## 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（陈列：具体时哪些临床数据），从而寻找有意义的影响因素。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，我们定义贫血的改善标准为随访时血红蛋白上升10g/ml, 贫血的恶化标准为血红蛋白下降10g/ml，血红蛋白波动在-10g/ml到10g/ml之间的患者，我们认为他们的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，实验室检查结果等。

(建议画个表格，陈列所有纳入的变量)

### 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（表，这些数据主要是实验室检查结果和免疫指标）。

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（贫血改善组/贫血恶化组/贫血稳定组）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。

#随机森林评估重要性（先空着）

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

# 将免疫学指标加上

# 将表格的编码改回来