

厚生労働科学研究費補助金（難治性疾患等政策研究事業）
難治性の肝・胆道疾患に関する調査研究
分担研究報告書

スコアリングシステムおよびデータマイニング法による劇症肝炎の
予後予測モデルの再検証

分担研究者 持田 智 埼玉医科大学消化器内科・肝臓内科 教授
研究協力者 清水 雅仁 岐阜大学消化器病態学 教授

研究要旨：わが国の急性肝不全は昏睡型でスコアリングシステム4点以上が、Status Iとして脳死肝移植で優先されることになる。従って、最近の症例を対象に同システムとこれを補完するデータマイニングによる予後予測モデルのvalidationを行い、その精度を再評価しておく必要がある。そこで我々は、スコアリングシステムとデータマイニングによる予後予測モデルに関して、全国集計に登録の2010~16年発症の急性肝不全、LOHFのデータを用いて予測死亡率から、予後を判定して実際の転帰と照合、予測精度を2004~2008年の症例でのvalidation結果と比較した。スコアリングシステム5点以上を死亡の判定基準とし、2010~16年発症に適應すると、2010年以降の発症例では、感度およびNPVが低下していた。そこで、重度基礎疾患を有する症例とSIRS基準に合致し、かつ3つ以上の合併症を有する症例を除く65歳未満の症例に限定すると、正診率とNPVは、2004~2008年の発症例での検証結果とほぼ同様の精度であった。データマイニングによる予後予測モデルは50%以上の予後予測値を死亡の判定基準とし、2014~16年の症例に適應すると、BP判別と同数値以外では正診率に開発当時と大きな差異はなかった。しかし、NPVはBPの2モデルの他、決定木でも低下し、この低下は高精度のハイブリッド統合モデルにおいても同様であった。PPVはいずれのモデルでも高率に保たれていた。スコアリング法は最近の症例においても、原則的に肝移植の適應となる65歳未満で感染症など合併症のない場合の有用性は高いと考えられた。データマイニングによる予後予測モデルは、ドナー数の少ない現状に適合するようにNPVの向上を目指したモデルへの修正が今後の検討課題である。

< 共同研究者 >

末次 淳 岐阜大学第一内科 講師
内木 隆文 岐阜市民病院 中央検査部長
中山 伸朗 埼玉医科大学消・肝内科 准教授

A. 研究目的

厚労省研究班は「劇症肝炎の肝移植適應ガイドライン（日本急性肝不全研究会）」¹⁾に代わる予後予測法として多変量解析に基づくスコアリングシステムを発表した²⁾。しかし、このシステムによる予測死亡率は、スコアの合計が4点では50~60%であり、肝移植適應の判断が困難になるという問題があった。一方、我々はデータマイニング手法を利用して、1998~2003年発症の症例データに基づき、決定木法³⁾、radial basis function

（RBF）法およびback propagation（BP）法による4種類の予後予測モデル（基本モデル）を作成した。さらに4種類の抽出Cプログラムモデルを複合して、予測死亡率を算出するハイブリッド統合モデルも設計した。2004~08年に発症した劇症肝炎およびLOHFの症例⁴⁾でこれら5種類の予後予測モデルの検証を行ったところ、基本モデルでは「BP数値」モデルの正診率が最も高かったが、ハイブリッド統合モデルはより高い正診率を示すことを発表した。

わが国では今後、急性肝不全は昏睡型でスコアリングシステム4点以上が、Status Iとして脳死肝移植で優先されることになる。従って、同システムとこれを補完するデータマイニングのモデルを最近の症例を対象にvalidationを行い、その妥当性を再評価する必要がある。そ

ここで今年度は多変量解析に基づくスコアリングシステムとデータマイニングによる予後予測モデルに関して、最近の症例における有用性を検討した。

B. 方法

1) スコアリングシステムの検証

厚生労働省研究班の全国調査に登録された2010~16年発症の急性肝不全，LOHF計1,898例のうち，昏睡例かつ，肝移植非実施の711例（急性型405例，亜急性型260例，LOHF46例）を対象とした^{6, 7)}。6項目の合計スコア別に死亡率を算出し，実際の転帰と照合，予測精度を2004~2008年の症例でのvalidation結果と比較した。

2) データマイニングによる予後予測モデル：基本モデルおよびハイブリット型統合モデルの検証

厚生労働省研究班の全国調査に登録された2014~16年発症の急性肝不全，LOHF計838例のうち，65歳以下の昏睡例かつ肝炎例で，肝移植非実施の122例（急性型66例，亜急性型46例，LOHF10例）を対象とした^{6, 7)}。II度以上の昏睡出現時の所見を基に，決定木法，radial basis function (RBF) 法，back propagation (BP) 法，これらを統合したハイブリットモデルで予測死亡率を算出，50%以上を死亡予測と判定して実際の転帰と照合，予測精度を2004~2008年の症例でのvalidation結果と比較した。

C. 成績

1) スコアリングシステムの検証

2010~16年発症の急性肝不全，LOHF計1,898例のうち，昏睡例かつ，肝移植非実施の711例（急性型405例，亜急性型260例，LOHF46例）のうち，442例（救命率33.3%）でスコアリング法での解析が可能であった。

死亡率はスコア計5点では81.9%（59/72），計4点では57.3%（43/75），計3点では53.3%（40/75）であった（図1）。スコアが計5点以上を死亡予測と仮

定すると，2010~2013年発症例の268例における正診率は68%，感度，特異度は61%と80%，positive predictive value (PPV) と negative predictive value (NPV) は49%と68%，2014~2016年発症例の171例では，それぞれ順に66%，54%，91%，93%，49%であった（表1）。2004~2008年発症例の正診率77%，感度74%，特異度81%，PPV86%，NPV65%と比較すると，2010年以降の発症例では，感度およびNPVが低下していた。4点未満の症例に，高齢の基礎疾患を有し，複数の合併症を発症する予後不良な症例が少なからず含まれ，正診率が低下していると考えられた（表2）。

そこで，2010~16年発症の急性肝不全，LOHFで，昏睡例かつ肝移植非実施でスコアリング法の解析可能な442例から，心不全，腎不全，固形癌など重度基礎疾患を有する症例とSIRS基準に合致し，かつ3つ以上の合併症を有する症例を除いた225例でvalidationを行った。スコアが計5点以上を死亡予測と判定すると，正診率71%，感度61%，特異度86%，PPV87%，NPV60%，4点以上の死亡予測判定で，それぞれ73%，79%，65%，76%，68%となり，正診率とNPVは改善した（表3）。さらにvalidationに用いる対象症例を65歳未満の162例に限定すると（図2），正診率75%，感度60%，特異度87%，PPV78%，NPV74%，4点以上の死亡予測判定で，それぞれ73%，81%，67%，66%，83%となり，正診率とNPVはさらに改善した（表4）。

2) データマイニングによる予後予測モデル：基本モデルおよびハイブリット型統合モデルの検証

Validationに用いた症例の救命率は全体で41.0%（50/122），急性型53.0%（35/66），亜急性型28.3%（13/46），LOHF20.0%（2/10）であった。これらの症例で予後予測を実施すると，決定木モデルの正診率は70.5%（感度69.4%，特異度72.0%，PPV78.1%，NPV62.1%）で，以前の正診率71.2%（それぞれ75.0%，67.1%，70.6%，71.8%）で正診率と特異度に差はなかったが，感度とNPVがやや低下していた（表5）。RBFモデル

は正診率 72.1% (73.6%, 70.0%, 77.9%, 64.8%) で以前は 69.4% (71.0%, 67.8%, 70.2%, 68.6%), BP 判別モデルは 59.8% (55.6%, 66.0%, 70.2%, 50.8%) と 71.1% (76.3%, 65.5%, 70.3%, 72.2%), BP 数値モデルは 68.0% (63.9%, 74.0%, 78.0%, 58.7%) と 75.6% (75.3%, 75.9%, 76.9%, 74.2%), ハイブリッドモデルは 73.0% (70.8%, 76.0%, 81.0%, 63.3%) と 77.2% (82.8%, 71.3%, 75.5%, 79.5%) で, RBF モデルでは差異は見られなかったが, BP 判別と数値モデルでは正診率, 感度, NPV がそれぞれ低下し, ハイブリッドモデルも感度と NPV がそれぞれ低下していた。一方, PPV はいずれのモデルでも高率のままであった。

D. 考案

2010~16 年発症の急性肝不全, LOHF の非移植例データを用いて, スコアリングシステムを検証した結果, 2004~2008 年の発症例に比して, 感度と NPV が低下していることが明らかとなった。しかし, validation に用いるデータを, 重度基礎疾患を有する症例と SIRS 基準に合致し, かつ 3 つ以上の合併症を有する 65 歳未満の症例に限定すると, 正診率と NPV は, 2004~2008 年の発症例での検証結果とほぼ同様の精度であった。多くの施設で, 肝移植の適応となるのは, 主に 65 歳未満で感染症など合併症のない症例であり, これらの症例においては, 現状のままでスコアリング法の有用性は十分であると考えられた。

データマイニング法の 4 種類の基本モデルを最近の症例で検証すると, BP 判別と数値モデル以外ではいずれも 2004~2008 年の発症例での検証と同様に正診率が高率であるが, NPV は BP の 2 モデルの他, 決定木でも低下が見られた。その結果, 基本モデルを統合したハイブリッド型統合モデルにおいても, 2004~2008 年発症例の検証と比較して, 最近の症例の検証では NPV が低下していた。

ドナー数が多い場合には「死亡予測例が死亡する率である PPV」に優れる予後予

測モデルの有用性が高いが, ドナー数が少ない現状では「生存予測例が生存する率である NPV」が高率のモデルが理想である。今後, データマイニング法の基本モデルおよびハイブリッド統合モデルにおいて特に NPV の向上を目標にして, モデルの修正を検討する必要がある。

E. 結語

スコアリング法は 2010~16 年発症の症例においても, 原則的に肝移植の適応となる 65 歳未満で感染症など合併症のない場合の有用性は高いと考えられた。

データマイニングによる予後予測モデルは 2014~16 年の症例に適応すると, BP 判別と数値モデル以外では正診率に開発当時と大きな差異はなかった。しかし, NPV は BP の 2 モデルの他, 決定木でも低下し, 高精度のハイブリッド統合モデルにおいても影響がみられた。ドナー数の少ない現状に適合するように NPV の向上を目指したモデルの修正が今後の検討課題である。

F. 参考文献

1. Mochida S, *et al.* Re-evaluation of the Guideline published by the Acute Liver Failure Study Group of Japan in 1996 to determine the indications of liver transplantation in patients with fulminant hepatitis. *Hepatol Res* 2008; 38: 970-979.
2. Naiki T, *et al.* Scoring System as a Useful Model to Predict the Outcome of Patients with Acute Liver Failure: Application to Indication Criteria for Liver transplantation. *Hepatol Res* 2012; 42: 68-75.
3. Nakayama N, *et al.* Algorithm to determine the outcome of patients with acute liver failure: a data mining analysis using decision trees. *J Gastroenterol* 2012; 47: 664-677.

4. Oketani M, et al. Etiology and prognosis of fulminant hepatitis and late-onset hepatic failure in Japan: Summary of the annual nationwide survey between 2004 and 2009. *Hepatol Res* 2013; 43: 97-105.

5. 中山伸朗, 持田 智. 肝移植適応ガイドラインの改変 -高精度の予後予測システムの開発-. 日本消化器病学会雑誌 2015; 112: 28-35.

6. Nakao M et al. Nationwide survey for acute liver failure and late-onset hepatic failure in Japan. *J Gastroenterol* 2018; 53: 752-769.

7. 持田 智, 中山伸朗. 我が国における急性肝不全および遅発性肝不全 (LOHF) の実態 (2016年): 平成29年度全国調査. 厚生労働省科学研究費補助金 (難治性疾患等政策研究事業) 「難治性の肝・胆道疾患に関する調査研究」平成29年度総括・分担研究報告書, 2018; pp85-104.

G. 学会発表

1. Nakayama N, *et al.* Factors Associated with Favorable Outcome in Patients with Acute Liver Failure and Late-Onset Hepatic Failure following Liver Transplantation Evaluated Based on Nationwide Surveys in Japan. *The 69th annual meeting of the American association for the study of liver disease (AASLD)*, San Francisco, 2018; Nov.

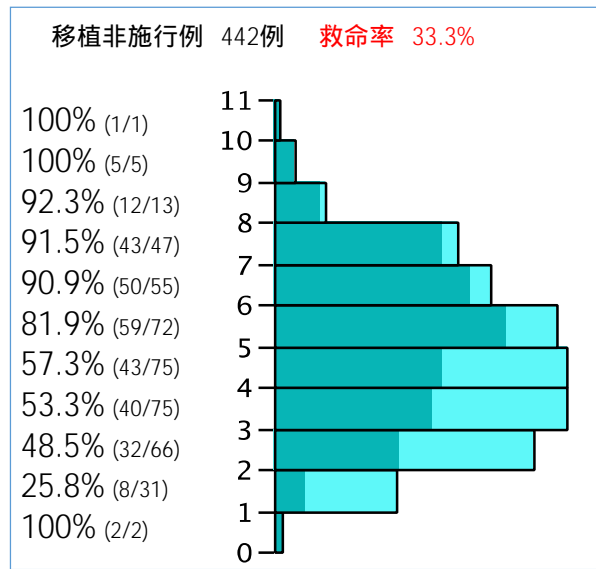


図1. 2010-2016 急性肝不全(昏睡型) 死亡率 (スコアリング解析可能例 580症例(64.8%))

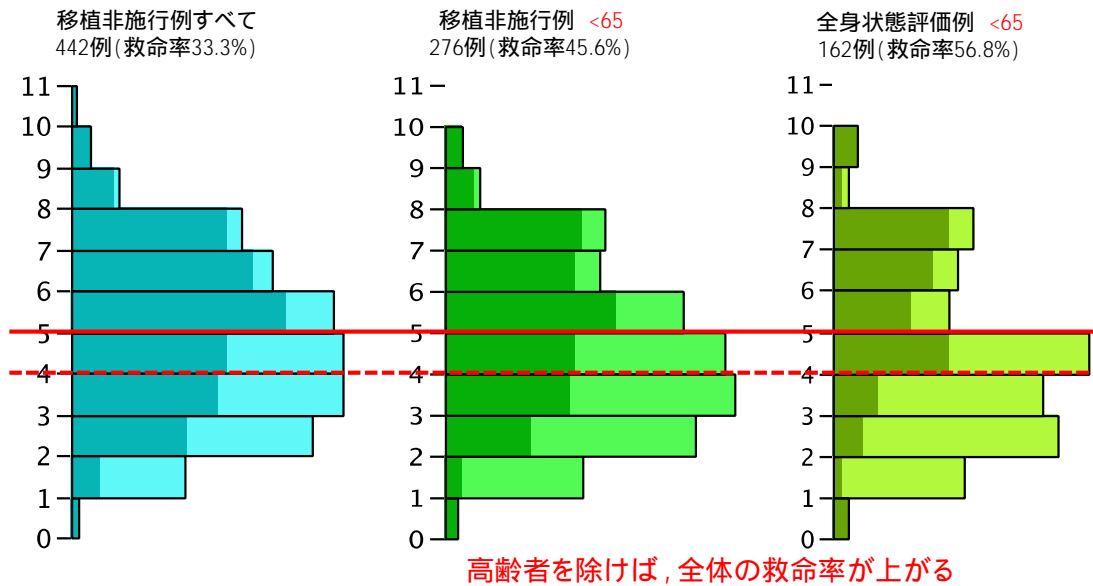


図2. 2010-2016 急性肝不全(昏睡型) 死亡率 (65歳未満のみスコアリング)

表1. スコアリング診断性能(5点cut)

*非移植例のみの死亡率の検討

年度 (n)	感度 Sensitivity	特異度 Specificity	陽性適中率 PPV	陰性適中率 NPV	正診率 accuracy
1998-2003 (421)	0.80	0.76	0.84	0.70	0.78
2004-2008 (231)	0.74	0.81	0.86	0.65	0.77
2010-2013 (268)	0.61	0.80	0.86	0.49	0.68
2014-2016 (171)	0.54	0.91	0.93	0.49	0.66

感度およびNPVの低下
4点以下での生存率の低下(特に3点以下)

表2. 4点以下の症例で致死率を高めている要因

項目	死亡	生存	P	項目	死亡	生存	P
年齢	58.6	41.5	<0.0001	O-C	6.8	13.0	N.S.
ALB	3.0	3.3	0.0020	PT%	30.5	27.9	N.S.
BUN	52.2	19.3	(0.0779)	TB	8.6	7.0	N.S.
Cre	2.25	1.37	0.0031	D/T	0.66	0.63	N.S.
PLT	12.2	13.8	N.S.	AST	3144	3086	N.S.
合併症	2.5	1.0	<0.0001	ALT	2407	2481	N.S.
基礎疾	27.9%	6.2%	<0.0001	LDH	2672	1917	N.S.
SIRS	69.4%	44.6%	0.0023	CK	1843	773	N.S.
				NH3	146	149	N.S.
				HGF	2.7	2.6	N.S.

*非移植例のみの生存率の検討 単変量解析

表3. スコアリング診断性能

*非移植例のみの死亡率の検討

年度 (n)	感度 Sensitivity	特異度 Specificity	陽性適中率 PPV	陰性適中率 NPV	正診率 accuracy
5点判定 (440)	0.57	0.84	0.88	0.49	0.66
4点判定 (440)	0.72	0.62	0.79	0.52	0.69
選別後 5点判定 (255)	0.61	0.86	0.87	0.60	0.71
選別後 4点判定 (255)	0.79	0.65	0.76	0.68	0.73

いずれの性能値, 特に正診率が改善する

表4. スコアリング診断性能 **65歳未満**

*非移植例のみの死亡率の検討

年度 (n)	感度 Sensitivity	特異度 Specificity	陽性適中率 PPV	陰性適中率 NPV	正診率 accuracy
5点判定 (276)	0.57	0.83	0.80	0.62	0.69
4点判定 (276)	0.72	0.63	0.70	0.65	0.67
選別後 5点判定 (162)	0.60	0.87	0.78	0.74	0.75
選別後 4点判定 (162)	0.81	0.67	0.66	0.83	0.73

移植推奨スコアは全身状態評価後の
5点が最良か？

表5. データマイニング手法による予後予測モデル4種と総合ハイブリッドモデルの検証結果
 昏睡発現時:昏睡型の非移植例

		決定木	RBFモデル	BP判別	BP数値	ハイブリッド型モデル
検証症例 2014-2016	正診率	70.5	72.1	59.8	68.0	73.0
	感度	69.4	73.6	55.6	63.9	70.8
	特異度	72.0	70.0	66.0	74.0	76.0
	PPV	78.1	77.9	70.2	78.0	81.0
	NPV	62.1	64.8	50.8	58.7	63.3
		決定木	RBFモデル	BP判別	BP数値	ハイブリッド型モデル
検証症例 2004-2008	正診率	71.2	69.4	71.1	75.6	77.2
	感度	75.0	71.0	76.3	75.3	82.8
	特異度	67.1	67.8	65.5	75.9	71.3
	PPV	70.6	70.2	70.3	76.9	75.5
	NPV	71.8	68.6	72.2	74.2	79.5