力扑素 初步统计计划

06/09/2023

宋秋冬

统计分析方法：

对数据进行数据清洗，包含重新编码，分组，处理缺失值，处理离群值，处理杠杆点等。

一. 描述性分析

包含单变量和双变量分析。对基线信息，力扑素联合PD-1应用现状，既往治疗情况，安全性做描述性分析。应对小基数categorical数据，使用Fisher's exact test去测试各变量跟outcome（PFS）是否有相关性。应对小基数continuous数据，使用Spearman's rank test 去检测相关性。

具体分析如下：

1）基线信息

年龄（中位数，范围），性别（男，女，占比N%），病理类型（鳞癌，腺癌，占比N%），临床分期（III,IV，占比%）

2）力扑素联合PD-1应用现状

总剂量（中位数，范围），力扑素用药周期数（中位数，范围），PD-1 单抗用药周期数（中位数，范围）

联合的 PD-1 种类（列出种类名称，数量，占比N%），联合的铂类化疗（补给）（列出种类名称，数量，占比N%）。选做：可绘制直方图或条形图来表示种类和占比。

3）既往治疗情况（例数、比例）

既往手术情况（例数、比例N%），既往放疗情况（例数、比例N%），统计后线化疗药物，

有手术名称+放疗部位代表做过手术

4）安全性

列出1，2，3级及以上不良事件（数量，占比N%）；列出与免疫相关的不良事件（数量，占比%）

选做：可用条形图（种类多）或散点图（种类少）列出各不良事件（种类，数量），可做折线图看趋势。

二. 生存分析

1）对总体人群、鳞癌、腺癌 算出疾病无恶化存活期

对缺失值做删失处理，根据时间和截尾建立KM生存分析模型，计算中位存活期（mPFS），半年存活期，一年存活期，两年存活期。选做：可使用logrank test做探索性分析用以比较两个或多个组之间的每个观察时间点上的观察事件数和预期事件数。

2）制作K-M 曲线做数据可视化

对于KM模型，核查K-M模型的构建假设：截尾原因不可有用信息（informative），独立性，右截尾，区间截尾，恒定风险。

对于总体人群、鳞癌、腺癌，制作KM曲线。选做：可把鳞癌、腺癌放在同一张图做对比。可使用logrank test

3）选做：可制作cox proportional hazard 模型如果想要探究协变量与风险比（HR）之间的关系

选做：可制作HR表格根据不同的协变量，可报告logrank test的P-value。

三. 疗效评估

针对总体人群，鳞癌，腺癌列出CR、PR、SD表格（数量，占比N%），并算出客观缓解率ORR（CR+PR）和疾病控制率DCR（CR+PR+SD）。