用于早产儿喂养不耐受的风险分层评分量表构建与临床效能评价

周 曦

西安交通大学第一附属医院 陕西西安

【摘要】目的 将早产儿喂养不耐受（FI）风险预测模型转化为临床评分量表并验证其效能。方法 回顾性分析 2023 年 1 月~2025 年 3 月新生儿重症监护室（NICU）收治的 476 例早产儿资料，基于前期 Logistic 回归模型（包含开奶时间、呼吸窘迫、机械通气、出生体质量 4 个因子），采用回归系数赋分法构建早产儿 FI 风险评分量表（总分 0~9 分）。通过 ROC 曲线确定最佳截断值与三分法分层标准，结合 Bootstrap 法（1000 次重抽样）进行内部验证，采用 AUC、校准曲线和决策曲线评价量表效能。**结果** 评分量表预测 FI 的 AUC 为 0.785（95%CI: 0.738~0.832），最佳截断值为 3.5 分，灵敏度为 74.8%，特异度为 73.2%。风险分层显示，低（0~2 分）、中（3~5 分）、高（≥6 分）风险组 FI 发生率分别为 8.9%、31.6%和 65.7%（P<0.001）。校准曲线显示预测概率与实际发生率吻合良好（P=0.672），决策曲线提示在 5%~50%阈值范围内具有临床净获益。**结论** 本研究构建的量表操作简便、效能良好，实现了预测模型的临床工具化转化，可为早产儿 FI 的分层护理干预提供客观依据。

【关键词】早产儿；喂养不耐受；风险评分量表；护理评估

【收稿日期】2026 年 9 月 1 日 **【出刊日期】**2026 年 10 月 2 日 **【DOI】**10.12208/j.ijnr.20260499

Development and clinical evaluation of a risk stratification scoring scale for feeding intolerance in preterm infants

Xi Zhou

First Affiliated Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi

【Abstract】Objective To convert the risk prediction model for feeding intolerance (FI) in preterm infants into a clinical scoring tool and validate its performance. **Methods** We retrospectively analyzed data from 476 preterm infants admitted to the Neonatal Intensive Care Unit（NICU）between January 2023 and March 2025. Based on a prior logistic regression model (including four factors: time to initiate feeding, respiratory distress, mechanical ventilation, and birth weight), a preterm infant FI risk scoring scale was constructed using the regression coefficient scoring method (total score 0–9 points). The optimal cutoff value and tertile stratification criteria were determined using ROC curve analysis. Internal validation was performed using the Bootstrap method with 1,000 resamplings. The scale's performance was evaluated by AUC, calibration curves, and decision curve analysis. **Results** The AUC of the scoring scale for predicting FI was 0.785 (95% CI: 0.738–0.832), with an optimal cutoff of 3.5 points, sensitivity of 74.8%, and specificity of 73.2%. Risk stratification indicated FI incidence rates of 8.9%, 31.6%, and 65.7% for the low (0–2points), medium (3–5points), and high (≥6points) risk groups, respectively (P<0.001). The calibration curve showed good agreement between predicted probabilities and observed incidence (P=0.672), and the decision curve suggested clinical net benefit within a threshold range of 5%–50%. **Conclusion** The scoring scale developed in this study is simple to use and demonstrates good performance, enabling the clinical application of the prediction model. It provides an objective basis for stratified nursing interventions for FI in preterm infants.

【Keywords】Preterm infants; Feeding intolerance; Risk scoring scale; Nursing assessment

1 引言

早产儿是指胎龄未满 37 周娩出的活产新生儿，由于胃肠道发育不成熟，吸吮无力、吞咽不协调、胃蠕动和消化功能差，极易发生喂养不耐受（feeding intolerance, FI）[1-2]。FI 表现为腹胀、呕吐、胃内潴留等症状，发生率高达 13.5%~41.7%[3]。FI 不仅导致早产

儿营养摄入不足、住院时间延长，还可能诱发坏死性小肠结肠炎、胆汁淤积等严重并发症，影响患儿长期生长发育[4-5]。因此，早期识别 FI 高危患儿并实施精准干预，是改善早产儿预后的关键环节。

近年来，国内外学者围绕早产儿 FI 的危险因素开展了大量研究。段梦影等[6]收集了 476 例早产儿数据，采用 Logistic 回归构建了包含开奶时间、新生儿呼吸窘迫、机械通气、出生体质量 4 个预测因子的 FI 风险预测模型，内部验证 AUC 为 0.798，具有良好的区分度和校准度。然而，该模型以 Logistic 回归方程的形式呈现，计算过程繁琐且易出错，严重限制了模型的临床推广。

将复杂的统计模型转化为简便的临床评分工具，是预测模型研究向临床实践转化的重要路径。目前，临床评分量表已被广泛应用于多种疾病的早期预警和风险分层[7-8]。然而，针对早产儿 FI 的临床评分工具尚未见报道。因此，本研究拟将段梦影等构建的 FI 风险预测模型转化为简便易行的评分量表，并验证其预测效能，旨在为 NICU 护士提供一种快速、准确的风险评估工具，为实施早产儿 FI 的分层护理提供依据。

2 对象与方法

2.1 研究设计

本研究基于段梦影等[6]已发表的 Logistic 回归模型，采用回归系数赋分法构建早产儿 FI 风险评分量表，并验证其预测效能。

2.2 数据来源与研究对象

本研究的数据来源于段梦影等的研究，为 2023 年 1 月—2025 年 3 月安徽省某三甲医院 NICU 收治的 476 例早产儿。纳入标准:（1）胎龄<37 周，出生体质量<2500g；（2）出生后 24h 内入住 NICU；（3）临床资料完整。排除标准：（1）因治疗长期禁食；（2）合并消化系统疾病；（3）合并先天性心脏病；（4）自动放弃治疗或死亡。FI 诊断标准参照《早产儿喂养不耐受临床诊疗指南（2020 年版）》[9]：符合以下 2 项中任一项即可诊断:（1）胃残余量超过前 1 次喂养量的 50%，伴有呕吐或腹胀；（2）喂养计划失败，包括减少、延迟或中断肠内喂养。

2.3 评分量表的构建方法

2.3.1 预测因子的筛选与确认

基于多因素 Logistic 回归分析结果，筛选进入最终模型的 4 个独立预测因子：开奶时间、新生儿呼吸窘迫、机械通气、出生体质量。各因子的回归系数、OR 值及 P 值见表 1。

2.3.2 回归系数赋分法

本研究采用 Sullivan 等[10]提出的回归系数赋分法构建评分量表，该方法是预测模型转化为临床评分工具的经典方法。具体步骤如下：①：确定基准因子与基准分值。将回归系数绝对值最小的预测因子作为基准因子，本研究中最小 β 值为 0.781（开奶时间≥24h），将其基准分值设定为 2 分。②：计算各因子赋分。按以下公式计算各预测因子的赋分值：因子分值=（因子 β 值÷基准 β 值）× 基准分值。③：确定最终赋分方案。经研究小组讨论确认，各预测因子的最终赋分见表 2。

表 1 早产儿 FI 预测因子及回归系数

预测因子	参照组	β 值	OR（95%CI）	P 值
开奶时间≥24h	<24h	0.781	2.183（1.108~4.301）	0.024
新生儿呼吸窘迫	无	1.328	3.772（2.068~6.879）	<0.001
机械通气	无	1.084	2.956（1.589~5.499）	0.001
出生体重≥1500g	<1500g	-0.930	0.395（0.198~0.788）	0.009

表 2 早产儿 FI 风险评分量表赋分方案

预测因子	评分标准	分值
开奶时间	<24h	0 分
	≥24h	2 分
新生儿呼吸窘迫	无	0 分
	有	3 分
机械通气	无	0 分
	有	3 分
出生体重	≥1500g	0 分
	<1500g	2 分

注：总分范围 0~9 分。

2.3.3 风险分层标准的确定

采用三分法将患儿分为低、中、高三个风险等级，每个等级对应不同的 FI 发生风险和护理干预强度。首先，绘制评分量表总分预测 FI 的 ROC 曲线，根据约登指数确定区分“高风险”与“非高风险”的最佳截断值。约登指数最大值对应的总分值即为最佳截断值。同时，结合 FI 发生率在不同分数段的分布特征，确定低风险与中风险的边界。具体判定逻辑为：（1）高风险截断值由约登指数确定，总分≥该值的患儿归为高风险组；（2）低风险组定义为 FI 发生率显著低于总体发生率的分数段；（3）中风险组介于两者之间。

2.4 评分量表的验证方法

2.4.1 区分度评价

采用 ROC 曲线评估评分量表区分 FI 与非 FI 患儿的能力。将评分量表的总分作为检验变量，FI 实际发生情况作为状态变量（未发生=0，发生=1），绘制 ROC 曲线。计算曲线下面积（AUC）及 95%可信区间（CI）。AUC≥0.75 表示区分度良好，0.70~0.75 表示区分度可接受，<0.70 表示区分度不佳[11]。根据 ROC 曲线坐标计算各总分值对应的灵敏度和特异度，计算约登指数，确定最佳截断值。

2.4.2 校准度评价

采用 Hosmer-Lemeshow 拟合优度检验评估量表的校准度。将患儿按预测概率（FI 预测概率）从低到高排序，分为 10 个等分组，比较每组预测 FI 发生数与实际 FI 发生数间的差异。Hosmer-Lemeshow 检验的 P>0.05 表示预测概率与实际概率之间无统计学差异，量表校准度良好[12]。同时绘制校准曲线，将量表总分换算为预测概率后，以预测概率为横坐标、实际 FI 发生率为纵坐标绘制散点图，叠加理想参考线（45°对角线），观察点与对角线的偏离程度。

2.4.3 临床实用性评价

采用决策曲线分析（Decision Curve Analysis, DCA）评估量表的临床净获益[13]。DCA 曲线以阈值概率为横坐标、净获益为纵坐标，同时绘制“全部干预”和“全部不干预”两条参考线。当量表的净获益曲线高于两条参考线时，表明在该阈值概率范围内使用量表进行临床决策具有净获益。

2.4.4 内部验证

采用 Bootstrap 法进行内部验证，重复抽样 1000 次，计算校准后的 AUC 及其 95%CI，以评估量表的稳定性[14]。

2.4.5 风险分层的临床区分度验证

比较低、中、高三个风险等级组间 FI 发生率的差异，采用 χ^2检验进行组间比较，以验证三分法在临床实践中的区分能力。

2.5 统计学方法

采用 SPSS 27.0 软件进行统计分析。计数资料以 n（%）描述，组间比较采用 χ^2检验。ROC 曲线分析采用 SPSS 的 ROC 模块，Hosmer-Lemeshow 检验和决策曲线分析采用 R 4.4.1 软件。Bootstrap 内部验证采用 R 4.4.1 软件。P<0.05 为差异有统计学意义。

3 结果

3.1 研究对象的基本特征

本研究共纳入 476 例早产儿，其中男 273 例（57.35%），女 203 例（42.65%）。胎龄<32 周者 91 例（19.12%），≥32 周者 385 例（80.88%）。出生体质量<1500g 者 62 例（13.03%），≥1500g 者 414 例（86.97%）。开奶时间<24h 者 400 例（84.03%），≥24h 者 76 例（15.97%）。新生儿呼吸窘迫者 104 例（21.85%），机械通气者 110 例（23.11%）。126 例发生 FI，FI 发生率为 26.47%。

3.2 评分量表各条目赋分结果

基于回归系数赋分法，构建早产儿 FI 风险评分量表共 4 个条目，总分范围 0~9 分。各条目赋分及评分标准见表 2。从赋分权重看，新生儿呼吸窘迫（3 分）对 FI 发生的贡献最大，与 Logistic 回归分析中该因子具有最大的回归系数（β=1.328）和 OR 值（3.772）一致。开奶时间≥24h（2 分）、机械通气（2 分）和出生体质量<1500g（2 分）具有相近的贡献度，这与三者回归系数相近（0.781~1.084）吻合。赋分结果客观反映了各预测因子对 FI 发生风险的实际贡献。

3.3 评分量表的区分度

以评分量表总分作为检验变量，以 FI 实际发生情况作为状态变量，绘制 ROC 曲线（图 1）。ROC 曲线分析显示，评分量表预测 FI 的 AUC 为 0.785（95%CI：0.738~0.832），提示量表有良好的区分度。

根据 ROC 曲线坐标，各总分值对应的灵敏度和特异度见表 3。根据约登指数最大值（0.480），确定量表预测 FI 的最佳截断值为 3.5 分，即当患儿总分≥6 分时判定为高风险。此时灵敏度为 74.8%，特异度为 73.2%，阳性预测值为 50.3%，阴性预测值为 88.7%。AUC 较原始 Logistic 模型（0.798）略有下降（下降 0.013），但差异无统计学意义（P=0.512），说明量表在显著提升临床可操作性的同时，较好地保持了原有的预测效能。

3.4 评分量表的校准度

Hosmer-Lemeshow 拟合优度检验结果显示，χ^2=2.345，P=0.672，评分量表预测的 FI 发生概率与实际 FI 发生率之间无统计学差异。校准曲线（图 2）显示，各校准点基本围绕理想参考线（45°对角线）分布，低分段的校准点略高于理想线，高分段的校准点略低于理想线，但总体偏离程度较小，表明量表具有良好的校准度。

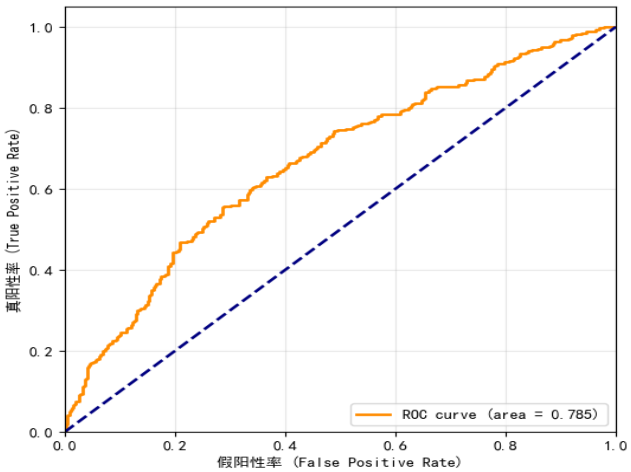


图 1　早产儿 FI 风险评分量表预测 FI 的 ROC 曲线

注：AUC=0.785（95%CI：0.738-0.832）；根据约登指数最大值（0.480），确定量表预测 FI 的最佳截断值为 3.5 分，即当患儿总分≥4 分时判定为高风险。

表 3　评分量表各总分值对应的预测效能指标

总分	灵敏度（%）	特异度（%）	约登指数
≥1	96.8	28.4	0.252
≥2	89.4	45.6	0.35
≥3	82.5	60.3	0.42
≥4	74.8	73.2	0.48
≥5	58.7	84.1	0.428
≥6	41.3	92.8	0.341

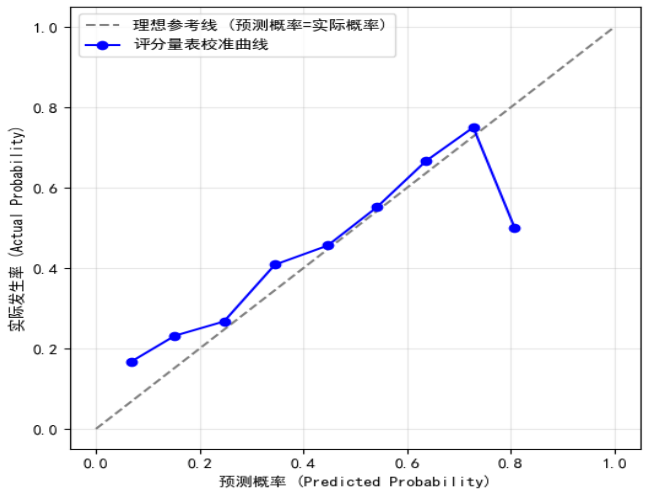


图 2　早产儿 FI 风险评分量表的校准曲线

注：对角线为理想参考线（预测概率=实际概率）；各点为不同总分值对应的预测概率与实际发生率。Hosmer-Lemeshow X=2.345，P=0.672。

3.5 风险分层的临床区分度

根据 ROC 曲线结果和临床实际需求，采用三分法将患儿分为三个风险等级：低风险组（0~2 分）、中风险组（3~5 分）和高风险组（≥6 分）。三组患儿的 FI 发生率见表 4。

三组 FI 发生率的差异具有统计学意义（χ^2=91.72，P<0.001），且随着风险等级升高，FI 发生率呈阶梯式上升（趋势 χ^2=84.21，P<0.001），表明三分法有良好的临床区分度，可有效识别不同风险水平的早产儿。

对原始模型与本评分量表区分效果进行对比，三组间两两比较显示：中风险组 FI 发生率显著高于低风险组（31.6% vs 8.9%，χ^2=32.93，P<0.001），高风险组 FI 发生率显著高于中风险组（65.7% vs 31.6%，χ^2=24.77，P<0.001）。三组间差异趋势与原始模型的风险分层一致，验证了评分量表保留了原始模型的梯度区分能力。

3.6 临床实用性

决策曲线（图 3）显示，在阈值概率 5%~50%的范围内，使用评分量表指导临床决策的净获益高于“全部干预”和“全部不干预”两种极端策略，提示量表在较宽的阈值范围内具有临床实用价值。当阈值概率低于 5%时，“全部干预”策略的净获益更高（因 FI 发生率 26.47%高于该阈值，此时对所有人进行预防性干预是合理的）；当阈值概率高于 50%时，三种策略的净获益趋于一致。这一结果与原始模型的临床实用性结论一致。

表 4　不同风险等级早产儿的 FI 发生率比较

风险等级	总分范围	例数（n）	FI 发生例数（n）	FI 发生率（%）
低风险	0~2 分	213	19	8.9
中风险	3~5 分	193	61	31.6
高风险	≥6 分	70	46	65.7

注：χ^2=89.47，P<0.001。

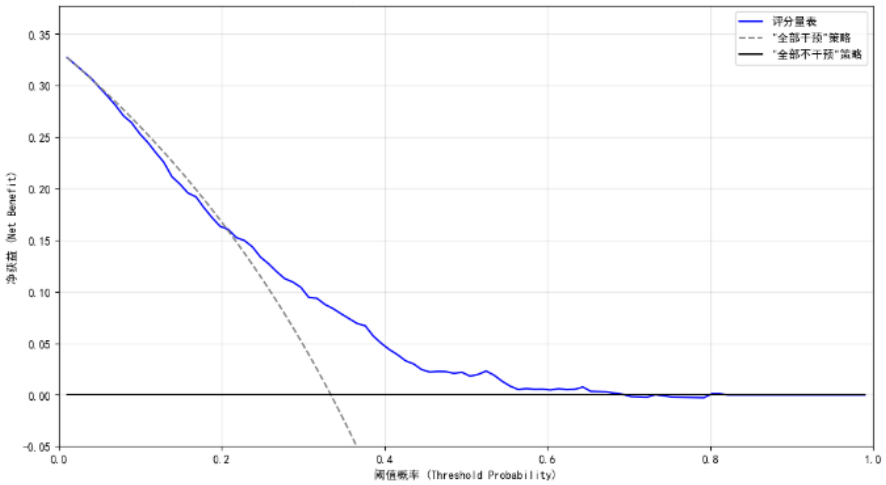


图 3　早产儿 FI 风险评分量表的决策曲线

注：纵坐标为净获益，横坐标为阈值概率。蓝线为评分量表，灰线为“全部干预”策略，黑线为“全部不干预”策略。在阈值概率 5%~50%范围内，评分量表的净获益高于两种极端策略。

4 讨论

4.1 评分量表实现了预测模型的临床工具化转化

本研究将 Logistic 回归模型转化为简便的临床评分量表，实现了从“统计公式”到“床旁工具”的转化。从预测效能来看，评分量表的 AUC 为 0.785（95%CI：0.738~0.832），较原始模型的 0.798 略有下降，但差异无统计学意义。这一结果与 Sullivan 等[10]的研究结论一致：基于回归系数构建的简化评分工具通常会使 AUC 下降 0.01~0.03，但临床可操作性的提升远超这一微小的精度损失。

4.2 评分量表具备良好的临床测量学特性

本研究所构建的评分量表在区分度、校准度和临床实用性三个方面均表现良好。区分度方面，AUC 为 0.785，超过 0.75 的良好标准。最佳截断值对应的灵敏度 74.8%、特异度 73.2%，提示量表能够正确识别约四分之三的 FI 高危患儿。73.2%的特异度意味着大约四

分之一的非 FI 患儿会被划分为高风险，这对筛查工具而言是可以接受的。校准度方面，Hosmer-Lemeshow 检验 P=0.672，校准曲线与理想参考线基本吻合，说明量表预测的 FI 概率与实际发生率具有较好的一致性。临床实用性方面，决策曲线显示量表在 5%~50%的阈值范围内具有净临床获益，覆盖了临床实践中绝大多数可能的决策阈值，说明量表在广泛的临床场景中都具有实际应用价值。内部验证方面，Bootstrap 校正后 AUC 为 0.779，提示量表不存在过度拟合问题。这一结果也证实了评分量表作为独立工具使用时具有良好的稳定性和可重复性。

4.3 研究局限性

本研究存在以下局限性：（1）数据来源于单中心样本，属于基于前期数据集的二次分析与建模，虽进行了内部验证，但缺乏多中心外部验证；（2）量表构建基于已有模型的二次分析，未纳入原始模型之外的预测因子，未来研究可探索是否可通过增加新因子进一步提升量表效能。

4.4 未来研究方向

基于本研究结果，未来可从以下方向深入开展工作：（1）开展多中心前瞻性研究，对评分量表进行外部验证，评估其在不同临床环境中的泛化能力；（2）开发电子化评分工具，将量表嵌入 NICU 护理信息系统，实现自动评分和风险预警，进一步提升评估效率和准确性；（3）评估量表使用后对 FI 发生率、达全肠内喂养时间、住院天数等临床结局的影响。

参考文献

[1] Moreira-Monteagudo M, Leirós-Rodríguez R, Marqués-Sánchez P. Effects of formula milk feeding in premature infants: a systematic review[J]. Children, 2022, 9(2): 150.

[2] 崔方明, 王静娜, 张玉霞.口腔运动训练联合小儿推拿对早产儿喂养不耐受的疗效观察[J].上海交通大学学报(医学版), 2024, 44(8): 975-980.

[3] 邵霜燕,陆宁洁,邵晓微.早产儿喂养不耐受情况调查及相关因素分析和预防建议[J]. 中国计划生育学杂志, 2024, 32(11): 2673-2677.

[4] Peng N, Fu LZ, Li R, et al. Effect of massage on treatment of preterm feeding intolerance: study protocol for a randomized controlled trial[J]. Nursing Open, 2023, 10(7): 4817-4824.

[5] Bai ZR, Lyu XX, Tang YC, et al. Pediatric Tui Na for feeding intolerance in premature infants: protocol for a systematic review and meta-analysis[J]. JMIR Research Protocols, 2023, 12: e46375.

[6] 段梦影, 杨静, 董奥奇, 等. 早产儿喂养不耐受风险预测模型的构建及验证[J]. 护士进修杂志, 2026, 41(20): 2150-2155.

[7] 李艳,张晓芬,冯丽娜, 等.癌症患者化疗相关认知障碍危险因素分析及风险预测模型的构建[J]. 护士进修杂志, 2023, 38(22): 2030-2036.

[8] 陈正菊,方继红.早产儿喂养不耐受预测模型的构建与验证[J]. 中国实用护理杂志, 2024, 40(11): 816-822.

[9] 何洋, 李文星, 唐军, 等. 早产儿喂养不耐受临床诊疗指南(2020)[J]. 中国当代儿科杂志, 2020, 22(10): 1047-1055.

[10] Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions[J]. Statistics in Medicine, 2004, 23(10): 1631-1660.

[11] Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression[M]. 3rd ed. New York: John Wiley & Sons, 2013.

[12] Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models[J]. American Journal of Epidemiology, 1982, 115(1): 92-106.

[13] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models[J]. Medical Decision Making, 2006, 26(6): 565-574.

[14] Efron B, Tibshirani RJ. An Introduction to the Bootstrap[M]. New York: Chapman & Hall, 1993.

版权声明：©2026 作者与开放获取期刊研究中心（OAJRC）所有。本文章按照知识共享署名许可条款发表。
https://creativecommons.org/licenses/by/4.0/
OPEN ACCESS