## 研究目的

本文的研究假设是，胃癌合并贫血患者，与没有合并贫血的胃癌患者，具有不同的临床特征和实验室检查结果，这些差异可能进一步影响胃癌患者的贫血状况的进展。

为了回答这个这个研究假设，我们（1）评估胃癌合并贫血的患者，与未合并贫血的患者，在癌症非进展期的临床特征（具体是什么临床特征呢，列出来）是否存在差异，并（2）探索纳入患者在随访时的贫血状况（叫外周血血红蛋白含量变化吗）变化与基线数据和实验室检查结果变化的相关关系。

## 材料和方法

### 实验设计

从 2015 年至 2022 年北京中医药大学东直门医院、北京朝阳三环肿瘤医院、 北京市朝阳区桓兴肿瘤医院、中国中医科学院西苑医院就诊的晚期胃癌患者中筛选病例进行回顾性分析，这些患者的数据需要包括一个基线时间点和一次随访。为了回答研究目的（1），基线数据首先被分析。患者按照基线贫血与否被分成贫血组和非贫血组，比较两组患者的一般资料和临床数据（需要陈列：具体是哪些临床数据），从而寻找影响贫血分组的因素。对于研究目的（2），我们结合随访数据进一步评价血红蛋白变化和实验室检查数据变化的相关性。首先，根据“患者纳入”一节关于胃癌患者在随访时关于贫血状况改善，恶化和稳定的定义，患者被分成三组，分别是贫血改善组，贫血恶化组和贫血状况平稳组。比较三组患者的一般资料的差异，探索实验室检查数据较基线数据的变化在三组患者间的差异，并定性地考察这种差异的变化方向。更进一步地，通过随机森林模型评估这些实验室检查变化对于贫血预后的重要性。最后，通过桑基图评估胃癌患者关于贫血状态的预后，即展示在同一批患者中，基线数据中的两个分组（贫血组/非贫血组）是如何转移到随访数据中的三个分组（贫血改善组/贫血恶化组/贫血稳定组）的。

（# 画一个流程图来展示这个研究设计）

### 患者纳入

研究中纳入胃癌患者共计（）名，其中基线（初次入组时）胃癌合并贫血的患者（）名，胃癌未合并贫血的患者（）名。胃癌的诊断标准基于（）指南，所有患者处于稳定期。贫血的诊断标准基于（）指南，为血红蛋白（）；所有患者基线数据收集后的（）周进行随访，根据（）指南，我们定义随访时血红蛋白显著上升的标准为上升10g/ml，贫血状况可能得到改善；随访时血红蛋白下降10g/ml，则认为血红蛋白明显下降，贫血状况可能出现恶化；血红蛋白波动在-10g/ml到10g/ml之间的患者，我们认为他们的血红蛋白含量稳定，贫血状况并没有改善或者恶化（斟酌一下这个说辞）。

### 数据收集

在基线和随访的时间点，收集所有纳入患者的一般资料，包括性别，年龄，KPS评分；胃癌疾病信息，实验室检查结果，免疫学指标等。

表1：纳入的基线一般数据

|  |  |  |  |
| --- | --- | --- | --- |
| 变量名 | 变量类型 | 单位/分类 | 备注 |
| 年龄 | 数值型 | 岁 |  |
| 性别 | 二分类 | 男/女 |  |
| bmi | 数值型 | 比值，无单位 |  |
| kps | 数值型 |  | 量表结果 |
| 确诊部位 | 无序多分类 | 残胃癌，胃底，胃窦， 贲门 |  |
| 病理类型 | 无序多分类 | 鳞癌， 腺癌，粘液细胞腺癌，印戒细胞癌 |  |
| 分化程度 | 有序多分类 | 低分化<中低分化<中分化<高分化 |  |
| 用药时间 | 二分类 | 大于三个月/小于三个月 |  |
| 基线是否贫血 | 二分类 | 贫血组/非贫血组 |  |

表2：基线和随访的实验室检查结果和免疫学指标

|  |  |  |  |
| --- | --- | --- | --- |
| 变量名 | 变量类型 | 单位/分类 | 备注 |
| 基线血红蛋白 | 数值型 |  |  |
| CRP |  | 变量被进行了自然对数转换 |
| IL6 |  |
| mcv |  |
| mch |  |
| mchc |  |
| 中性粒细胞绝对值 |  |
| 淋巴细胞绝对值 |  |
| 中性粒比淋巴细胞 |  |
| CD3差值 |  |
| CD3CD4 |  |
| CD3CD8 |  |
| CD3CD4CD25.CD3CD4 |  |
| CD3CD4CD25FoxP3.CD3CD4 |  |
| CD3/CD19 |  |
| CD3/CD16 |  |
| CD4/CD8 |  |

### 排除标准

（什么样子的排除标准呢？未能随访的人？随访后肿瘤恶化的人？）

### 统计分析方法

本研究中的所有计量资料均表示为中位数（四分位间距），计数资料表示为频数（百分比）。为了克服部分非正态分布的变量，由于尺度散布较大对数据分析带来的负面影响，我们对这些变量做了自然对数转换。

通过假设检验来评估基线数据的变量在贫血组和非贫血组之间是否存在差异。计量资料差异由非参数检验（Mann-Whitney检验）进行评估，两分类的计数资料的差异由卡方检验来评估，多分类的计数资料差异借助Fisher精确检验来评估，而P值低于0.2认为该变量在两组间存在显著差异。为了进一步鉴别基线数据中贫血组和非贫血组的独立危险因素，我们纳入在单因素分析中显著的变量，进行多因素logistics回归，并汇报比率比（Odds Ratio）和对应的置信区间作为结果。此时，P值低于0.05认为该变量对结果有显著影响。

通过单因素方差分析来评估变量在三个组之间（Hb下降组/ Hb稳定组/Hb上升组）是否存在差异，计量资料和计数资料的假设检验方法同上文，p值低于0.2视为显著。更进一步地，对于实验室检查和免疫指标数据，进行组间比较，并汇报未调整的p值。

随机森林模型是一种用于拟合分类问题常见机器学习模型，它借助许多随机生成的决策树“投票”决定最后的分类的结果，并允许我们评估自变量对的分类因变量的重要程度。更具体地，当一棵决策树基于某个变量对样本进行分类时，优化的目标是使得分类前后两类样本（在这里，是患者子集合）在基于Gini系数下更加同质化，即让样本 的Gini系数下降。在这些随机生成的决策树形成的“森林”中，每个变量的平均下降的Gini系数（mean decrease Gini）可以作为变量重要性的评估。为了更进一步评估随访时刻的贫血是否改善（Hb下降组/ Hb稳定组/Hb上升组）是如何被实验室检查和免疫指标的变化所影响，我们以随访时的三个分组为因变量，实验室检查和免疫指标的变化自变量，构建随机森林模型，并根据平均下降的Gini系数（mean decrease Gini）将自变量的重要程度排序作为结果给出。

本文所有数据清洗，统计分析和可视化都是借助R软件（版本4.2.0, 美国Bell实验室）完成的。

### 伦理学批准

## 结果

### 基线数据单因素分析

为了分析患者的基线数据，包括一般人口学资料，肿瘤治疗相关的信息，以及实验室检查结果和免疫学指标，患者首先根据贫血与否被分成两组。单因素分析首先被进行，它允许我们通过假设检验比较单个变量的分布（频数或者分位数）在两个分组间是否存在统计学差异。假设检验的结果在表3和表4被呈现，以p<0.2为阈值，我们可以发现年龄，kps，确诊部位，用药时间，CD3CD4，mcv和mch，淋巴细胞绝对值是显著的。关于用药，我们可以看到两组患者选择是否接受中医治疗是有不同倾向的，91%的合并贫血患者治疗时间超过三个月，而未合并贫血的患者中，这个数目仅有49%。

表3: 比较基线时刻, 胃癌合并贫血与未合并贫血的患者的一般资料和临床特征 (单因素分析).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variables | Total (n = 101) | 贫血组 (n = 54) | 非贫血组 (n = 47) | p | statistic |
| 年龄, Median (Q1,Q3) | 67 (58, 76) | 71.5 (58, 77) | 64 (58.5, 71.5) | 0.138 | 1487 |
| 性别, n (%) |  |  |  | 0.716 | 0.133 |
| 女 | 68 (67) | 35 (65) | 33 (70) |  |  |
| 男 | 33 (33) | 19 (35) | 14 (30) |  |  |
| bmi, Median (Q1,Q3) | 20.7 (18.37, 22.95) | 20.85 (18.94, 22.72) | 20.24 (17.81, 23.19) | 0.536 | 1360.5 |
| 化疗有无, n (%) |  |  |  | 0.959 | 0.003 |
| 无 | 12 (12) | 7 (13) | 5 (11) |  |  |
| 有 | 89 (88) | 47 (87) | 42 (89) |  |  |
| kps, Median (Q1,Q3) | 80 (70, 90) | 70 (60, 80) | 90 (72.5, 90) | < 0.001 | 733 |
| 确诊部位, n (%) |  |  |  | 0.948 | Fisher |
| 残胃癌 | 1 (1) | 0 (0) | 1 (2) |  |  |
| 胃体 | 39 (39) | 20 (38) | 19 (40) |  |  |
| 胃底 | 1 (1) | 1 (2) | 0 (0) |  |  |
| 胃窦 | 29 (29) | 16 (30) | 13 (28) |  |  |
| 贲门 | 30 (30) | 16 (30) | 14 (30) |  |  |
| 病理类型, n (%) |  |  |  | 0.722 | Fisher |
| 不详 | 7 (7) | 4 (7) | 3 (6) |  |  |
| 印戒细胞 | 2 (2) | 2 (4) | 0 (0) |  |  |
| 印戒细胞癌 | 3 (3) | 2 (4) | 1 (2) |  |  |
| 粘液细胞腺癌 | 1 (1) | 1 (2) | 0 (0) |  |  |
| 腺癌 | 87 (86) | 44 (81) | 43 (91) |  |  |
| 鳞癌 | 1 (1) | 1 (2) | 0 (0) |  |  |
| 分化程度, n (%) |  |  |  | 0.885 | Fisher |
| 中低分化 | 18 (20) | 9 (18) | 9 (23) |  |  |
| 中分化 | 21 (24) | 13 (26) | 8 (21) |  |  |
| 低分化 | 48 (54) | 27 (54) | 21 (54) |  |  |
| 高分化 | 2 (2) | 1 (2) | 1 (3) |  |  |
| 用药时间, n (%) |  |  |  | < 0.001 | 19.461 |
| 大于三个月 | 72 (71) | 49 (91) | 23 (49) |  |  |
| 小于三个月 | 29 (29) | 5 (9) | 24 (51) |  |  |

表4: 比较基线时刻, 胃癌合并贫血与未合并贫血的患者的实验室检查结果和免疫指标 (单因素分析).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variables | Total (n = 101) | 贫血组 (n = 54) | 非贫血组 (n = 47) | p | statistic |
| CRP, Median (Q1,Q3) | 1.17 (0.08, 2.58) | 1.58 (0.2, 2.55) | 1.13 (0.12, 2.54) | 0.563 | 1354 |
| IL.6.0.7pg.ml., Median (Q1,Q3) | 2.41 (1.4, 3.41) | 2.41 (1.4, 3.4) | 2.41 (1.68, 3.42) | 0.585 | 1188.5 |
| mcv, Median (Q1,Q3) | 90.2 (85.6, 94.4) | 88.9 (81.93, 92.55) | 91.9 (88.05, 96.15) | 0.023 | 935 |
| mch, Median (Q1,Q3) | 3.41 (3.34, 3.45) | 3.39 (3.26, 3.45) | 3.42 (3.38, 3.48) | 0.008 | 878.5 |
| mchc, Median (Q1,Q3) | 5.78 (3.56, 5.82) | 5.77 (5.68, 5.8) | 5.79 (3.54, 5.83) | 0.412 | 1148 |
| 中性粒细胞绝对值, Median (Q1,Q3) | 1.28 (0.79, 1.63) | 1.27 (0.51, 1.67) | 1.28 (1.02, 1.63) | 0.622 | 1196 |
| 淋巴细胞绝对值, Median (Q1,Q3) | 0 (-0.42, 0.26) | -0.1 (-0.58, 0.23) | 0.05 (-0.39, 0.32) | 0.125 | 1043 |
| 中性粒/淋巴, Median (Q1,Q3) | 1.32 (0.82, 1.95) | 1.42 (0.73, 2.09) | 1.21 (0.9, 1.63) | 0.622 | 1342 |
| CD3, Median (Q1,Q3) | 4.27 (4.11, 4.33) | 4.28 (4.15, 4.33) | 4.26 (4.1, 4.33) | 0.764 | 1313.5 |
| CD3CD4, Median (Q1,Q3) | 3.69 (3.58, 3.83) | 3.68 (3.6, 3.89) | 3.69 (3.53, 3.79) | 0.173 | 1469.5 |
| CD3CD8, Median (Q1,Q3) | 3.16 (2.9, 3.54) | 3.24 (2.91, 3.54) | 3.15 (2.91, 3.52) | 0.788 | 1309 |
| CD3CD4CD25.CD3CD4, Median (Q1,Q3) | 3.04 (2.66, 3.44) | 3.14 (2.69, 3.58) | 3.03 (2.66, 3.24) | 0.28 | 1428 |
| CD3CD4CD25FoxP3.CD3CD4, Median (Q1,Q3) | 1.46 (1.16, 1.84) | 1.46 (1.16, 1.93) | 1.46 (1.27, 1.83) | 0.741 | 1220 |
| CD3.CD19., Median (Q1,Q3) | 2.28 (1.96, 2.47) | 2.28 (1.89, 2.38) | 2.3 (2.02, 2.78) | 0.252 | 1100.5 |
| CD3.CD16., Median (Q1,Q3) | 2.4 (1.92, 3.06) | 2.49 (1.92, 3.08) | 2.24 (1.88, 2.75) | 0.268 | 1432 |
| CD4.CD8, Median (Q1,Q3) | 0.74 (0.1, 0.88) | 0.79 (0.11, 0.99) | 0.64 (0.1, 0.8) | 0.303 | 1420.5 |

注：此时的实验室检查结果和免疫指标都时自然对数转换后的结果，即log转换。

### 基线数据多因素分析结果

正如上文所提及，作为单因素分析的结果，以p<0.2为阈值, 可以看到到年龄，kps，CD3CD4，mcv和mch，淋巴细胞绝对值，以及关于用药时间在两个分组之间是存在显著差异的。考虑将这些因素作为自变量纳入多因素分析模型中，以基线贫血与否的分组作为因变量，构建二分类的logistics回归模型。

在构建logistics回归模型的过程中，有需要注意的细节。一方面，由于mcv和mch高度相关，由于logistics回归模型假设自变量不会出现共线性，我们需要避免将高度相关的自变量放入模型，所以取mcv纳入多因素分析的模型中即可。另一方面，正如上一节所讨论的，为了评估两组患者在用药时间是否超过三个月的选择上存在较强的不均衡是如何影响logistics回归结果的，我们构建了两个嵌套的logistics回归模型。嵌套的模型意味着，其中一个模型纳入了年龄，kps，CD3CD4，mcv，淋巴细胞绝对值作为自变量，另一个模型在纳入第一个模型的自变量的同时，还纳入用药时间作为自变量。因此，从所纳入的自变量的集合来看，前者是“嵌套”在后者里面的。

两个嵌套模型的结果在表5给出。在这个表格中, 非贫血组编码为1, 贫血组被编码为0。考察第一个模型（即不纳入用药时间的模型），以kps为例，它的OR为1.03，代表在其他自变量都不变的情况下，当kps每升高一个单位，从贫血组（编码为0）转移到非贫血组（编码为1）的优势比将乘以1.03。直觉地讲，即kps越高, 越不容易贫血（病例落到非贫血组的概率越大）。比较两个嵌套模型，对于第一个模型，以p<0.05为阈值，则kps和淋巴细胞绝对值是显著的。第二个模型在纳入用药时间后，以p<0.05为阈值，则kps和用药时间是显著的。然而，留意到用药时间的OR值置信区间为（6.67, 169），意味着很大的标准误和抽样误差。因此，我们最终采纳不纳入用药时间的logistics二分类回归模型的结果。

表5: 构建两个嵌套的logistics回归模型

|  | **不纳入'用药时间'** | | | **纳入'用药时间'** | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **OR**1 | **95% CI**1 | **p-value** | **OR**1 | **95% CI**1 | **p-value** |
| 年龄 | 1.00 | 0.97, 1.04 | >0.9 | 1.02 | 0.98, 1.08 | 0.3 |
| kps | 1.03 | 1.01, 1.07 | 0.019 | 1.06 | 1.02, 1.11 | 0.002 |
| mcv | 1.04 | 0.99, 1.10 | 0.12 | 1.06 | 1.01, 1.14 | 0.046 |
| 淋巴细胞绝对值 | 2.24 | 1.05, 5.18 | 0.045 | 1.83 | 0.72, 5.00 | 0.2 |
| CD3CD4 | 0.36 | 0.06, 1.84 | 0.2 | 0.26 | 0.04, 1.59 | 0.15 |
| 用药时间 |  |  |  |  |  |  |
| 大于三个月 |  |  |  | — | — |  |
| 小于三个月 |  |  |  | 27.6 | 6.67, 169 | <0.001 |
| 1OR = Odds Ratio (比率比), CI = Confidence Interval (置信区间) | | | | | | |

~~在这个表格中, 非贫血组为1, 贫血组为0, OR值以1为分隔, OR>1的表示正相关, OR<1的表示负相关.~~

* ~~以”kps”为例, 它的OR为1.03, 代表在其他自变量都不变的情况下, 当kps每升高一个单位, 从贫血组 (编码为0) 到非贫血组 (编码为1)的优势比将乘以1.03. 直觉地讲, 就是kps越高, 越不容易贫血 (病例落到非贫血组的概率越大).~~

~~(这个表格其实有细节问题, 即0-1的编码反了. 在这个表格中, 非贫血组为1, 贫血组为0, 一般来说, 会标记有病 (阳性) 的那个组为1, 因此这个表格还会返工.~~

### 随访数据单因素分析结果

为了分析随时血红蛋白的变化（Hb下降组/ Hb稳定组/Hb上升组）与一般资料，临床信息，基线血红蛋白，以及实验室检查和免疫指标变化之间的关系，我们通过单因素方差分析来评估这些变量在三个组之间是否存在差异。比较的结果呈现于表6，表7和表8。以p<0.2为阈值，我们看到性别，kps，确诊部位，用药时间，基线是否贫血，基线血红蛋白在三组间存在显著差异。关于随访的数据和基线数据差异，我们可以找到CD3CD4CD25FoxP3.CD3CD4差异值，CD3.CD16差异值，CRP差异值，中性粒细胞绝对值差异值，淋巴细胞绝对值差异值在三组间存在显著差异。

表6: 评估随访时, 一般资料, 临床信息和基线血红蛋白在三组患者中是否存在差异的结果 (单因素分析)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Total (n = 101) | Hb上升组 (n = 55) | Hb下降组 (n = 23) | Hb不变组 (n = 23) | p | statistic |
| 年龄, Median (Q1,Q3) | 67 (58, 76) | 65 (55.5, 76) | 66 (59, 72.5) | 70 (65.5, 77) | 0.221 | 3.022 |
| 性别, n (%) |  |  |  |  | 0.153 | 3.75 |
| 女 | 68 (67) | 36 (65) | 13 (57) | 19 (83) |  |  |
| 男 | 33 (33) | 19 (35) | 10 (43) | 4 (17) |  |  |
| bmi, Median (Q1,Q3) | 20.7 (18.37, 22.95) | 20.86 (18.62, 23.1) | 20.42 (16.8, 22.67) | 19.84 (17.07, 21.99) | 0.287 | 2.5 |
| 化疗有无, n (%) |  |  |  |  | 0.597 | Fisher |
| 无 | 12 (12) | 5 (9) | 3 (13) | 4 (17) |  |  |
| 有 | 89 (88) | 50 (91) | 20 (87) | 19 (83) |  |  |
| kps, Median (Q1,Q3) | 80 (70, 90) | 80 (70, 90) | 80 (62.5, 90) | 80 (70, 90) | 0.569 | 1.128 |
| 确诊部位, n (%) |  |  |  |  | 0.018 | Fisher |
| 残胃癌 | 1 (1) | 0 (0) | 0 (0) | 1 (4) |  |  |
| 胃体 | 39 (39) | 23 (43) | 9 (39) | 7 (30) |  |  |
| 胃底 | 1 (1) | 0 (0) | 0 (0) | 1 (4) |  |  |
| 胃窦 | 29 (29) | 21 (39) | 5 (22) | 3 (13) |  |  |
| 贲门 | 30 (30) | 10 (19) | 9 (39) | 11 (48) |  |  |
| 病理类型, n (%) |  |  |  |  | 0.87 | Fisher |
| 不详 | 7 (7) | 3 (5) | 3 (13) | 1 (4) |  |  |
| 印戒细胞 | 2 (2) | 1 (2) | 0 (0) | 1 (4) |  |  |
| 印戒细胞癌 | 3 (3) | 3 (5) | 0 (0) | 0 (0) |  |  |
| 粘液细胞腺癌 | 1 (1) | 1 (2) | 0 (0) | 0 (0) |  |  |
| 腺癌 | 87 (86) | 46 (84) | 20 (87) | 21 (91) |  |  |
| 鳞癌 | 1 (1) | 1 (2) | 0 (0) | 0 (0) |  |  |
| 分化程度, n (%) |  |  |  |  | 0.427 | Fisher |
| 中低分化 | 18 (20) | 11 (22) | 3 (19) | 4 (18) |  |  |
| 中分化 | 21 (24) | 15 (29) | 1 (6) | 5 (23) |  |  |
| 低分化 | 48 (54) | 24 (47) | 12 (75) | 12 (55) |  |  |
| 高分化 | 2 (2) | 1 (2) | 0 (0) | 1 (5) |  |  |
| 用药时间, n (%) |  |  |  |  | < 0.001 | 29.513 |
| 大于三个月 | 72 (71) | 50 (91) | 7 (30) | 15 (65) |  |  |
| 小于三个月 | 29 (29) | 5 (9) | 16 (70) | 8 (35) |  |  |
| 基线是否贫血, n (%) |  |  |  |  | < 0.001 | 14.857 |
| 贫血组 | 54 (53) | 39 (71) | 8 (35) | 7 (30) |  |  |
| 非贫血组 | 47 (47) | 16 (29) | 15 (65) | 16 (70) |  |  |
| 血红蛋白, Median (Q1,Q3) | 100 (85, 116) | 91 (80.5, 102) | 125.6 (90.5, 152.35) | 115 (92, 117) | < 0.001 | 18.64 |

表7: 评估随访时免疫指标的变化在三组患者中是否存在差异 (单因素分析)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Total (n = 101) | Hb下降组 (n = 23) | Hb不变组 (n = 23) | Hb上升组 (n = 55) | p | Statistic |
| CD3差值, Median (Q1,Q3) | 0.02 (-0.1, 0.14) | -0.01 (-0.14, 0.07) | 0 (-0.09, 0.1) | 0.06 (-0.07, 0.18) | 0.218 | 3.048 |
| CD3CD4差值, Median (Q1,Q3) | 0 (-0.27, 0.14) | -0.02 (-0.3, 0.17) | -0.06 (-0.28, 0.1) | 0 (-0.24, 0.2) | 0.682 | 0.766 |
| CD3CD8差值, Median (Q1,Q3) | 0.01 (-0.27, 0.3) | 0.08 (-0.24, 0.25) | 0.12 (-0.16, 0.38) | 0 (-0.34, 0.32) | 0.787 | 0.479 |
| CD3CD4CD25.CD3CD4差值, Median (Q1,Q3) | 0.06 (-0.44, 0.44) | 0.05 (-0.42, 0.27) | 0 (-0.39, 0.4) | 0.11 (-0.44, 0.51) | 0.688 | 0.747 |
| CD3CD4CD25FoxP3.CD3CD4差值, Median (Q1,Q3) | 0.12 (-0.35, 0.55) | 0.42 (-0.03, 0.79) | 0.12 (-0.45, 0.46) | -0.1 (-0.38, 0.5) | 0.04 | 6.447 |
| CD3.CD19.差值, Median (Q1,Q3) | 0.04 (-0.35, 0.36) | -0.12 (-0.27, 0.3) | 0.01 (-0.61, 0.5) | 0.07 (-0.34, 0.37) | 0.79 | 0.471 |
| CD3.CD16.差值, Median (Q1,Q3) | 0.02 (-0.7, 0.7) | 0.18 (-0.42, 0.71) | -0.58 (-1.06, 0.33) | 0.02 (-0.7, 0.78) | 0.117 | 4.296 |
| CD4.CD8差值, Median (Q1,Q3) | -0.07 (-0.5, 0.26) | -0.08 (-0.61, 0.31) | -0.15 (-0.5, 0.19) | 0 (-0.46, 0.33) | 0.831 | 0.37 |

注: 这些指标在基线和随访的数据都是先通过自然对数转换, 而后再做差. 即log(随访数据) – log(基线数据) = log(随访数据/基线数据)

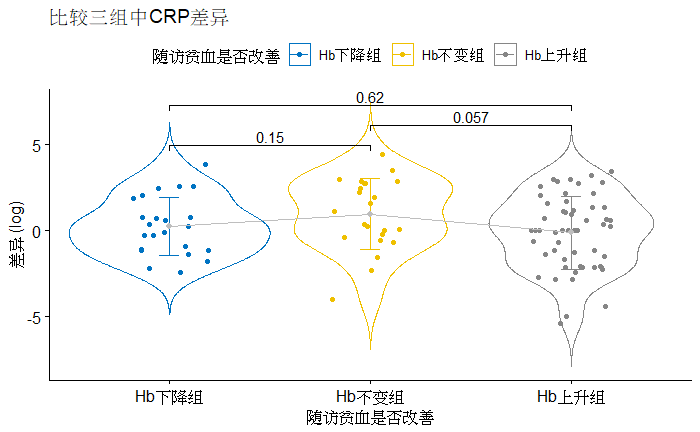
表8: 评估随访时实验室检查结果的变化在三组患者中是否存在差异 (单因素分析)

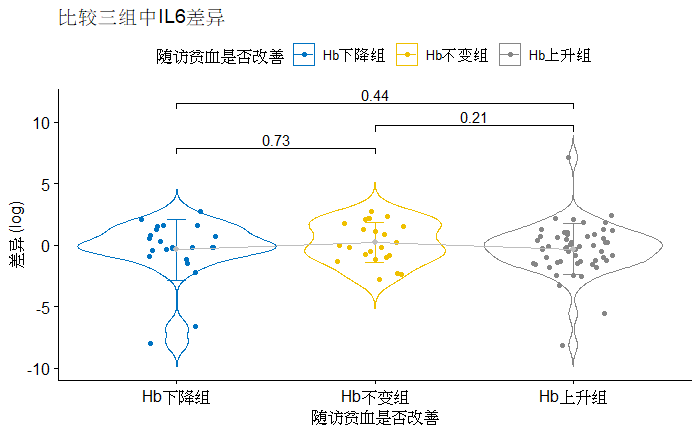
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Total (n = 101) | Hb下降组 (n = 23) | Hb不变组 (n = 23) | Hb上升组 (n = 55) | p | statistic |
| CRP差异, Median (Q1,Q3) | 0.24 (-1.17, 1.91) | 0.25 (-1.12, 1.32) | 1.1 (-0.33, 2.72) | 0.02 (-1.89, 1.33) | 0.138 | 3.967 |
| IL6差异, Median (Q1,Q3) | -0.17 (-1.17, 1.01) | -0.17 (-0.64, 1.06) | 0 (-0.91, 1.65) | -0.23 (-1.31, 0.71) | 0.422 | 1.727 |
| mcv差异, Median (Q1,Q3) | -0.2 (-6.7, 8.9) | 5.2 (-3.3, 12.55) | 0.3 (-5.9, 7.65) | -1.1 (-9.22, 6.05) | 0.34 | 2.159 |
| mch差异, Median (Q1,Q3) | 0.01 (-0.09, 0.09) | 0.03 (-0.08, 0.12) | -0.01 (-0.08, 0.08) | 0.01 (-0.08, 0.09) | 0.916 | 0.176 |
| mchc差异, Median (Q1,Q3) | -0.02 (-2.16, 0.02) | -0.03 (-0.07, 0.02) | -0.02 (-0.07, 0.01) | -0.02 (-2.19, 0.02) | 0.981 | 0.038 |
| 中性粒细胞绝对值差异, Median (Q1,Q3) | -0.04 (-0.68, 0.71) | -0.35 (-1.09, 0.24) | -0.04 (-0.37, 0.89) | 0.02 (-0.77, 0.82) | 0.185 | 3.377 |
| 淋巴细胞绝对值差异, Median (Q1,Q3) | -0.06 (-0.58, 0.49) | -0.5 (-1.07, -0.07) | 0.23 (-0.46, 0.49) | 0.07 (-0.35, 0.66) | 0.013 | 8.756 |
| 中性粒比淋巴差异, Median (Q1,Q3) | -0.05 (-0.78, 0.89) | 0.34 (-0.54, 0.64) | -0.26 (-0.73, 0.95) | -0.16 (-0.82, 0.88) | 0.904 | 0.201 |

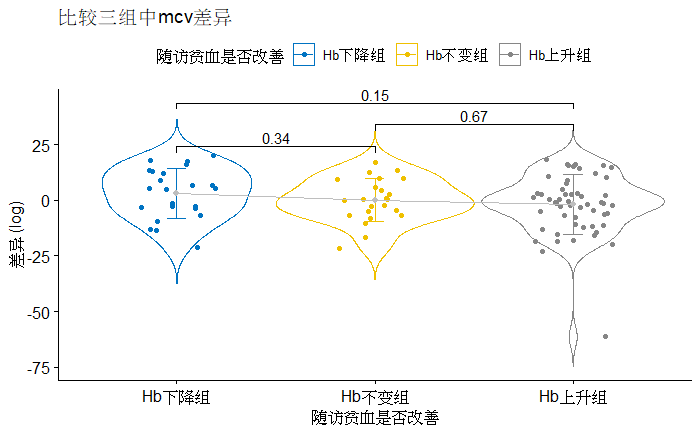
注: 这些指标在基线和随访的数据都是先通过自然对数转换, 而后再做差. 即log(随访数据) – log(基线数据) = log(随访数据/基线数据)

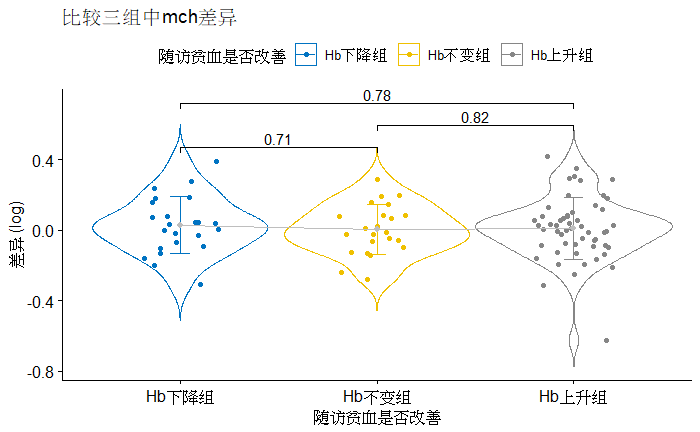
### 随访数据差异的组间比较

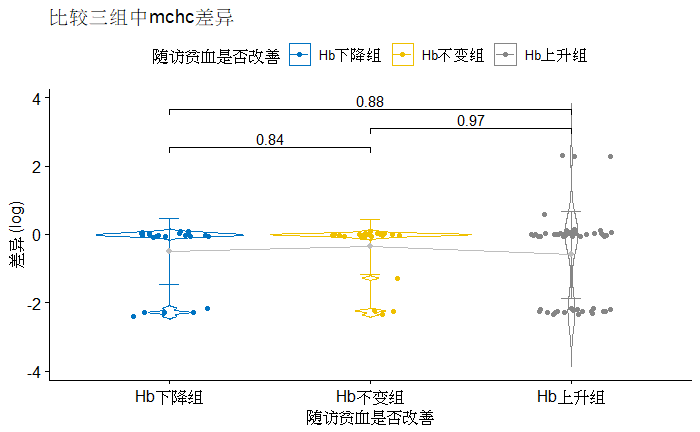
由于随访数据单因素分析结果中的p值是三组间只要任意一对存在差异，p值就会小于0.05。举例说明，假如hb下降组和hb不变组存在差异，但是hb下降组和hb上上升组不存在差异，在这种情况下，p值还是小于0.05。因此，需要进一步做组间比较，下面的图被用来进一步展示组间差异。

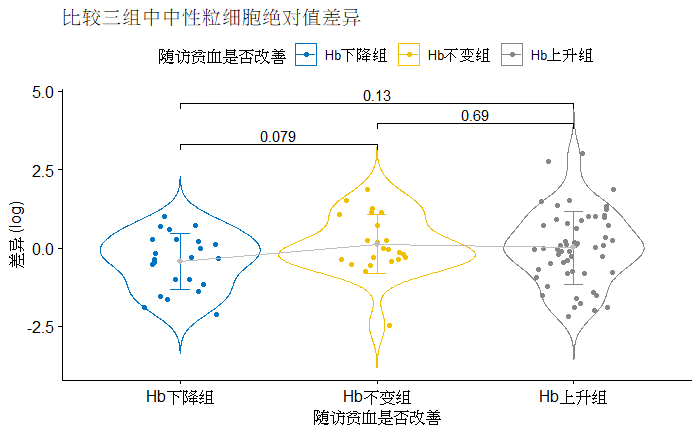


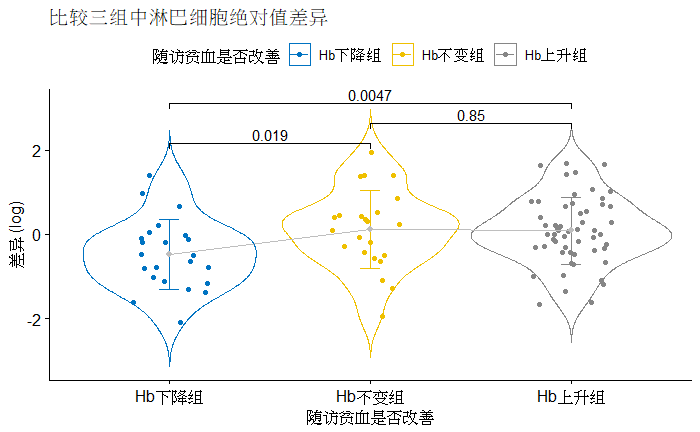


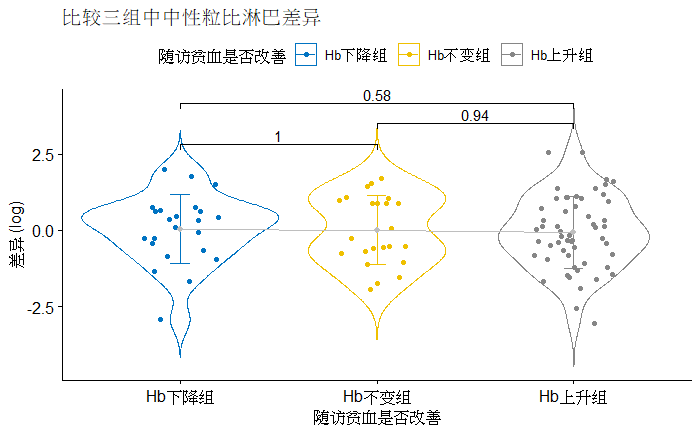


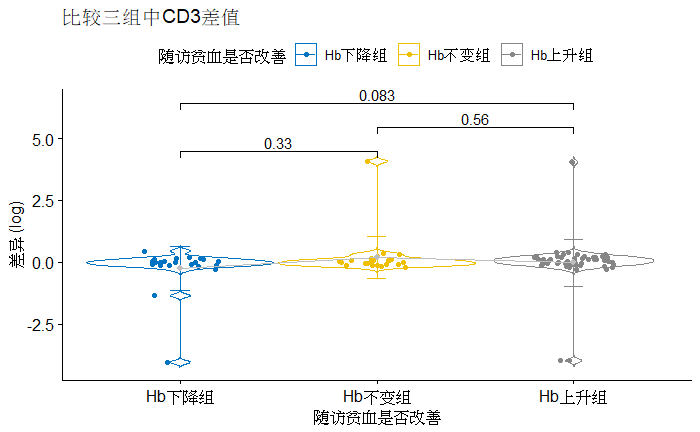


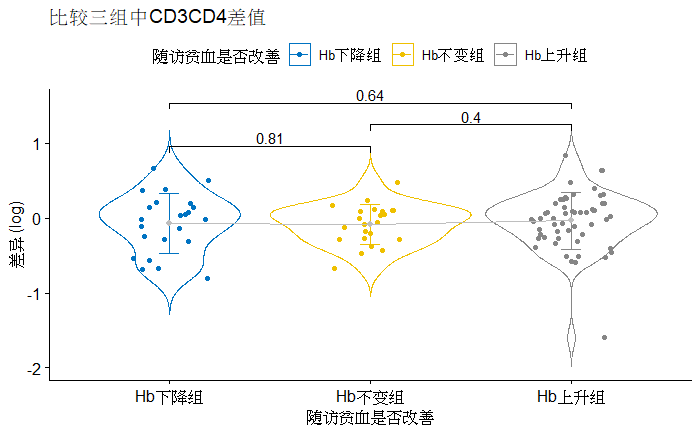


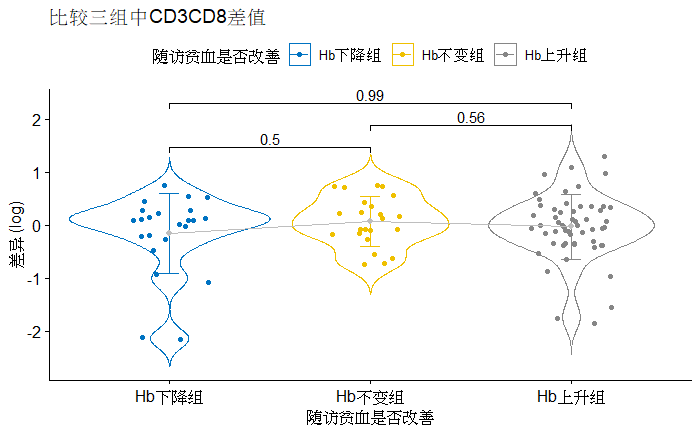


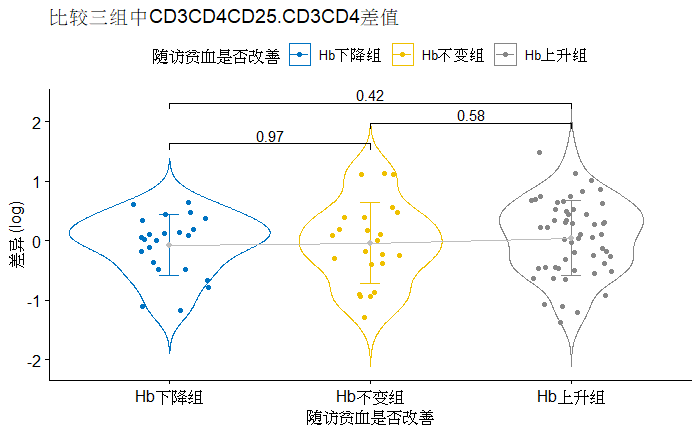


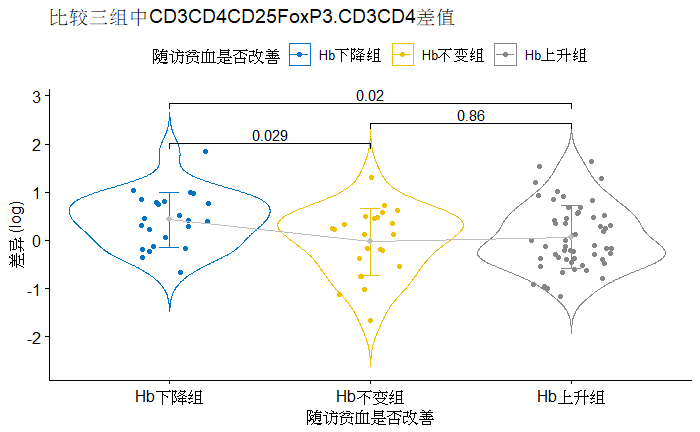


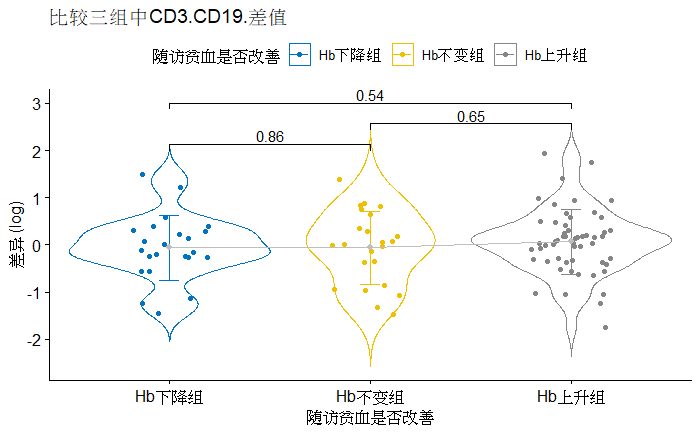


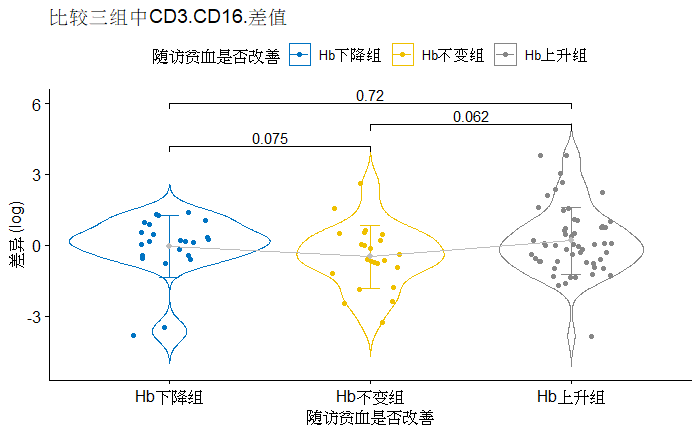


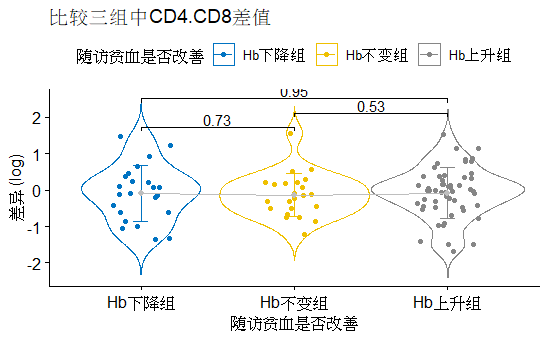












### 随访血红蛋白变化的变量重要性排序

借助随机森林模型的平均下降Gini系数（mean decrease Gini），我们可以得到感兴趣的变量，即随访实验室检查结果的变化值，以及免疫指标变化值，对随访患者落在三个分组中的哪个组别（Hb下降组/ Hb稳定组/Hb上升组）的贡献大小，即可进行特征重要性排序。排序的结果如图所示。

