基于生存优化模型的 NIPT 时点选择研究

张辉，肖扬，卜子轩
陆军航空兵学院，北京 100000

DOI: 10.61369/SSSD.2025180011

摘　　要： 高 BMI 孕妇 NIPT 检测时点选择问题亟待解决。本研究构建 AFT 模型与风险优化体系，基于 1689 例临床数据量化 BMI 等指标与检测时点的关联，通过网格搜索法确定 7 个 BMI 分组的最佳检测周数。结果显示，分组策略使总体检测风险降低 45.8%，BMI ≥ 40 组降幅达 54.2%，显著提升高 BMI 人群筛查时效性。该成果为个性化检测时点选择提供依据，助力产前筛查精准化与资源优化。

关 键 词： NIPT；生存分析；加速失效时间模型；风险优化；BMI 分组；时点优化；网格搜索

Research on NIPT Time Point Selection Based on Survival Optimization Model

Zhang Hui, Xiao Yang, Bu Zixuan
Army Aviation Academy, Beijing 100000

Abstract: The problem of selecting the timing of NIPT testing for high BMI pregnant women urgently needs to be solved. This study constructs an AFT model and risk optimization system, quantifies the correlation between BMI and other indicators and detection time points based on 1689 clinical data, and determines the optimal detection weeks for 7 BMI groups through grid search method. The results showed that the grouping strategy reduced the overall testing risk by 45.8%, with a decrease of 54.2% in the BMI ≥ 40 group, significantly improving the timeliness of screening for high BMI populations. This achievement provides a basis for personalized testing timing selection, helping to achieve precision in prenatal screening and optimize resources.

Keywords: **NIPT; survival analysis; accelerated failure time model; risk optimization; bmi grouping; time optimization; grid search**

非侵入性产前检测（NIPT）是胎儿染色体异常筛查的核心技术，但 BMI ≥ 30 的高危孕妇假阴性率高达 15%-20%[1]。研究表明，BMI 与胎儿游离 DNA（cffDNA）浓度呈负相关（$r=-0.261, P<0.05$）[2]，导致检测窗口期需延长。然而，现行指南对肥胖孕妇存在矛盾：美国医学遗传学会（ACMG）建议慎用 NIPT[3]，中国技术规范将其列为慎用人群[4]，但均未提供替代方案。

现有研究存在三大局限：①未整合年龄、体重等多维指标[5]；②忽视检测时效性与假阴性风险的动态平衡[6]；③缺乏针对中国人群的优化模型[7]。本研究基于加速失效时间（AFT）模型与网格搜索算法，构建了 BMI 分组的动态检测时点预测模型。通过 1689 例临床数据分析，为高 BMI 孕妇的 NIPT 个性化时点选择提供依据，对提升筛查精准性与医疗资源优化配置具有重要临床价值。

一、模型方法选择

研究数据存在显著左删失（12 周前）和右删失（30 周后）现象，导致传统线性回归模型因无法处理截尾数据产生偏倚。Cox 比例风险模型因协变量（BMI、年龄等）对 Y 染色体浓度的影响随孕周动态变化，违反比例风险假设（PH），亦不适用。为此，本研究采用加速失效时间（AFT）模型：通过参数化对数生存时间，直接量化多因素对检测时点的差异化影响，无需 PH 假设即可解析时变作用，最终实现高 BMI 孕妇的精准检测时点选择[8]。

（一）AFT 模型介绍

AFT 模型是生存分析的重要工具，专用于处理含右删失的时间至事件数据。其核心是通过参数化模型直接关联协变量（如 BMI、年龄）与事件发生时间，假设对数生存时间服从广义极值分布或 Weibull 分布，并通过引入协变量线性组合刻画变量间关系。相较于 Cox 比例风险模型，AFT 无需满足比例风险假设（PH），可灵活解析多维风险因素的时变作用。

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

1. 网格搜索法

为最小化风险函数，本文采用网格搜索法优化各风险项的权重组合。网格搜索法是一种通过系统性遍历预设参数组合来优化模型性能的超参数调优技术。其核心思想是为每个待优化参数定义候选值范围，生成所有可能的参数组合形成“搜索网格”，并通过交叉验证对每组参数进行性能评估，最终选择使模型达到最优指标（如最小验证误差）的参数组合作为最终配置。该方法通过穷举搜索确保覆盖全局最优解空间，尤其适用于多维参数空间中存在非线性关系的复杂模型优化场景[9]。

二、生存优化模型建立

（一）AFT(加速失效)模型

为了建立对于不同特征（身高，年龄，体重，BMI 等）的样本的时间与Y染色体浓度的关系，我们采用了加速失效时间(AFT)模型来对数据进行拟合，得到多变量条件下的限制关系，通过回归系数量化各变量对时间的具体贡献，并且可以处理删失数据。

$$\log(T) = \beta_0 + \beta_1 \cdot BMI + \beta_2 \cdot WEIGHT + \beta_3 \cdot HIGH + \beta_4 \cdot AGE + \epsilon \quad (2\text{-}1)$$

其中 T 是胎儿Y染色体浓度首次达到 Y_0 的达标时间，$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ 是模型的回归系数，表示各自变量对达标时间的影响，ò 是误差项，通常假设为正态分布[10]。

（二）时点置信度

计算时点 T 对应的达标置信度：

$$\alpha(T) = \frac{N_{Y<0.04}}{N_{Y>0.04}} \quad (2\text{-}2)$$

其中 $N_{Y<0.04}$ 与 $N_{Y>0.04}$ 分别是Y染色体达标与未达标的样本数。

（三）风险优化函数

风险函数 $R_1(T, BMI)$ 用于评估在检测时间 T 下，因未能及时发现异常，Y染色体浓度不足检测失败或者因随机误差导致失败而带来的风险。它主要由以下部分构成：

未达标风险（检测过早，$Y < 0.04$，检测无效）

$$R_1(T, BMI) = P(Y(T, BMI) < 0.04) \quad (2\text{-}3)$$

晚检风险（检测过晚，异常胎儿发现延迟，缩短干预窗口）

$$R_2(t) = \alpha \cdot I(T > T_{late}) \quad (2\text{-}4)$$

检测误差风险（由于测序/检测误差，出现假阴性/假阳性）：

假阴性风险（未发现异常）：

$$R_3(t) = \beta \cdot P(falset) \quad (2\text{-}5)$$

假阳性风险（误判异常）：

$$R_4(t) = \gamma \cdot P(falset) \quad (2\text{-}6)$$

其中 T_{late} 是窗口期（比如12周以前晚检风险低，$13-27$ 周晚检风险中，27周后晚检风险高）的分界点，β 与 γ 是权重。

将上述部分加权组合有：

$$R(t, BMI) = w_1 \cdot R_1(t, BMI) + w_2 \cdot R_2(t)3 + w_3 \cdot R_3(t) + w_4 \cdot R_4(t) \quad (2\text{-}7)$$

其中 $w1, w2, w3, w4$ 是风险项的权重，用于平衡每个风险在总风险函数中的重要性。

三、实验结果及分析

本文对2025年数学建模国赛C题所提供的1083例男胎数据与606例女胎数据进行拟合与分析。

表2 AFT 模型参数

变量	系数（β）	标准差	p 值
BMI	−0.031	0.005	<0.05
年龄	0.012	0.003	<0.05
体重	−0.005	0.001	<0.05
身高	0.008	0.002	<0.05
截距	2.456	0.12	<0.05

根据模型计算，最终得到的最佳参数表如上表。由此得到模型基本预测函数：

$$log(T) = 2.456 - 0.031 * BMI + 0.012 * AGE - 0.005 * WEIGHT + 0.008 * HIGH$$

模型拟合结果显示，BMI 对 Y 染色体达标时间具有显著负向影响（β=−0.031，p<0.05），即 BMI 每增加1个单位，达标时间延长约3.1%。年龄（β=0.012，p<0.05）和身高（β=0.008，p<0.05）呈正向关联，体重（β=−0.005，p<0.05）呈负向关联，与已有研究结论一致。通过 Weibull 分布假设下的 AFT 模型验证，模型拟合优度良好（AIC=2453.2），残差分析支持参数合理性。测试集验证表明，模型在95%置信区间内可靠预测 Y 染色体浓度达标时间，或者以95%置信度确定最佳 NIPT 检测时点。

通过风险函数最小化方法，我们确定了7个 BMI 分组的最佳检测时间：

表3 BMI 分组的最佳检测时间

BMI 分组	达标时间中位数（周）	最佳检测时间（周）	归一化风险系数
<23.0	14.3	13.2	0.0222
23.0−24.9	15.9	15.3	0.0489
25.0−29.9	17.3	16.6	0.0567
30.0−34.9	19.1	18.2	0.0924
35.0−37.4	20.1	19.8	0.1361
37.5−39.9	21.0	20.6	0.1678
≥40.0	22.1	21.9	0.2136

分析显示，采用分组特异性检测时间后，总体检测风险平均降低45.8%。特别是在超重和肥胖组中，风险降低效果最为显著，

 Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

这与高 BMI 孕妇传统检测时间选择不当导致的高假阴性率问题相符。

四、结论

本研究构建了基于 AFT 模型与风险优化的 NIPT 检测时间分组策略，通过 BMI 细分（7组）确定个性化检测周数，寻找最低风险的检测时点。结果显示，该策略使检测风险降低45.8%，尤其显著改善 BMI ≥ 40组（风险降幅54.2%），为高 BMI 孕妇提供精准筛查方案。

参考文献

[1] CESARELLI M，ROMANO M，BIFULCO P. Comparison of short term variability indexes in cardiotocographic foetal monitoring[J]. Computers in Biology & Medicine，2009，39(2)：106.

[2] 王杰，等 . 无创产前筛查中胎儿游离 DNA 比例的影响因素分析 [J]. 中华医学遗传学杂志，2018，35(3)：390 - 394.

[3] KLEIN J P，MOESCHBERGER M L. Survival Analysis：Techniques for Censored and Truncated Data[M]. 2nd ed. New York：Springer，2003.

[4] AMERICAN COLLEGE OF MEDICAL GENETICS AND GENOMICS. ACMG statement on noninvasive prenatal screening for fetal aneuploidy[J]. Genetics in Medicine，2016，18(10)：1056 - 1059.

[5] 国家卫生计生委办公厅 . 胎儿染色体非整倍体无创基因检测技术规范 (试行)[Z]. 2016.

[6] THERNEAU T M，GRAMBSCH P M. Modeling Survival Data：Extending the Cox Model[M]. New York：Springer，2000.

[7] HARRELL F E. Regression Modeling Strategies：With Applications to Linear Models，Logistic Regression，and Survival Analysis[M]. New York：Springer，2001.

[8] 刘萧，李静，许珂 . 基于生存分析的智能电网安全告警事件持续时间预测模型 [J]. 计算机应用与软件，2024，41(1)：328-335.DOI：10.3969/j.issn.1000-386x.2024.01.048.

[9] 王能发，杨哲，刘自鑫 . 基于生存理论的两类博弈模型 [J]. 重庆师范大学学报（自然科学版），2024，41(1)：1-7.

[10] 张娜，陈文倩，白雪松，等 . 基于时空优化模型的 PM_(2.5) 遥感估测研究 [J]. 中国环境科学，2024(4).

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.