研究队列中共10,291名新生儿，其中仅273名新生儿最终确诊宫内感染，感染发生率约为2.56%，与已有的文献报道相似（或者比国内其他报道的低，表明国妇婴的医疗水平高）。Table1中展示了感染与未感染新生儿及其母亲在基线时的临床特征差异。基线期的单因素分析表明产妇和新生儿的多个临床特征均与新生儿宫内感染有相关性。

为了排除变量间共线性和混杂因素对结局的影响，我们挑选单因素P＜0.05的临床特征进行LASSO回归筛选更具有临床解释性的预测因子。LASSO回归结果显示（Table 2）与新生儿感染独立相关的临床特征有8个，分别为产程开始至胎膜早破的时间差，母亲宫腔培养，孕周，白细胞计数，中性粒细胞比例，关联CRP，B族链球菌和羊水情况。其中产程开始至胎膜早破时间差、孕周近足月、孕周远足月、白细胞计数、关联CRP异常、B族链球菌异常和未检测、羊水有胎粪是新生儿感染的危险因素（OR＞1）。母亲宫腔培养未检查、中性粒细胞比例、关联CRP未检查是新生儿感染的保护因素（OR＜1）。值得注意的是，这8个与结局独立相关的重要变量中，关联CRP的缺失较严重，有9,290名（90.27%）产妇未测量关联CRP，主要原因是关联CRP的测量不是常规的临床检验项目（临床可以更多解释）。

用筛选出来的8个预测因子，我们分别通过logistic回归方法（用普通的logistic回归就可以，不需要用加权的）和随机森林算法建立模型，并比较了两个模型的预测能力（Figure 3，Figure4）。Figure 3是两个模型的ROC曲线比较，结果显示RF（随机森林）模型的AUC（曲线下面积）比LR（logistic回归）模型的高，且具有统计学意义（*P* = 0.02）。Figure 4是两个模型的阳性PR曲线比较，结果显示RF（随机森林）模型的AUC（曲线下面积）比LR（logistic回归）模型的高，且具有统计学意义（*P* = 0.002）。以上结果可以看出，构建临床预测模型时，非线性的RF算法在预测能力和防过拟合能力方面都优于传统的logistic回归方法。在RF模型（Model 1）的PR曲线中，当cutoff为53%时，模型预测感染的召回率为60.97%，与基于经验的临床实践（60%，175/273）相当；精确度达到26.04%，比基于经验的临床实践（21%，175/836）略高。

由于关联CRP的缺失较严重，因此我们在数据较完整的队列（测量CRP患者群体）中进行亚组分析，以确定关联CRP对结局的真正影响。该亚组共包含1001名患者，新生儿确诊感染的人数有157名。该亚组中的感染发生率约15.68%，高于整体队列的感染率，因为通常产科医生仅对高危的胎膜早破产妇测量CRP(需要医生临床解释)。对该亚组进行倾向性评分匹配，以消除混杂因素对CRP的影响。匹配后的队列共包含688名患者。单因素分析（Table 3）结果显示，匹配后的队列基线基本对齐，而CRP的两组患者感染发生率仍然具有统计学上的显著性差异（P ＜ 0.001）。考虑到今后临床会将CRP作为必测指标，因此将模型在CRP组上的预测效果进行评估，ROC曲线下面积和整体队列人群的相当，分别为0.849和0.859（Figure.5，right）。但是在阳性PR曲线中，CRP患者的曲线完全包含整体队列的曲线，分别为0.504和0.227（Figure.5，left）。当cutoff为55%时，对阳性的预测召回率为75%，与基于经验的临床实践（78%）相当；阳性的预测精确度为49.21%，是基于经验的临床实践（22%）的2倍。结果表明如果患者CRP数值可以获取的情况下，预测模型可以在保证召回率与基于经验的临床实践相当的同时，大幅度降低对新生儿感染的误判率，从而降低对新生儿抗生素的误打率，这具有很重要的临床意义。

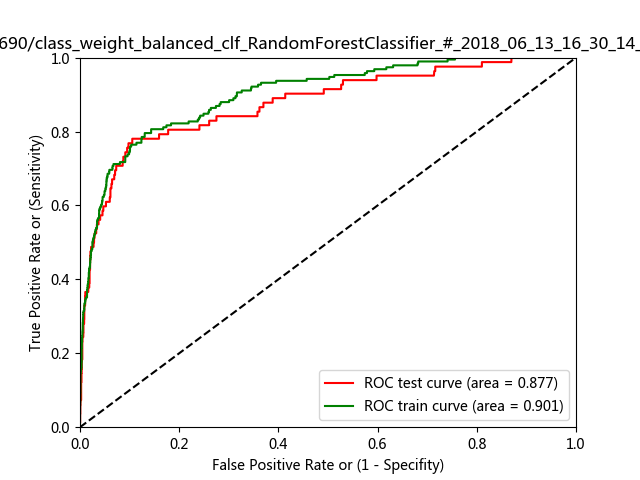
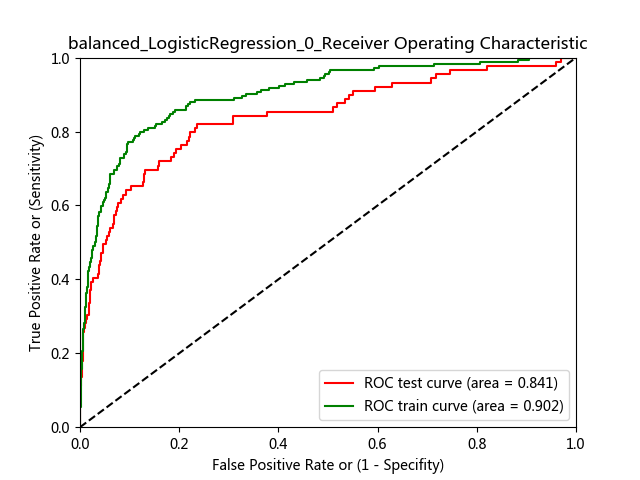
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 1: Demographic and Clinical Characteristics of 10, 291 Maternal and Neonatal** | | | | | |
|  | **Variables** |  | **Neonatal without infection**  **（N =10,018; 97.35%）** | **Neonatal with infection**  **（N =273; 2.65%）** | ***P* value** |
|  | **产妇临床特征** |  |  |  |  |
|  | 体温 |  | 36.94 (0.30) | 36.94 (0.30) | 0.029 |
|  | 产程开始至胎膜早破时间差/h |  | 14.01 (14.03) | 14.01 (14.03) | < 0.001 |
|  | 孕前体重 |  | 55.25 (7.72) | 55.25 (7.72) | 0.110 |
|  | 身高 |  | 162.54 (4.91) | 162.54 (4.91) | 0.210 |
|  | BMI |  | 20.89 (2.58) | 20.89 (2.58) | 0.010 |
|  | 年龄 |  | 29.60 (3.35) | 29.60 (3.34) | 0.053 |
|  | 白细胞计数 |  | 8.80 (2.47) | 8.80 (2.47) | < 0.001 |
|  | 中性粒细胞比例 |  | 72.40 (7.09) | 72.40 (7.09) | < 0.001 |
|  | 血小板计数 |  | 193.87 (51.04) | 193.87 (51.04) | 0.278 |
|  | 胎膜破裂至分娩时间差/h |  | 20.46 (19.75) | 20.46 (19.75) | < 0.001 |
|  | FT3 |  | 4.24 (0.53) | 4.24 (0.53) | 0.007 |
|  | FT4 |  | 11.75 (1.47) | 11.75 (1.47) | 0.173 |
|  | TT3 |  | 2.03 (0.36) | 2.03 (0.36) | 0.008 |
|  | TT4 |  | 110.08 (20.23) | 110.08 (20.23) | 0.085 |
|  | TSH |  | 1.60 (0.84) | 1.60 (0.84) | 0.190 |
|  | ogtt |  | 4.42 (0.37) | 4.42 (0.37) | 0.972 |
|  | 生产次数 |  |  |  | 0.260 |
|  |  | 1 | 8998 (89.82%) | 256 (93.77%) |  |
|  |  | 2 | 982 (9.8%) | 16 (5.86%) |  |
|  |  | 3 | 24 (0.24%) | 1 (0.37%) |  |
|  |  | 4 | 1 (0.01%) | 0 (0%) |  |
|  |  | 未知 | 13 (0.13%) | 0 (0%) |  |
|  | 白带培养 |  |  |  | < 0.001 |
|  |  | 无致病菌生长 | 35 (0.35%) | 19 (6.96%) |  |
|  |  | 有致病菌生长 | 57 (0.57%) | 8 (2.93%) |  |
|  |  | 没有检查 | 9,926 (99.08%) | 246 (90.11%) |  |
|  | 宫腔培养 |  |  |  | < 0.001 |
|  |  | 无致病菌生长 | 1,813 (18.1%) | 122 (44.69%) |  |
|  |  | 有致病菌生长 | 550 (5.49%) | 53 (19.41%) |  |
|  |  | 没有检查 | 7,655 (76.41%) | 98 (35.9%) |  |
|  | 高血压家族史 |  |  |  | 0.103 |
|  |  | No | 8,706 (86.9%) | 228 (83.52%) |  |
|  |  | Yes | 1,312 (13.1%) | 45 (16.48%) |  |
|  | 糖尿病家族史 |  |  |  | 0.109 |
|  |  | No | 9,496 (94.79%) | 252 (92.31%) |  |
|  |  | Yes | 522 (5.21%) | 21 (7.69%) |  |
|  | 心脏病家族史 |  |  |  | 0.940 |
|  |  | No | 9,903 (98.85%) | 270 (98.9%) |  |
|  |  | Yes | 115 (1.15%) | 3 (1.1%) |  |
|  | 妊娠期高血压 |  |  |  | 0.002 |
|  |  | No | 9,857 (98.39%) | 261 (95.6%) |  |
|  |  | Yes | 161 (1.61%) | 12 (4.4%) |  |
|  | 妊娠期糖尿病 |  |  |  | 0.187 |
|  |  | No | 9,995 (99.77%) | 271 (99.27%) |  |
|  |  | Yes | 23 (0.23%) | 2 (0.73%) |  |
|  | 贫血 |  |  |  | 0.994 |
|  |  | No | 9,429 (94.12%) | 257 (94.14%) |  |
|  |  | Yes | 589 (5.88%) | 16 (5.86%) |  |
|  | 甲状腺疾病史 |  |  |  | 0.997 |
|  |  | No | 9,327 (93.1%) | 254 (93.04%) |  |
|  |  | Yes | 691 (6.9%) | 19 (6.96%) |  |
|  | 乙肝病史 |  |  |  | 0.231 |
|  |  | No | 9,710 (96.93%) | 268 (98.17%) |  |
|  |  | Yes | 308 (3.07%) | 5 (1.83%) |  |
|  | 子宫肌瘤病史 |  |  |  | 0.962 |
|  |  | No | 9,668 (96.51%) | 263 (96.34%) |  |
|  |  | Yes | 350 (3.49%) | 10 (3.66%) |  |
|  | 分娩方式 |  |  |  | < 0.001 |
|  |  | 自产 | 8,793 (87.77%) | 218 (79.85%) |  |
|  |  | 产钳 | 972 (9.7%) | 52 (19.05%) |  |
|  |  | 剖腹产 | 238 (2.38%) | 3 (1.1%) |  |
|  |  | 未知 | 15 (0.15%) | 0 (0%) |  |
|  | 是否难产 |  |  |  | < 0.001 |
|  |  | No | 8,762 (87.46%) | 213 (78.02%) |  |
|  |  | Yes | 1256 (12.54%) | 60 (21.98%) |  |
|  | 孕周 |  |  |  | < 0.001 |
|  |  | 足月 | 9,387 (93.7%) | 175 (64.1%) |  |
|  |  | 近足月早产 | 572 (5.71%) | 57 (20.88%) |  |
|  |  | 远足月早产 | 35 (0.35%) | 41 (15.02%) |  |
|  |  | 未知 | 24 (0.24%) | 0 (0%) |  |
|  | 是否抗生素 |  |  |  | < 0.001 |
|  |  | No | 2,592 (25.87%) | 37 (13.55%) |  |
|  |  | Yes | 7,426 (74.13%) | 236 (86.45%) |  |
|  | 关联CRP |  |  |  | < 0.001 |
|  |  | 正常 | 592 (5.91%) | 65 (23.81%) |  |
|  |  | 异常 | 252 (2.52%) | 92 (33.7%) |  |
|  |  | 未查 | 9174 (91.58%) | 116 (42.49%) |  |
|  | B族链球菌 |  |  |  | < 0.001 |
|  |  | 阴性 | 9,154 (91.38%) | 160 (58.61%) |  |
|  |  | 阳性 | 412 (4.11%) | 25 (9.16%) |  |
|  |  | 未查 | 452 (4.51%) | 88 (32.23%) |  |
|  | 羊水情况 |  |  |  | < 0.001 |
|  |  | 清 | 6,934 (69.22%) | 183 (67.03%) |  |
|  |  | 混 | 523 (5.22%) | 21 (7.69%) |  |
|  |  | 胎粪 | 548 (5.47%) | 33 (12.09%) |  |
|  |  | 未知 | 2,013 (20.09%) | 36 (13.19%) |  |
|  | **新生儿临床特征** |  |  |  |  |
|  | 出生后1分钟Apgar评分 |  |  |  | 0.001 |
|  |  | 8到10分 | 9,213 (91.96%) | 246 (90.11%) |  |
|  |  | 4到7分 | 106 (1.06%) | 10 (3.66%) |  |
|  |  | 0到3分 | 632 (6.31%) | 16 (5.86%) |  |
|  |  | 未知 | 67 (0.67%) | 1 (0.37%) |  |
|  | 出生后5分钟Apgar评分 |  |  |  | < 0.001 |
|  |  | 8到10分 | 9,930 (99.12%) | 266 (97.44%) |  |
|  |  | 4到7分 | 12 (0.12%) | 5 (1.83%) |  |
|  |  | 0到3分 | 9 (0.09%) | 1 (0.37%) |  |
|  |  | 未知 | 67 (0.67%) | 1 (0.37%) |  |

Data are N (%) or mean (SD). Continuous variables used t test or Mann-Whitney U test and Categorical variable used χ2 test for compare the baseline characteristics of neonatal without infection and with infection.

**Table 2: 8 Predictive Factors in LASSO Regression Model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **P value** | **OR** | **95% C.I.** | |
| **lower** | **upper** |
| 产程开始\_胎膜早破时间差\_h | 0.0324 | 1.004 | 1.000 | 1.008 |
| 母亲宫腔培养无致病菌 | Reference |  |  |  |
| 母亲宫腔培养有致病菌生长 | 0.0564 | 1.478 | 0.989 | 2.207 |
| 母亲宫腔培养没有检查 | ＜ 0.001 | 0.422 | 0.307 | 0.581 |
| 孕周\_足月 | Reference |  |  |  |
| 孕周\_近足月早产 | ＜ 0.001 | 4.030 | 2.650 | 6.127 |
| 孕周\_远足月早产 | ＜ 0.001 | 5.050 | 2.368 | 10.773 |
| 孕周\_未知 | 0.9982 | 0.000 | 0.000 |  |
| 白细胞计数 | ＜ 0.001 | 1.137 | 1.085 | 1.192 |
| 中性粒细胞比例 | 0.0464 | 0.987 | 0.975 | 1.000 |
| 关联CRP正常 | Reference |  |  |  |
| 关联CRP异常 | ＜ 0.001 | 4.216 | 2.692 | 6.605 |
| 关联CRP未知 | ＜ 0.001 | 0.406 | 0.277 | 0.596 |
| B族链球菌正常 | Reference |  |  |  |
| B族链球菌异常 | ＜ 0.001 | 3.915 | 2.429 | 6.308 |
| B族链球菌未知 | ＜ 0.001 | 3.463 | 2.197 | 5.459 |
| 羊水情况清 | Reference |  |  |  |
| 羊水情况\_混 | 0.1444 | 1.507 | 0.869 | 2.612 |
| 羊水情况\_胎粪 | ＜ 0.001 | 3.293 | 2.125 | 5.104 |
| 羊水情况\_未知 | 0.1259 | 0.728 | 0.484 | 1.093 |

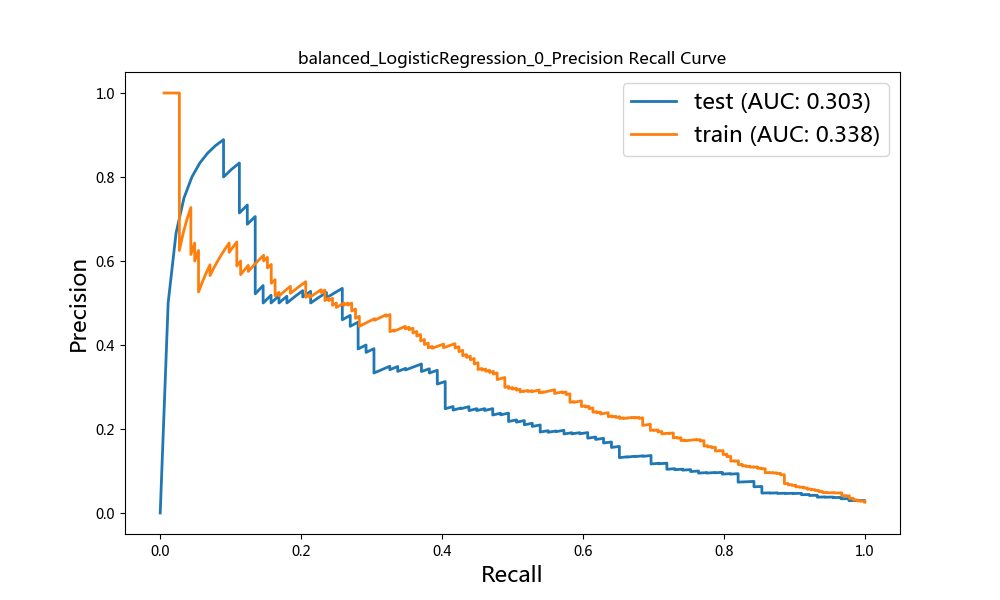
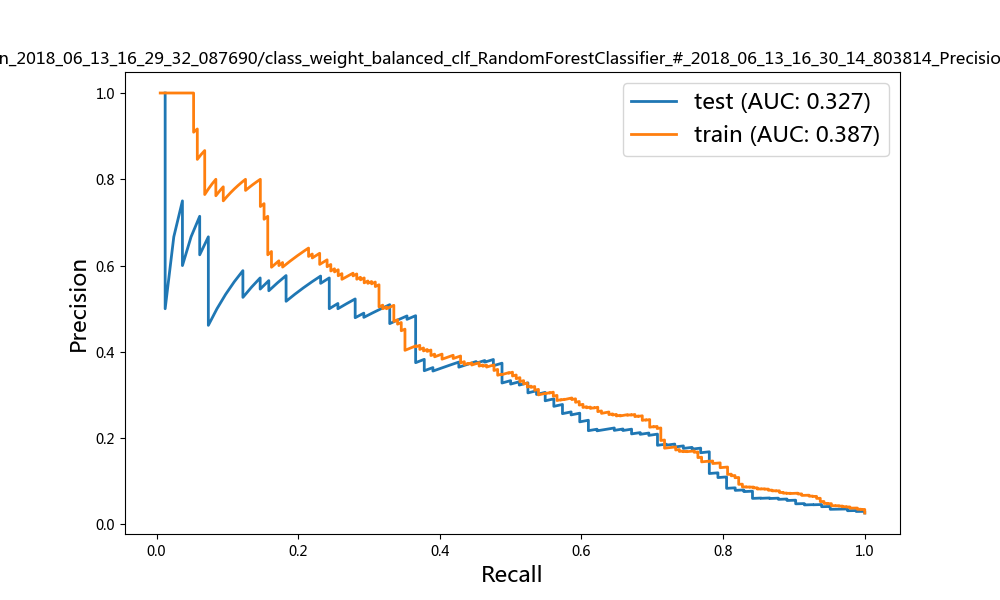
**Figure 3: ROC Curves of Logistic Regression Model and Random Forest Model**



Left: the ROC crues of training and test logistic regression model. The AUC are 0.902 (green, train) and 0.841 (red, test) respectively.

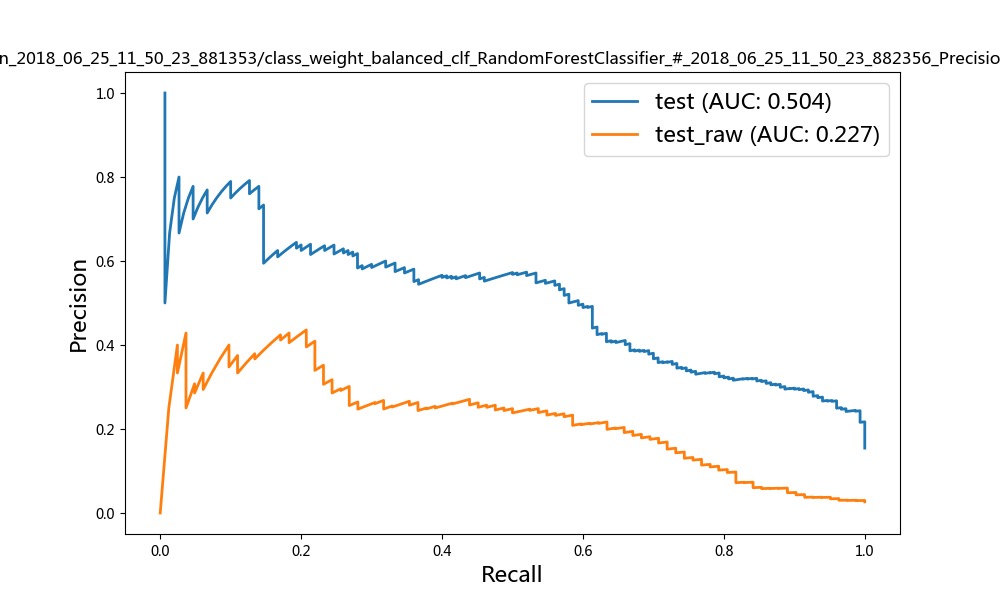
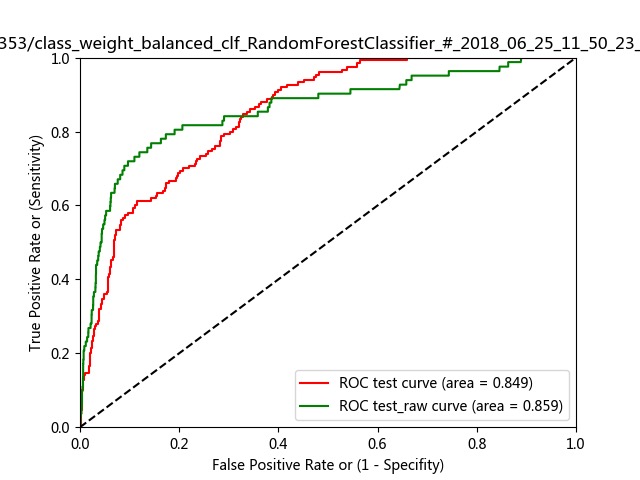
Right: the ROC crues of training and test random forest model (Model 1). The AUC are 0.901 (green, train) and 0.877 (red, test) respectively. The difference of AUC on test set between LR model and RF model is statistically significant. (Z test, *P* = ??? )

**Figure 4: Precision Recall (PR) Curves of Logistic Regression Model and Random Forest Model**

Left: the PR crues of training and test logistic regression model. The AUC are 0.338 (orange,train) and 0.303 (blue, test) respectively. Right: the ROC crues of training and test random forest model (Model 1). The AUC are 0.387 (orange,train) and 0.327 (blue, test) respectively. The difference of AUC on test set between LR model and RF model is statistically significant. (Z test, *P* = ??? )

**Figure 5: ROC and Precision Recall (PR) Curves of Random Forest Model in CRP Cohort**



Left: ROC curve of random forest model in neonatal who measured CRP. The AUC are 0.849 (red, CRP patient) and 0.859 (green, all patient) respectively.Right: PR curve of random forest model in neonatal who measured CRP . The AUC are 0.504 (blue, CRP patient) and 0.227 (orgien, all patient) respectively.

**Table 3：亚组患者（测量CRP患者，N = 1001）的倾向性评分匹配前后的基线分析**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **匹配前**  **（CRP正常/异常=657/344）** | **匹配后** | | |
| **变量** |  | **P value** | **关联CRP正常（N=344）** | **关联CRP异常（N=344）** | **P value** |
| 体温 |  | ＜ 0.001 | 31.58 (13.42) | 32.38 (12.38) | 0.425 |
| 产程开始至胎膜早破时间差/h |  | 0.943 | 21.01 (37.17) | 22.02 (32.25) | 0.709 |
| BMI |  | ＜ 0.001 | 13.82 (10.46) | 13.96 (10.95) | 0.866 |
| 白细胞计数 |  | ＜ 0.001 | 11.29 (4.56) | 11.95 (5.42) | 0.09 |
| 中性粒细胞比例 |  | ＜ 0.001 | 76.67 (16.46) | 77.17 (15.38) | 0.685 |
| 胎膜破裂至分娩时间差/h |  | 0.638 | 29.06 (37.25) | 30.68 (32.24) | 0.551 |
| FT3 |  | 0.831 | 0.92 (2.57) | 0.98 (2.59) | 0.759 |
| TT3 |  | 0.589 | 0.01 (1.46) | 0.05 (1.49) | 0.709 |
| 母亲白带培养 |  | 0.001 |  |  | 0.011 |
|  | 无致病菌生长 |  | 18 (5.42%) | 13 (3.92%) |  |
|  | 有致病菌生长 |  | 4 (1.2%) | 17 (5.12%) |  |
|  | 没有检查 |  | 310 (93.37%) | 302 (90.96%) |  |
| 母亲宫腔培养 |  | ＜ 0.001 |  |  | 0.099 |
|  | 无致病菌生长 |  | 145 (43.67%) | 135 (40.66%) |  |
|  | 有致病菌生长 |  | 41 (12.35%) | 61 (18.37%) |  |
|  | 没有检查 |  | 146 (43.98%) | 136 (40.96%) |  |
| 妊娠期高血压 |  | 0.026 |  |  | 0.364 |
|  | No |  | 324 (97.59%) | 320 (96.39%) |  |
|  | Yes |  | 8 (2.41%) | 12 (3.61%) |  |
| 分娩方式 |  | ＜ 0.001 |  |  | 0.254 |
|  | 自产 |  | 241 (72.59%) | 222 (66.87%) |  |
|  | 产钳 |  | 75 (22.59%) | 93 (28.01%) |  |
|  | 剖腹产 |  | 16 (4.82%) | 17 (5.12%) |  |
| 孕周 |  | ＜ 0.001 |  |  | ＜ 0.001 |
|  | 足月 |  | 270 (81.33%) | 291 (87.65%) |  |
|  | 近足月早产 |  | 51 (15.36%) | 16 (4.82%) |  |
|  | 远足月早产 |  | 11 (3.31%) | 24 (7.23%) |  |
|  | 未知 |  | 0 (0%) | 1 (0.3%) |  |
| 是否抗生素 |  | ＜ 0.001 |  |  | 0.793 |
|  | No |  | 33 (9.94%) | 31 (9.34%) |  |
|  | Yes |  | 299 (90.06%) | 301 (90.66%) |  |
| B族链球菌 |  | ＜ 0.001 |  |  | 0.196 |
|  | 阴性 |  | 272 (81.93%) | 276 (83.13%) |  |
|  | 阳性 |  | 23 (6.93%) | 13 (3.92%) |  |
|  | 未查 |  | 37 (11.14%) | 43 (12.95%) |  |
| 产后1分钟Apgar |  | 0.176 |  |  | 0.596 |
|  | 8到10分 |  | 303 (91.27%) | 297 (89.46%) |  |
|  | 4到7分 |  | 7 (2.11%) | 11 (3.31%) |  |
|  | 0到3分 |  | 22 (6.63%) | 24 (7.23%) |  |
| 产后5分钟Apgar |  | 0.145 |  |  | 0.484 |
|  | 8到10分 |  | 327 (98.49%) | 323 (97.29%) |  |
|  | 4到7分 |  | 2 (0.6%) | 5 (1.51%) |  |
|  | 0到3分 |  | 3 (0.9%) | 4 (1.2%) |  |
| 羊水情况 |  | ＜ 0.001 |  |  | 0.993 |
|  | 清 |  | 234 (70.48%) | 235 (70.78%) |  |
|  | 混 |  | 27 (8.13%) | 28 (8.43%) |  |
|  | 胎粪 |  | 29 (8.73%) | 27 (8.13%) |  |
|  | 未知 |  | 42 (12.65%) | 42 (12.65%) |  |
| 感染\_结局名称？ |  | ＜ 0.001 |  |  | ＜ 0.001 |
|  | No |  | 298 (89.76%) | 245 (73.8%) |  |
|  | Yes |  | 34 (10.24%) | 87 (26.2%) |  |

Data are N (%) or mean (SD). Continuous variables used t test or Mann-Whitney U test and Categorical variable used χ2 test for compare the baseline characteristics of neonatal without infection and with infection.