

一般口演 | 医療データ解析

一般口演13 モデル構築・検証

2021年11月20日(土) 14:10 ~ 16:10 D会場 (2号館1階212)

[3-D-1-03] 深層学習で作成したリスクスコアを用いた治療効果比較研究

*松居 宏樹¹、康永 秀生¹（1. 東京大学大学院医学系研究科公共健康医学専攻臨床疫学・経済学）

*Hiroki Matsui¹, Hideo Yasunaga¹（1. 東京大学大学院医学系研究科公共健康医学専攻臨床疫学・経済学）

キーワード：Deep Learning, Administrative data, Causal Inference, Target trial emulation

【背景】観察データを用いた治療効果比較研究は、重要度を増している。精度の高い患者リスクスコアは、重症度調整を行う上で重要である。我々は過去、日本のDPCデータを用い、入院初日の医療行為と診断情報から院内死亡を深層学習で予測するモデルを構築した。モデルから得たリスクスコアは高い死亡予測精度と、実リスクとの対応を有していた。しかし、このリスクスコアを用いた治療効果比較研究は未だない。そこで、本研究では、構築したリスクスコアを用いて、心不全患者における早期食事開始が院内死亡率に与える影響を検証し、過去に同一データソースを用いて実施された研究と比較する。これにより、リスクスコアの可用性を検討した。

【方法】厚生労働科学研究DPCデータ調査研究班のDPCデータを用いた。2010年7月-2017年3月の期間中に対象病院(1492病院)にNYHA II以上の心不全で入院した20歳以上の患者をデータベースから抽出した。在院日数が2日以下の症例を対象から除外した。入院2日以内に食事を開始した症例を早期食事開始群、2日目以降に開始した症例を対照群とした。アウトカムを院内死亡とし、ロジスティック回帰を用いて、リスクスコアを調整した早期食事開始のオッズ比と信頼区間を計算した。重症度の高い症例と低い症例で早期食事開始の効果が異なると考えられたため、リスクスコアと早期食事開始の交互作用項をモデルに投入した。【結果】354820名を解析対象とした。モデルから計算された早期食事開始のオッズ比は0.49(95%信頼区間0.46-0.51)であった。早期食事開始とリスクスコアの間には強い交互作用が認められた(オッズ比32.2(95%信頼区間26.3-39.3))。【考察】得られた結果は、過去に実施された研究と矛盾しない結果であった。リスクスコアを用いた因果推論の実施可能性を示すものであると考えられた。

深層学習で作成したリスクスコアを用いた治療効果比較研究

松居宏樹^{*1}、康永秀生^{*1}

^{*1} 東京大学大学院医学系研究科公共健康医学専攻臨床疫学・経済学

A comparative effectiveness study using risk scores generated by deep learning

Hiroki Matsui ^{*1}, Hideo Yasunaga ^{*1}

^{*1} Department of Clinical Epidemiology and Health Economics, School of Public Health,
The University of Tokyo, Tokyo, Japan

This study used risk scores generated by deep learning to investigate the effectiveness of early dietary initiation on reducing in-hospital mortality in heart failure patients and compared the results with those in previous studies using the same data source. We extracted data on hospitalized patients aged 20 years or older with heart failure (NYHA II or higher) from the Diagnosis Procedure Combination database (July 2010-March 2017). We excluded patients who underwent surgery and those whose past history of hospitalization was unknown. We also excluded patients who underwent mechanical ventilation, intra-aortic balloon pumping, or extracorporeal membrane oxygenation, and those who started meals on the day of hospitalization. Patients who started eating on the second day of hospitalization were defined as the early dietary initiation group, and those who started eating on the third day or later were defined as the control group. The outcome was in-hospital mortality. The results from the deep learning method were compared with those in previous studies using conventional methods. In the 190,268 eligible patients, the early dietary initiation group showed a significantly lower mortality (adjusted odds ratio, 0.53; 95% confidence interval, 0.51-0.55). The results of this study indicate the feasibility of using risk scores for causal inference.

Keywords: Deep Learning, Administrative data, Causal inference, Target trial emulation

1. 緒論

近年、大規模な医療観察データを用いて、治療効果比較研究が実施される頻度が増えている[1]。これは、医療情報システムの発展に伴い、実際の医療現場のデータ(Real world data)の利活用が進んだことに起因している。

観察研究で治療効果比較研究が実施する場合は、治療群と対照群の患者背景や疾患特異的な重症度情報を基にリスク補正を行う必要がある。従来の臨床疫学研究においては、基本的な患者背景情報に加え、退院時サマリ情報やカルテ等から取得した疾患特異的な重症度情報を用いてリスク補正を実施してきた[2]。しかしながら、疾患特異的な重症度情報は、情報を入力する医師を始めとする専門家の労力や、調査自体にかかる費用の関係から無制限の収集は困難である。

我々は過去に、疾患特異的な重症度情報を利用することなく、日常診療記録から取得される入院初日の医療行為と診断情報等の医療管理データ(Administrative data)から院内死亡を深層学習で予測するモデルを構築した[3]。モデルから得たリスクスコアは高い死亡予測精度と実リスクとの対応を有しており、疾患特異的な重症度情報の代替としてリスク補正に使用できる可能性がある。しかし、このリスクスコアを用いた治療効果比較研究は未だ実施されていない。

2. 目的

本研究では、心不全患者に対する早期食事開始が院内死亡率に減少させる効果を検証するために、先行研究において深層学習を用いて構築したリスクスコアを用いたリスク補正を実施した。その結果を、疾患特異的な重症度情報(New York Heart Association functional classification (NYHA)による心機能分類)を用いたリスク補正方法を適用した先行研究の結果と比較することにより、深層学習を用いて構築したリスクスコアを用いたリスク補正方法の妥当性を検証する。

3. 方法

心不全患者において、カヘキシア等の栄養不良は生命予後を悪化させることが知られている。金子らは、過去に DPC データを用いて早期の食事開始が心不全入院症例の生命予後を改善させることを明らかにした[2]。当該研究では、NYHA A分類という疾患特異的な重症度情報を用いてリスク補正が行われた。

そこで本研究では、NYHA分類を用いることなく、深層学習を用いて構築したリスクスコアを用いて、心不全患者における早期食事開始が院内死亡率に与える影響を検証した。さらに両研究の結果を比較することにより、深層学習を用いて構築したリスクスコアの治療効果比較研究での可用性を検討した。

3.1 使用データベース

本研究では2010年7月-2017年3月の期間中の厚生労働科学研究 DPC データ調査研究班の DPC データを用いた。このデータベースは、全国の1492病院から収集された入院症例の、入院中の診療実績および、退院時サマリが記録されている。データベースに含まれている情報としては、入退院日や退院時転帰に加え、患者背景として、年齢・性別・body mass index (BMI)・過去の入院歴・喫煙歴、診断情報として、入院時病名・入院時併存症・入院後合併症、診療実績として入院中各日の入院中の診療行為・薬剤処方等、患者重症度として、入院時意識状態(Japan Coma Scale, JCS)・入院時の日常生活動作(Barthel index)・NYHA 分類等が含まれている。

3.2 研究デザイン

本研究では Hernan らの提唱する target trial emulation のフレームワークを用いて因果推論を実施した[1,4,5]。想定する target trial は、入院初日に食事を開始していない症例におい

て二日目に食事を開始するか、三日目以降に食事を開始するかを検討する前向き介入研究を想定し、観察データから仮想介入研究を可能な限り模倣(emulate)する。被推計量(estimand)として、入院時に食事開始を行わない症例における2日目食事開始の効果を治療企図分析(intention to treat analysis)により推計する。target trialの諸条件は以下とする。

1)入院初日(day 0)を起点とし、同時に適格症例を決定する。
2)入院2日目(day 1)に食事を開始する症例を早期食事開始群、入院3日目(day 2)以降に食事を開始する症例を非早期食事開始群としランダムに割り付ける。3)評価するアウトカムは割り付け完了後以降の死亡退院とする。

3.3 症例選択

データベースから、適格基準である1)NYHA II以上の心不全(ICD10コード:I500,I501,I509)で入院した症例、2)20歳以上の患者を抽出した。そのうえで、1)入院期間中の全身麻酔を伴う手術の実施症例、2)過去の入院歴が不明な症例を除外し、入院初日情報を用いて3)入院初日の人工呼吸、4)入院初日の大動脈内バルーンパンピング(intra-aortic balloon pumping, IABP)実施、5)入院初日の対外膜型人工肺(extracorporeal membrane oxygenation, ECMO)実施、6)入院初日の食事開始症例を除外とした。

3.4 研究に用いた変数と統計解析

Day 1での食事療養費(鼻腔栄養含む)の算定がある症例を早期食事開始群、ない症例を非早期食事開始群とした。アウトカムはDay 2以降の在院死亡とした。リスク補正として、Day 0における患者背景要因により補正を行った。既存研究の重症度情報を用いた傾向スコア(propensity score)もしくは、リスクスコア(risk score)を用いたマッチングを実施した。傾向スコアの計算にはロジスティック回帰モデルを用い、従属変数には早期食事開始、独立変数には年齢、性別、BMI(<18.5, 18.5-24.9, 24.9-29.9, ≥30kg/m², または欠損値)、6カ月以内の入院履歴、NYHA分類(II, III, または IV)、入院時 Barthel index (:<60, ≥60, または欠損値)、入院時 Barthel Indexにおける食事の自立(0, 5, 10, または欠損値)、喫煙歴(無しまたは有り)、JCS(0, 1, 2, 3, 10, 20, 30, 100, 200, 300, または欠損値)、週末入院、入院初日のβブロッカー・レニン-アンジオテンシン系阻害剤・ミネラルコルチコイド受容体拮抗薬・トルバプタンの投与、入院初日の強心剤・硝酸薬・フロセミド・カルペリチドの静注、入院初日の人工呼吸・血液透析の実施、入院初日の集中治療室入室、及び、高血圧症・糖尿病・慢性腎不全・慢性肝疾患・慢性呼吸器疾患・貧血・脳卒中・がん・心筋梗塞・拡張型心筋症の併存症の有無を用いた。リスクスコアの計算は、入院初日に取得可能な診療行為・病名情報、年齢と6カ月以内の入院履歴を用いて入院中死亡をアウトカムとした深層学習モデルを構築し、計算された予測死亡率(predicted mortality)をリスクスコアとした。交差検証の結果は良好であり、観測死亡率(observed mortality)と予測死亡率の対応を示す較正(calibration)の結果も良好であった[3]。マッチングは1:1マッチングをいずれも最近傍法(Nearest neighbor method)で行い、キャリパー(caliper)は標準偏差の1%とした。マッチング前後でそれぞれの群間の患者背景と重症度のバランスを標準化差(standardized mean difference, SMD)を用いて評価した。SMDの値は0.10以下でその変数のおおよその群間バランスが取れていると判断される[6]。マッチング前後でアウトカムの群間のオッズ比をFisherの正確確率検定を用いて計算した。

3.5 効果修飾の検出

リスクスコアを用いた効果修飾(effect modification)の検出の可否を検討するため、早期食事開始とリスクスコアの交互作用項を追加したロジスティック回帰モデルを構築した。

4. 結果

714,486症例がデータベースから抽出され、除外基準に従い524,222症例が除外され、190,268症例が解析対象となった(Fig.1)。119,949(63.0%)症例が早期食事開始群であった。Table 1にマッチング前の各群の患者背景を示す。早期食事開始群は非早期食事開始群に比較して、入院初日のリスクスコアが低く(早期/非早期:0.10/0.19)、平均年齢が低く(80.3歳/82.5歳) Barthel Indexにおける食事の自立の割合が高かった(42.2%/26.1%)。早期食事開始群ではNYHA IVの割合が低かった(早期/非早期 36.2%/49.7%)。傾向スコアマッチングにより両群61,203症例が、リスクスコアマッチングにより62,340症例がマッチングされた。マッチング前、傾向スコアおよび、リスクスコアでマッチングした後の各変数の群間のSMDをFig.2に示した。傾向スコアマッチングにより、マッチング前と比べ、傾向スコアの算出に用いたすべての変数はSMD<0.1となったものの、リスクスコアにおけるSMDが0.1以上であった(Fig.2)。リスクスコアマッチングにより、マッチング前と比べ、多くの変数のSMDは減少したものの、BMI、入院時 Barthel Index、JCSについて、SMDが0.1以上であった(Fig.2)。マッチング前、傾向スコアマッチ後、リスクスコアマッチ後の在院死亡に対する早期食事開始群のオッズ比はそれぞれ0.30(95%信頼区間 0.29-0.31)、0.50(95%信頼区間 0.48-0.51)、0.53(95%信頼区間 0.51-0.55)であった。早期食事開始とリスクスコアの交互作用項を追加したロジスティック回帰モデルでは、早期食事開始の主効果の係数が-1.35(95%信頼区間 -1.40,-1.31)であるのに対し、交互作用項の係数が3.19(95%信頼区間 3.02, 3.37)と強い交互作用を認めた。

5. 考察

本研究では、深層学習で構築したリスクスコアを用いたリスク補正と、退院時サマリ情報等から取得した疾患特異的な重症度情報を用いたリスク補正を行い、両者を比較した。比較を容易にするために、入院初日に食事を開始していない症例において二日目に食事を開始するか、三日目以降に食事を開始するかを検討する前向き介入研究をTarget trialとしたうえで、ベースラインの患者背景の補正を行うことでリスク補正を行った。いずれのリスク補正方法も、マッチング前の早期食事開始の効果推計値を縮小する結果となり、早期食事開始のオッズ比はマッチング前が0.30(95%信頼区間 0.29-0.31)に対し、マッチ後は0.50(95%信頼区間 0.48-0.51)、0.53(95%信頼区間 0.51-0.55)となり、早期食事開始は予後を改善していた。推定された治療効果は、リスクスコアを用いた補正でも、疾患特異的な重症度情報を用いた補正でも大きく変わらなかった。この結果は、疾患特異的な重症度情報の代わりに、日常診療記録から取得される入院初日の医療行為と診断情報を用いることで、治療効果研究を行う上で必須となるリスク補正を実施できる可能性を示唆していると考えられる。本研究で推定された心不全患者に対する早期食事開始の治療効果は、先行研究で推定された治療効果(オッズ比 0.75)と同じ方向を持つものの、そのサイズは先行研究より大きなものとなった[2]。これは、本研究と先行研究で模倣(emulate)しているtarget trialが異なることに起因すると考えられる。今回

提案したリスクスコアはその算出に入院初日の診療行為や処方を用いており、食事の摂取状況も含まれている。そのため、ベースラインの患者背景の補正に用いるためには、治療群への割り当ては Day 1 以降である必要があった。そこで、入院初日に食事を摂取していない心不全症例を対象を限定し、治療を Day 1 で食事を開始するかどうかにすることで、リスクスコアを用いたベースラインの患者背景調整を実施可能とした。しかしそのため、先行研究とは模倣(emulate)する target trial が異なっており、推定された治療効果が異なっていたと考えられる。

リスクスコアをリスク補正に用いる利点は複数ある。まず、退院時サマリ情報等から取得した疾患特異的な重症度情報を収集することはしばしば困難を伴う。例えば、国が保有するレセプト情報等データベース(NDB)は、そのデータの悉皆性は優れているものの、疾患特異的な重症度はデータベースに記録されない。今回示したように、日常的に収集される診療関連情報からリスクスコアを計算し、それを用いてリスク補正が可能であるならば、NDB 等の大規模データでの治療効果研究に寄与することが考えられる。また、モデルに投入する変数を減らすことで、モデル内での交互作用項の利用が解釈しやすくなる。今回明らかとなったように、心不全における早期食事開始は、死亡リスクが高い症例においては効果が小さくなる(もしくは逆転する)可能性が示唆された。先行研究においても、致命的な状態の患者においては栄養補給を遅らせることが推奨されている[7]。将来的には、モデルの精度を高めることで、治療選択の推奨を実施するシステムの構築が可能かもしれない。

本研究では傾向スコアマッチングとリスクスコアマッチングで共変量のバランスがマッチング前と比較してどの程度取れたかを SMD を用いて示した。結果として、マッチング前よりも共変量のバランスはいずれのマッチングでも取れていたものの、傾向スコアマッチングはリスクスコアのバランスが取れず、リスクスコアマッチングでは、傾向スコアに用いた重症度の一部のバランスが取れなかった。一部の共変量のインバランスが結果に影響を与える可能性が否定できない点は、リスクスコアマッチングの限界であると考えられる。

6. 結論

本研究では、深層学習で構築したリスクスコアを用いたリスク補正方法を、既存の臨床疫学研究で実施されたリスク補正方法と比較し、その妥当性を検証した。既存のリスク補正方法と同程度の結果が得られたことは、深層学習で構築したリスクスコアを用いたリスク補正方法の妥当性を示しているものと考えられる。

7. 文献

- Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* [Internet] 2016 Apr [cited 2021 Aug 25];183(8):758–764. PMID:26994063
- Kaneko H, Itoh H, Morita K, et al. Early Initiation of Feeding and In-Hospital Outcomes in Patients Hospitalized for Acute Heart Failure. *Am J Cardiol* [Internet] 2021 Apr [cited 2021 Jun 29];145:85–90. [doi: 10.1016/j.amjcard.2020.12.082]
- 松居宏樹, 康永秀生. 深層学習をレセプト記載情報に適用した入院予後予測モデルの構築と妥当性検証. 福岡県, 日本臨床疫学会 第三回学術集会(2019);
- Hernán MA, Sauer BC, Hernández-Díaz S, et al. Specifying a target trial prevents immortal time bias and other

self-inflicted injuries in observational analyses. *J Clin Epidemiol* [Internet] 2016 Nov [cited 2021 Aug 30];79:70–75. [doi: 10.1016/j.jclinepi.2016.04.014]

- Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. *BMJ* [Internet] 2018 Feb [cited 2021 Aug 30];360:k182. PMID:29419381
- Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* [Internet] 2009 Nov [cited 2021 Aug 31];28(25):3083–3107. PMID:19757444
- Reintam Blaser A, Starkopf J, Alhazzani W, et al. Early enteral nutrition in critically ill patients: ESICM clinical practice guidelines. *Intensive Care Med* [Internet] 2017 Mar [cited 2021 Aug 31];43(3):380–398. [doi: 10.1007/s00134-016-4665-0]

8. 図表

Table 1: 患者背景

	非早期食事開始群	早期食事開始群	SMD
N	70315	119949	
リスクスコア, 平均 (標準偏差)	0.19 (0.20)	0.10 (0.11)	0.579
年齢, 平均 (標準偏差)	82.45 (11.56)	80.30 (12.02)	0.182
女性, n(%)	38757 (55.1)	62746 (52.3)	0.056
Body mass index (kg/m ²), n (%)			0.241
< 18.5	15695 (22.3)	22299 (18.6)	
18.5-24.9	32540 (46.3)	61523 (51.3)	
24.9-29.9	8000 (11.4)	17467 (14.6)	
>= 30.0	6034 (8.6)	11632 (9.7)	
欠損値	8046 (11.4)	7028 (5.9)	
高血圧, n(%)	29061 (41.3)	61438 (51.2)	0.199
糖尿病, n(%)	16133 (22.9)	30490 (25.4)	0.058
慢性腎不全, n(%)	10062 (14.3)	17157 (14.3)	<0.001
慢性肝疾患, n(%)	1033 (1.5)	1813 (1.5)	0.003
慢性呼吸器疾患, n(%)	1833 (2.6)	3731 (3.1)	0.030
貧血, n(%)	5881 (8.4)	8754 (7.3)	0.040
脳卒中, n(%)	2409 (3.4)	2947 (2.5)	0.057
がん, n(%)	4423 (6.3)	6891 (5.7)	0.023
心筋梗塞, n(%)	2456 (3.5)	2875 (2.4)	0.065
拡張型心筋症, n(%)	1340 (1.9)	3493 (2.9)	0.066
喫煙歴, n(%)			0.070
無し	46251 (65.8)	74900 (62.4)	
有り	24063 (34.2)	45049 (37.6)	
欠損値	1 (0.0)	0 (0.0)	

	非早期食事開始群	早期食事開始群	SMD		非早期食事開始群	早期食事開始群	SMD
入院の既往, n(%)	20749 (29.5)	38913 (32.4)	0.063	血液透析, n(%)	724 (1.0)	1706 (1.4)	0.036
NYHA 分類, n(%)			0.278	集中治療室入室, n (%)	7693 (10.9)	12507 (10.4)	0.017
II	14064 (20.0)	32451 (27.1)					
III	21313 (30.3)	44095 (36.8)					
IV	34938 (49.7)	43403 (36.2)					
入院時 Barthel Index, n(%)			0.11				
< 60	62312 (88.6)	110303 (92.0)					
>= 60	7848 (11.2)	9399 (7.8)					
欠損値	155 (0.2)	247 (0.2)					
入院時の食事の自立, n(%)			0.49				
0	32742 (46.6)	31677 (26.4)					
5	11239 (16.0)	28067 (23.4)					
10	18331 (26.1)	50559 (42.2)					
欠損値	8003 (11.4)	9646 (8.0)					
Japan Coma Scale, n(%)			0.437				
0	46780 (66.5)	98136 (81.8)					
1	7661 (10.9)	11017 (9.2)					
10	2895 (4.1)	1679 (1.4)					
100	823 (1.2)	265 (0.2)					
2	3181 (4.5)	4154 (3.5)					
20	1059 (1.5)	462 (0.4)					
200	963 (1.4)	231 (0.2)					
3	4436 (6.3)	3443 (2.9)					
30	1040 (1.5)	388 (0.3)					
300	1475 (2.1)	173 (0.1)					
欠損値	2 (0.0)	1 (0.0)					
週末入院, n(%)	8231 (11.7)	13891 (11.6)	0.004				
β ブロッカー, n(%)	2405 (3.4)	5579 (4.7)	0.063				
レニン-アンジオテンシン系阻害剤, n (%)	5771 (8.2)	13875 (11.6)	0.113				
ミネラルコルチコイド受容体拮抗薬, n(%)	5840 (8.3)	12717 (10.6)	0.079				
トルバプタン, n(%)	1444 (2.1)	3306 (2.8)	0.046				
強心剤, n(%)	11316 (16.1)	11624 (9.7)	0.192				
硝酸薬, n(%)	13132 (18.7)	22741 (19.0)	0.007				
フロセミド, n(%)	44542 (63.3)	73986 (61.7)	0.034				
カルペリチド, n(%)	19839 (28.2)	39061 (32.6)	0.095				
人工呼吸, n(%)	52096 (74.1)	84208 (70.2)	0.087				

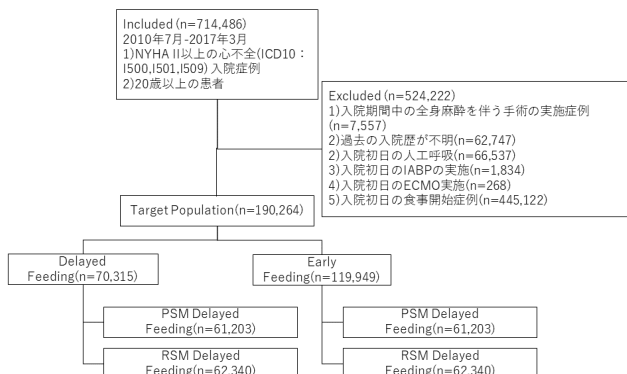


Figure 1: Patient flow diagram
症例選択とマッチングの流れ

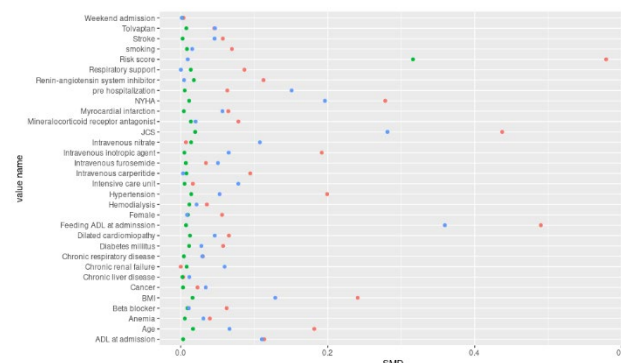


Figure 2: マッチング後の共変量のバランス