数据分析过程

1 收集5000万个最常见的搜索查询按每周计数的时间序列，一周内某一特定的搜索查询与搜索查询的总数之比作为一个查询分数

2 收集特定地区ILI相关的医生就诊占总就诊的百分比，该百分比作为医生就诊与ILI相关的概率

3 建立上述查询分数的对数概率与ILI相关的医生就诊对数概率之间的线性关系，系数待定

4 将某一个特定的搜索查询作为解释变量Q，以相关程度测量准确度

5 在对5000万个查询候选进行测试后，选择相关性最高的45个用于模型当中

6 使用2003-2007年间的数据进行拟合，得到适用于各个区域的系数

7 运用模型生成2007-2008年ILI估计数

所犯的错误

在上述第5步中直接使用了相关性最高的45个查询，但高相关性不代表这些查询一定与ILI的发生有因果关系。文中给出的后面55个查询可以说明这个问题

不严谨之处

使用搜索数据进行分析时，未考虑到一些其他情况，例如，出现与流感相近的症状时，人们可能会进行搜索，但此时他们很可能并未患上流感。