

---

# 临床数据需清洗后才能进行统计分析

作者：李楠，赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/6165.html>

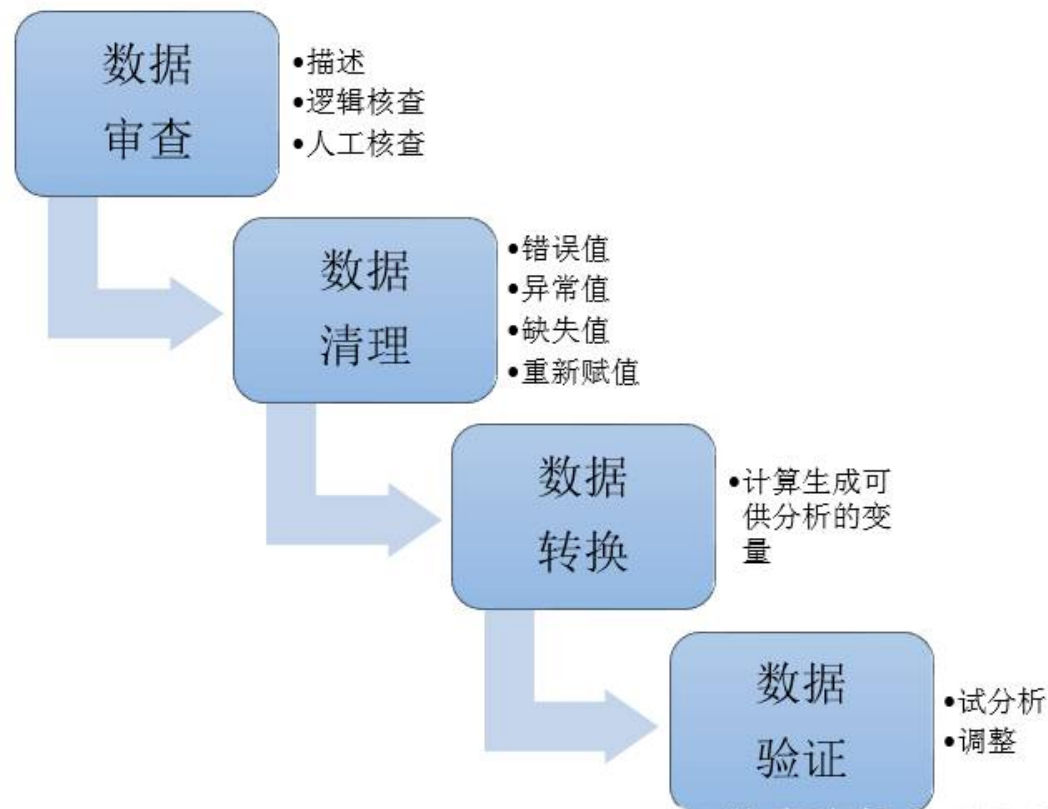
*本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！*

## 临床数据需清洗后才能进行统计分析

。在一般的研究过程中，得到数据库之后，接下来可能就要进入数据分析的阶段了。在经典的临床研究中，我们通常会下大力气在数据收集上，当我们拿到数据之后，肯定会迫不及待的想看看结果。这样会有问题么?答案是肯定的。

即便对于药物上市前的床试验(如三期临床试验)中，我们花费大量经费和人力、引入了完善质量控制措施，所得到的数据仍然存在大量的问题，需要严格的进行数据清洗之后才能进入到统计分析的步骤。相比之下，缺少如此严格数据管理和质量控制流程的一般临床研究，数据质量比较难以直接达到分析的需求。因此，对于任何来源于临床信息的数据，我们在使用前都应该进行数据清洗的过程。

数据的清洗本身是一套相对复杂的工作，临床医生在操作时常觉得找不到规律，无从下手。今天就让我们来简单梳理一下临床数据清洗的大致流程。让我们先来看一个流程图：



临床流行病学和循证医学

根据以往人们的认识和我们的实践经验，我们认为对收集到的数据进行应用之前，至少需要经过以上4个步骤的操作。

### 1、数据审查

，通过对核心数据进行描述、逻辑关系的核查，必要的时候结合人工翻阅数据，尽可能多地找到数据中可能存在的对分析的实施或结果有影响的错误。为下一步对问题数据进行清理奠定基础。

### 2、数据清理

，将审查过程中发现的问题进行分类汇总，结合错误类别、错误对结局的影响、错误的严重程度，对错误数据进行处理。多数情况下会进一步判断发现的“问题数据”是否真的是错误，还是仅仅是偏离了一般情况的正确数据。对于错误数据，通常会做删除处理。如果能够找到数据来源，并得到原始的正确数据，我们也会考虑使用正确的数据对错误数据进行替代。

### 3、数据转换

，在清理数据后，会留下不少因删除错误数据而带来的数据缺失。此时我们会评价缺失数据的类型，进而采取不同的处理方法(详见第116期)。此外我们还需要成一些衍生变量(比如BMI、变量重新赋值、时间日期计算等等)。

### 4、数据验证

，在完成上述步骤之后，我们就可以尝试对数据进行试分析了。试分析并验证数据的可靠性之后，才会最终进入数据分析的流程。

---

对于常见的临床研究，上述数据清理的过程是非常必要的。尤其是现在随着医院信息系统数据结构的完善，越来越多的医生和研究者尝试调用医院信息系统的数据进行数据挖掘和分析，从而回答某一科学问题。而由于数据的初始收集目的是为临床服务，不一定能够满足科学研究的准确性和精确性，此时数据清洗的工作就变得越发重要了。无论如何解决，由什么人解决数据问题，医学研究者至少还是需要对数据分析前必备的过程有一些了解。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发