルールによって抽出された例は 641 例であった。このうち 200 例をランダムサンプリングして判定対象とした。200 例のうち、男性が 64.5%、平均年齢は 66.4±14.9 歳であった。

診療録等を確認して急性の動脈解離であるかを判定し、真のケース、疑われるケース、その他のケースに分類した。真のケースは 62 例、疑われるケース 1 例、その他のケース 137 例であり、PPV は 31.0%であった(表 2)。疑われるケースを真のケースとして扱っても、PPV は 31.5%であり、大きな変化は認められなかった(表 2)。

真のケースとなった例の ICD-10 コードは、I710(大動脈解離)が 58 例、I728(上腸間膜動脈解離)が 2 例、S250(外傷性大動脈損傷)が 2 例であり、I710 が多かった。

さらに、I710 に該当した 58 例のうち、43 例が Stanford A 型、13 例が Stanford B 型であった。また、58 例中 16 例が来院時心肺停止状態であった。

一方、その他のケースに該当した例の ICD-10 コードは、I710 が 40 例、I72x(大動脈以外の動脈解離・動脈瘤)が 96 例、S250 が 2 例であり、その他のケースにも I710 が多く存在した。

その他のケースとなった I710 のうち、急性期治療を要しない慢性大動脈解離や、大動脈解離の既往歴が 30 例であった。また、疑い病名が 10 例存在した。

I72x の 96 例は、大動脈以外の真性動脈瘤であった。これらは動脈解離と ICD-10 コードが同一であるため、抽出されていた。同様に、S250 についても、解離を形成しない大動脈損傷が抽出され、これらも解離を伴う大動脈損傷と ICD-10 コードが同一であった。

次いで全データを投入して機械学習を実施した。機械学習アルゴリズムとしては、勾配ブースティング(Gradient Boosting Decision Tree, GBDT)を用いた。具体的には、統計解析ソフト R バージョン 3.5.1 のパッケージの XGboost を用いて機械学習を行った。ROC 解析については、pROC パッケージを使用し、AUC を評価した。使用した変数は、合計 1937 変数であった(表 3)。AUC は 0.944 であった (図 1)。急性大動脈解離 Stanford A 型の病名が最も強く関連し、他にも急性大動脈解離 B 型、急性大動脈解離、慢性大動脈解離等の、大動脈解離の病名が強く関連した。病名以外では、医療機器安全管理料、ハンブの関連が強かった。動脈解離の治療で頻用される薬剤では、ニカルジピン注射液が比較的強く関連した。

表 3 機械学習に用いた変数

| データ種別 | 変数の数 | 備考 |
|-------------|------|-------------------------------------|
| DPC病名 | 349 | 主病名、契機、資源1番目、資源2番目、入院時併存、入院後発症、から抽出 |
| 処方・注射 | 880 | 名称は商品名を表示 |
| 手術 | 106 | |
| 管理料 | 16 | |
| その他のDPC請求項目 | 586 | |
| 計 | 1937 | |

入院当日または翌日に退院した例も、治療法にかかわらず抽出対象とすることとした。

一方、大動脈以外の真正動脈瘤が多く抽出され、その他の例となっていた。大動脈以外の動脈瘤では、動脈解離では実施しない治療法である動脈塞栓術が施行されている例が多く存在した。そこで、動脈塞栓術がある例は抽出対象から除外することとした。疑い病名も多く存在したため、抽出対象から除外することとした。以上から、表4のような改良型アウトカムを作成した。

表 4 改良型アウトカム

| 条件 | 条件の内容 |
|---------|-------------------------------------|
| ① | 動脈解離のDPC病名がある(必須) |
| ② | 疑い病名を対象外とする(必須) |
| ③ | 動脈塞栓術がある場合、対象外とする(必須) |
| ④ | 入院当日または翌日に、上行大動脈手術または胸部大動脈ステント挿入がある |
| ⑤ | 入院当日または翌日に、ニカルジピンの注射がある |
| ⑥ | 入院当日または翌日に、退院がある |
| 条件組み合わせ | (① and ② and ③) and (④ or ⑤ or ⑥) |

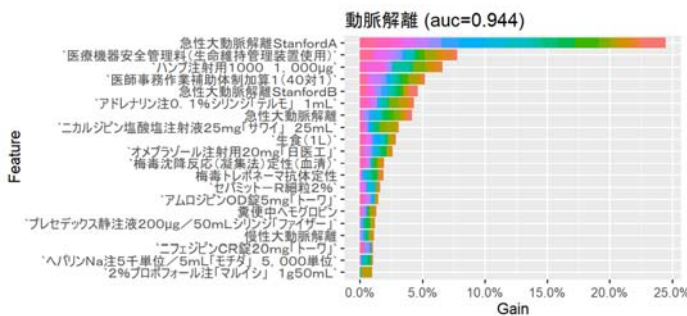


図 1 機械学習の結果

病名が強く関連したが、上述の通り、急性大動脈解離、慢性大動脈解離のいずれも ICD-10 コードは I710 であり、ICD-10 コードのみではそれらを区別するのが困難であった。そこで、病名以外の関連要因を抽出条件に加えることを検討した。

A 型大動脈解離と、B 型大動脈解離および大動脈以外の動脈解離では標準的な治療法が異なる。また、動脈解離による来院時心肺停止でも治療法は異なる。よって、それぞれの病態ごとに条件を検討した。

機械学習で関連が強かった医療機器安全管理料やハンプは大手術に関係する項目であり、まず手術に注目した。A 型大動脈解離では標準的な治療として、原則緊急手術を行うため、入院当日または翌日に手術があると予想された。一方、慢性動脈解離では待機的に手術を行うため、入院日から数日後に手術があると考えられた。そこで、入院当日または翌日に上行大動脈から胸部大動脈の手術またはステント挿入を実施した例を抽出対象とした。

B 型大動脈解離や大動脈以外の動脈解離では保存的に治療することが多く、その中でも血圧管理が重要となる。機械学習でニカルジピン注射が比較的関連したため、入院当日または翌日に投与した例を抽出対象とした。

大動脈解離はしばしば突然死の原因となり、心肺停止で来院した例も多く認められた。そこで、

改良型アウトカムにより、初期ルールで検討した 200 例中 73 例が抽出対象となり、真のケースは 62 例中全例抽出対象となった。PPV は 84.9% であり、ケース内感度 100%、特異度 92.0%、F 値 91.2% となった。

その他のケースとして残ったのは、大動脈以外の動脈瘤破裂による緊急手術、動脈解離を形成しない外傷による動脈損傷等であった。

D. 考察

初期ルールでは、急性動脈解離（大動脈解離および大動脈以外の解離を含む）の病名に関連する ICD-10 コードを条件として抽出した。しかし、PPV は 31% 程度と低く、多くの例がその他のケースとなっていた。真のケースには大動脈解離に関する ICD-10 コードである I710 が多く含まれたが、その他のケースでも I710 が少なくなかった。これは、大動脈解離の既往歴や、慢性大動脈解離も ICD-10 コードが同一であり、区別されないためであった。

また、疑い病名も含まれていた。血圧上昇を伴う急激な腰背部痛は急性大動脈解離の症状であり、急性大動脈解離を否定するため造影 CT を撮影することがある。このような場合に疑い病名が付与される。

大動脈以外の動脈解離である I72x に関しては、大半がその他のケースであった。これらは動脈解離と真性の動脈瘤が同じ ICD-10 コードであり、

大量に抽出された一方、多くがその他のケースとなっていた。しかしながら、真のケースも含まれていた。

機械学習でも急性大動脈解離 Stanford A 型をはじめとして、病名が強く関連した。対象病名を整理することで PPV の改善が見込まれた。しかしながら、大動脈解離はいずれの型や慢性大動脈解離でも ICD-10 コードが I710 であり、病名での区別が困難であった。

そこで病名以外の要因に注目した。機械学習で手術に関連する項目やニカルジピン注射との関連が示されたため、動脈解離の標準的な治療に注目した。A 型大動脈解離は緊急手術を行うことが原則であり、B 型大動脈解離および大動脈以外の動脈解離では保存的に治療することが多い。いずれも来院直後から治療を開始し、慢性大動脈解離では入院直後には手術しないため、入院当日または翌日に上行大動脈の手術や胸部大動脈のステント挿入がある例、またはニカルジピン注射の実施がある例を抽出対象とした。

大動脈解離は突然死の原因となることも多く、来院時心肺停止の例も多く認められた。そこで、入院当日または翌日に退院した例も抽出対象とした。

一方で、大動脈以外の動脈では、真性の動脈瘤が多く抽出されたため、PPV 低下の原因となった。大動脈以外の真性動脈瘤では動脈塞栓術がある例が多かった。動脈解離ではステント挿入によって治療することはあるが、塞栓術を行うことはない。そこで動脈塞栓術がある例は抽出対象外とした。疑い病名も多く存在したため、これも抽出対象外とした。これらの条件により、PPV は 85% まで改善が見込まれ、感度も保たれた。

本研究は急性大動脈解離を対象とした。動脈解離は病態や治療法が明確な疾患であり、抽出条件の設定が比較的容易であると考えられる。よって、単純に他の疾患に応用することはできない可能性がある。実際に、横紋筋融解症や深部静脈血栓症等では、抽出条件の設定が困難であった。しかしながら、これらは診断が難しい疾患であり、動脈解離と同様に標準的な診断、治療が確立されている疾患であれば、応用は十分に可能であると考えられる。

また、本研究は東北大学病院に入院した患者のデータのみを用いた。B 型大動脈解離や大動脈以外の動脈解離の治療では血圧管理が重要であり、降圧薬の注射が使用される。これにはカルシウム拮抗薬だけでなく、ニトロ化合物や β 遮断薬の注射も用いられる。しかし、本研究で使用されていたのはニカルジピン注射薬のみであった。東北大学病院ではニカルジピン注射薬が頻用されてお

り、他の医療機関では降圧薬の使用頻度が異なる可能性がある。医療機関の内情に応じて、注射薬を始め抽出条件の調整が必要と考えられる。

E. 結論

動脈解離の病名に基づいた ICD-10 コードのみを抽出条件とした初期ルールでは、PPV が低かった。真のケースとなりやすい病名とその他のケースとなりやすい病名がしばしば ICD-10 コードが同一であるため、ICD-10 コードによる条件の設定は困難であった。そこで手術や治療などの標準的な治療に基づいて抽出条件を再検討した。それにより PPV は顕著に改善した。

F. 健康危険情報：

(分担研究報告書では記入不要です)

G. 研究発表：

1. 論文発表

1. [Nakayama M](#), Takehana K, Kohro T, Matoba T, Tsutsui H, Nagai R. Standard Export Data Format for Extension Storage of Standardized Structured Medical Information Exchange. *Circulation Reports*. 2(10); 587–616. 2020.
2. Park J, Yamashita T, Takada A, Hotta T, Nojiri C, Izukura R, Fujimura Y, Kimura M, [Nakayama M](#), Ohe K, Orii T, Sueoka E, Suzuki T, Yokoi H, Kang D, Nakashima N. Development of Continuous Validation Model on Standard Codes Mapping for Multi-Institutional Collaborative Data-Driven Medical Study. *European Journal of Biomedical Informatics*. 16(3); 10–19. 2020.
3. 宮本 恵宏, 竹村 匡正, 竹上 未紗, 興梠 貴英, [中山 雅晴](#), 的場 哲哉, 小室 一成, 斎藤 能彦, 安田 聡, 宍戸 稔聡, 西村 邦宏, 平松 治彦, 上村 幸司, 辻田 賢一, 宇宿功市郎, 中村 文明. 電子カルテ情報をセマンティクス (意味・内容) の標準化により分析可能なデータに変換するための研究. 医療情報学 40(1) 32–33. 2020.

2. 学会発表

1. Prospective Randomized Trial of Telemedicine-based Collaborative Care Using A Prefectural Medical Information Network System. [Nakayama M](#), Inoue R, Miyata S, Shimizu H. AMIA 2020 VIRTUAL ANNUAL SYMPOSIUM. 2020/11/16, 国際, Poster.
2. Electronic phenotyping of heart failure from a national clinical information database. [Nakayama M](#), Inoue R. ESC Congress 2020 - The Digital

Experience. 2020/8/29-9/1, 国際, Poster.

3. 医療安全を担保するための Patient Summary 考. 中山雅晴. 第 40 回医療情報学連合大会 (第 21 回日本医療情報学会学術大会), 2020/11/22, 国内, 口頭.
4. Secondary Usage of Cardiovascular Clinical Data through Incorporation of a Standardized Data Format. Nakayama M, 第 84 回日本循環器学会学術集会, 2020/8/1, 国内, 口頭.
5. Safety-signal Detection For Liver Dysfunction Associated With A Drug : Pilot Studies Using Mid-net. Sawada S, Fujimura Y, Kimura M, Murata K, Nakashima N, Nakayama M, Ohe K, Orii T, Sueoka E, Suzuki T, Yokoi H, Ishiguro C, Yamada K, Itoh M, Uyama Y. 第 30 回 日本疫学会学術総会, 2020/2/21, 国内, Poster.
6. MID-NET を用いた急性冠症候群の検索精度に関する検討.井上隆輔、中山雅晴. 第 40 回医療情報学連合大会 (第 21 回日本医療情報学会学術大会), 2020/11/20, 国内, 口頭.

H. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得
2. 実用新案登録
3. その他