

一般口演

一般口演19

医療データ分析 7（DWH・臨床研究データベース）

2018年11月25日(日) 09:00 ～ 10:30 F会場 (5F 502+503)

[4-F-1-1] MID-NETを用いた心不全の検索精度に関する検討

○井上 隆輔¹, 大友 千晶¹, 中山 雅晴^{1,2} (1.東北大学病院メディカルITセンター, 2.東北大学大学院医学系研究科医学情報学分野)

【目的】医療情報データベース基盤事業において、東北大学病院に導入された医療情報データベース(以下 MID-NET)を用いて、薬剤疫学研究等を実施する際に適切な対象者の抽出が可能となるよう、抽出条件定義の妥当性を確認することを目的として、データ検証を行った。【方法】本研究では、アウトカムとして、心不全を定義し、MID-NETから抽出するための条件を設定した。入外を問わず、心不全の病名(疑いを含む)がある患者を対象とした。病名に加え、胸部 X線撮影の実施、心エコー検査の実施、脳性利尿ペプチド(BNP)検査の実施、利尿薬および循環器作動薬の投与の各条件が病名開始日の前後3ヶ月以内に行われた患者を抽出対象とした。MID-NETから抽出された該当者について、診療録を確認することにより、心不全ガイドラインに準じて作成した判定基準に基づき、真のケースとその他のケースに判定し、結果を基に陽性的中度 (PPV) を算出した。【結果】初期条件により2799例が抽出され、200例をランダム抽出して判定した。真のケースは70例であり、PPVは35%、感度は91.2%であった。真のケースであった70例のうち、40例(57.1%)が入院例であった。このため、入院が病名開始日から2週間の範囲内にあるケースが真である可能性が高いと推測し、対象を「病名開始日より2週間以内に入院した例」に限定したところ、200例中99例が該当した。この99例中、真のケースは55例であり、PPVは55.6%となった。さらに、機械学習を併用して BNPのカットオフ値と併用薬を調整することにより、PPV87.8%、感度69.7%となった。【結論】心不全の病名がつけられており、かつ真の心不全である例は全体の35%であり、病名の信頼性は低かった。入院日を病名開始日から14日以内に限定し検査値や併用薬を調整することにより、PPVの改善が認められた。その他の条件も最適化することにより、さらに PPVが改善すると考えられる。

MID-NET を用いた心不全の検索精度に関する検討

井上隆輔^{*1}、大友千晶^{*1}、中山雅晴^{*1,2}

*1 東北大学病院メディカル IT センター、*2 東北大学大学院医学系研究科医学情報学分野

Investigation on Retrieval Precision of Heart Failure with MID-NET

Ryusuke Inoue^{*1}, Chiaki Otomo^{*1}, Masaharu Nakayama^{*1,2}

*1 Medical Informatics Center, Tohoku University Hospital,

*2 Department of Medical Informatics, Tohoku University School of Medicine

[Purpose] To validate accuracy of extracting condition with MID-NET system in Tohoku University Hospital.

[Methods] We defined congestive heart failure (CHF) as an outcome and set conditions for extraction from MID-NET. Regardless of inpatients or outpatients, patients with heart disease name (including suspicious disease name) were targeted. In addition to disease name, each condition was included as extraction condition such as chest X-ray imaging, echocardiography examination, cerebral diuretic peptide (BNP) examination, diuretic drugs and cardiovascular drug administration within 3 months before and after the disease name start date. Based on the judgment criteria created according to the heart failure guideline, we confirmed the medical records for the patients extracted from MID-NET, and judged as a true case and other cases. Based on the result, a positive predictive value (PPV) was calculated.

[Results] 2799 cases were extracted according to the initial rule, and 200 cases were randomly extracted and judged. The true case was 70 cases, PPV was 35%, sensitivity was 91.2%.

Of the 70 true cases, 40 cases (57.1%) were inpatients. Thus, it was presumed that patients who were hospitalized within two weeks from the disease name start date is highly likely to be true cases. And the subjects were limited to "cases of hospitalization within 2 weeks from the disease name start date", 99 cases were applicable. Of these 99 cases, the true case was 55 cases, PPV was 55.6%.

Furthermore, by using machine learning, the cut-off value of BNP and concomitant medication were adjusted, resulting in PPV of 87.8% and sensitivity of 69.7%.

[Conclusion] Of 200 cases with name of heart failure, 35% of cases was true CHF and the reliability of disease name was low. Improvement of PPV was observed by limiting hospitalization date to within 14 days from the start date of disease name and by adjusting examination value and concomitant medication. By optimizing other conditions, it is considered that the PPV is further improved.

Keywords: MID-NET, heart failure, machine learning.

1.はじめに

医薬品の製造販売後の安全対策は従来、対照群を設定しない使用成績調査等の評価することで行われてきた。しかし、この方法には、いくつかの課題が知られており、より科学的な根拠に基づく安全対策を実施すべきとの意見もあることや、国際的にも医薬品の安全対策においてデータベースの活用が行われている事を踏まえ、本邦でも医療情報を二次利用可能な環境が整備されつつあるところである。

MID-NET は厚生労働省が主導して整備した医療情報のデータベースであり、レセプト及び DPC の情報に加えて、検査値も利用可能な点が特色である。平成 30 年度からは製薬企業等も MID-NET を利活用可能となり、医薬品の製造販売後調査に活用されることが想定されている。東北大学病院は、この MID-NET 事業の協力医療機関として指定されている。

データベースを利用する場合、データベースに含まれる情報を組み合わせて有害事象(アウトカム)を定義する必要がある。

データベースを利用した薬剤疫学研究を実施する場合には、より適切なアウトカム定義を作成することが、調査全体の質に影響する。本邦においてアウトカム定義のバリデーションはほとんど行われておらず、本研究では、MID-NET を用いて、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方法について検討した。具体的には、アウトカムとして心不全を定義し、アウトカム定義の妥当性の指標として陽性的中度(PPV: positive predictive value)を算出することを目的として実施した。

2.方法

本研究では、心不全を対象に、臨床ガイドライン[1]等を参考に心不全を発現した症例を網羅的に抽出するための条件を作成した。データ期間は 2013 年 4 月 1 日から 2013 年 12 月 31 日とした。初期の検索条件(初期ルール)は、表 1(巻末)にある、【病名 a の開始日、かつ病名 a の開始日の前後 3 ヶ月以内に、検査 a の実施日または検査 b の実施日または検査 c の実施日があり、かつ医薬品 a および医薬品 b の開始

日があること】とした。病名開始日は、SS-MIX2:診断日(開始日)、レセプト:診療年月、DPC:入院日とし、病名には疑い病名を含む。また、外来・入院は問わない。

MID-NETから当該条件を用いて抽出された症例のうち、ランダムサンプリングした 200 例について、医師が診療録等の記録を確認し、臨床ガイドライン[1]等を参考に作成した判定アルゴリズム(図 1)に基づいて真のケースを判定し、陽性的中度等を算出した。

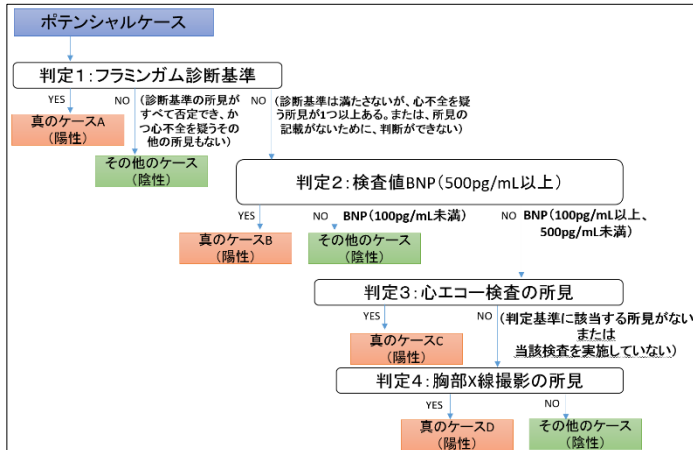


図1 心不全の判定アルゴリズム

また、初期ルールの感度分析のため、東北大学病院の循環器内科データを用い、2013年4月1日から2013年12月31日に循環器内科に入院したうっ血性心不全に該当する205例について検証した。

判定結果に基づき、真のケースが保有するデータに共通する要素を機械学習法により整理する方法を検討し、その機械学習法によりアウトカム定義の妥当性の改善に寄与する要素を特定した。機械学習法により特定された要素を組み合わせた改良型アウトカム定義を作成し、この定義の陽性的中度等を算出し、妥当性を評価した。

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 のパッケージ xgboost を用い機械学習を行った。ROC 解析については pROC パッケージを使用し AUC を評価した。

3.結果

病名 a に該当した例は 5282 例であった。このうち初期条件に該当した 2799 例が抽出され、200 例をランダム抽出して判定対象とした。真のケースは 70 例であり、PPV は 35%であった。

真のケースであった 70 例のうち、40 例(57.1%)が入院例であった。このため、入院が病名開始日から 1 ヶ月の範囲内にあるケースが真である可能性が高いと推測し、対象を「病名開始日より 1 ヶ月以内に入院した例」に限定したところ、200 例中 111 例が該当した。この 111 例中、真のケースは 56 例であり、111 例の中では PPV は 50.5%となった。

さらに、病名開始日から経過した時間ごとに入院数と真のケース数を調査したところ、0.5 ヶ月(14 日間)に大半の症例があり、PPV が 55.6%と高く、以後真のケース数の急激な増加はなく PPV も減少した(図 2)。対象を「病名開始日より 14 日

内に入院した例」に限定したところ、200 例中 99 例が該当した。この 99 例中、真のケースは 55 例であり、99 例の中では PPV は 55.6%となった。

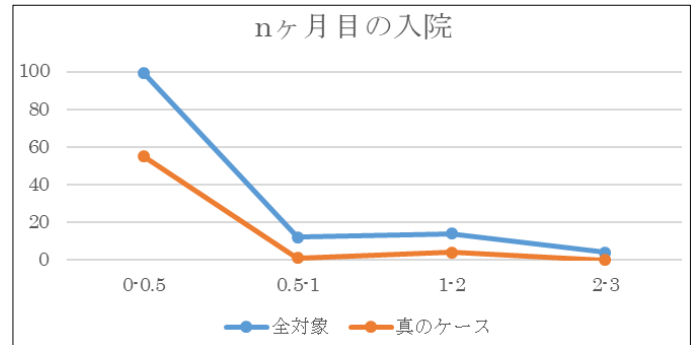


図2 病名開始日から n ヶ月経過した際の入院数

循環器内科データを用いた感度分析では、病名 a 該当が 193 例、初期ルール該当が 187 例であり、感度は 91.2%であった。偽陰性となった 18 例は明示的に心不全の病名が付与されず、心筋炎等の病名が付与されていたため、病名 a の付与がなかった例であった。

次に PPV を上昇させるために、機械学習を併用してアルゴリズムの見直しを行った。

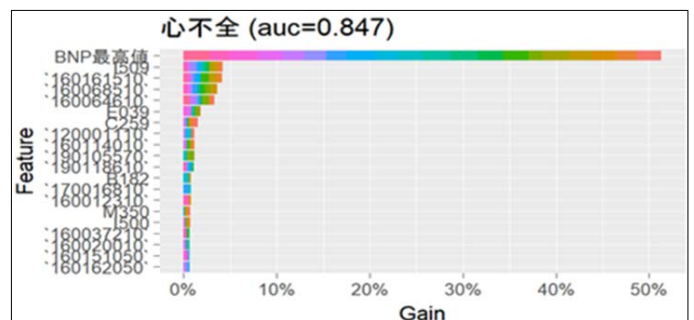


図3 機械学習の結果における BNP 最高値の寄与

まず、機械学習の結果から、脳性利尿ペプチド(BNP)最高値と心不全の相関は非常に強い(図 3)、BNP 最高値によって患者をグループ化し、それぞれの群における心不全の陽性適中率を検討した。すると BNP 値と心不全の陽性適中率の間に正の相関が認められ、BNP 値 400 以上は 80%以上が心不全であった(図 4)。逆に BNP 値 100 未満は心不全が 10%未満であった(図 4)。

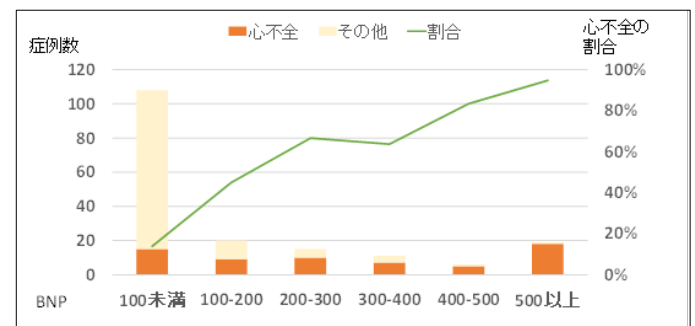


図4 BNP 最高値と心不全陽性的中率との相関

そこで、BNP 値 400 以上は他の条件によらず心不全、100 未満は心不全でないと定義し、100 以上 400 未満について標

標準治療をもとに検討した。すなわち、心不全治療で頻用される薬剤を条件に加え、再度機械学習を行った(図 5)。薬剤は 6 グループ(G)に分類し、カテコラミン(G1)、ヒト心房性利尿ペプチド(ハンプ)(G2)、利尿剤(G3)、スピロラクソン(G4)、アンジオテンシン受容体拮抗薬・アンジオテンシン変換酵素阻害薬・β 遮断薬群(G5)、ジギタリス(G6)とした(図 5)。

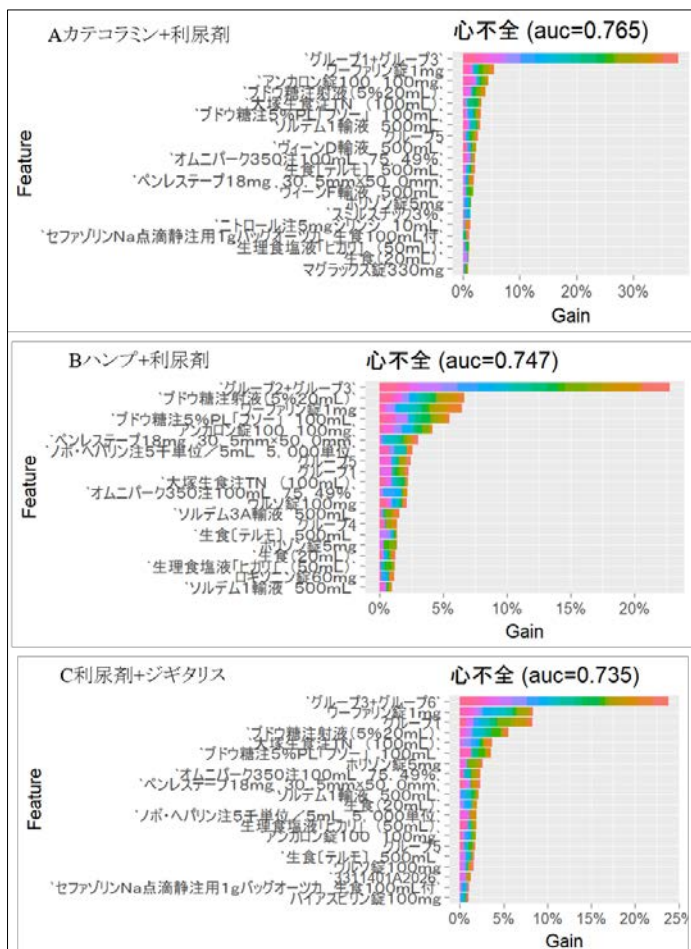


図 5 機械学習によるグループごとの心不全との相関

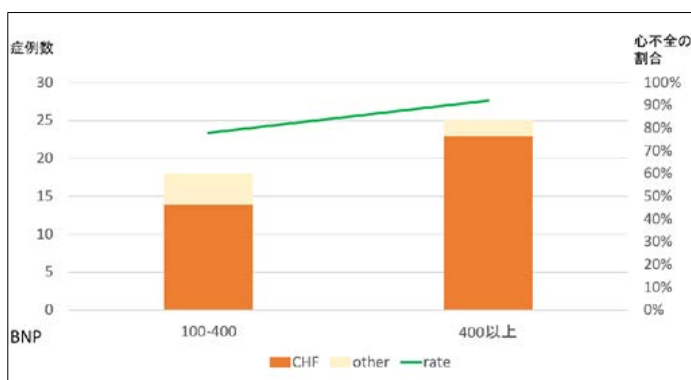


図 6 抽出した 43 症例の BNP と心不全陽性的中率

カテコラミン+利尿剤(G1+3)(図 5-A)、ハンプ+利尿剤(G2+3)(図 5-B)、利尿剤+ジギタリス(G3+6)(図 5-C)が、心不全との相関において上位となり、それら 3 条件を新たに加え改良型アウトカムとして再検討した。初期ルールに該当した 200 例中、BNP 値 100 以上で条件を満たすのは 43 例であり、

BNP100 以上 400 未満の群で陽性的中率の上昇が認められた(図 6)。

同様の条件で、2013 年 4 月 1 日から 2013 年 12 月 31 日の東北大学病院循環器内科独自データ 593 例で確認すると、PPV=87.8%(=196÷223×100)、感度=69.7%(=196÷281×100)であった(表 2)。

表 2 改良型アウトカム定義の該当者と真のケース等

		循環器内科独自データをゴールドスタンダードとした場合		合計
		真のケース	その他のケース	
改良型アウトカム定義	該当	196	27	223
	非該当	85	285	370
合計		281	312	593

4. 考察

心不全のガイドラインに基づいたアルゴリズムにより、初期ルールでも 90%以上の高い感度は得られた。一方で、PPV は 35%と低いため、上昇させるためのアルゴリズムを再考した。具体的には BNP のカットオフ値を調整することで、ある程度の正確な選択が可能となった。BNP 値が高い群で PPV は上昇しており、機械学習でも BNP は他の検査や処方と比較し、心不全に非常に高い相関を示したことから、患者群を BNP により分割し、選定条件を決定した。まず、BNP 値 400 以上は無条件に心不全とし、100 未満は心不全でないとした。BNP 値が 100 から 400 の間の症例に関しては、心不全に対する標準治療について分類を行った。

心不全の標準治療では、利尿剤が頻用される。また、しばしばハンプやカテコラミンが併用される。実際、機械学習でも相関が強い要素であった。さらに、内服薬では、アンジオテンシン受容体拮抗薬、アンジオテンシン変換酵素阻害薬、β 遮断薬が代表的な心不全薬であり、多くの症例で用いられる。加えて、心房細動を伴う例ではジギタリス製剤も選択される。それらも機械学習で心不全に関連の大きい因子として認められた。以上より、これらの薬剤をアルゴリズムに加えたところ、より真の心不全症例が抽出された。

今後、他の疾患を対象とする際も、機械学習の結果とともに、ガイドラインや実臨床における知見を参考にすることが重要である。

5. 結論

心不全も他疾患同様、病院情報システムにおける病名の正確性は低かった。入院日を病名開始日から 14 日以内に限定し、BNP 値や併用薬を調整することにより、PPV の改善が認められた。その他の条件も最適化することにより、さらに PPV が改善すると考えられる。

参考文献

- 1) 急性・慢性心不全診療ガイドライン(2017 年改訂版). 日本循環器学会/日本心不全学会. 2018.
<http://www.mhlw.go.jp/file/05-Shingikai-10901000-Kenkoukyoku-Soumuka/0000202651.pdf>.

表 1. 心不全の初期ルールに用いた条件

	データソース	条件に含まれる内容
病名 a	SS-MIX2/レセプト /DPC	I50 (心不全)、I110 (高血圧性心不全)、O291 (妊娠中の麻酔による心不全)、P290 (新生児心不全) など
検査 a	レセプト/DPC	胸部 X 線撮影 (実施)
検査 b	レセプト/DPC	心エコー検査 (実施)
検査 c	SS-MIX2	BNP (結果値) 100pg/mL 以上
医薬品 a	SS-MIX2/レセプト /DPC	ループ利尿薬、トルバプタンのいずれか
医薬品 b	SS-MIX2/レセプト /DPC	アンジオテンシン変換酵素阻害薬、アンジオテンシン受容体拮抗薬、 β 遮断薬、抗アルドステロン薬、強心薬、ヒト心房性利尿ペプチド、PDE 阻害薬、硝酸薬、カルシウムチャネル阻害薬のいずれか