

新冠疫情统计数据公布中的科学问题

上海的小乐

眼下，新冠疫情的最新状况应该是全国人民几乎都在关注的一个信息。在这个关注的过程中，我们还能实实在在地学一些科学知识/技能，也可以开展有关“科学/合理地表达”的讨论。

以下图片，都摘自百度“新型冠状病毒肺炎—疫情实时大数据报告”

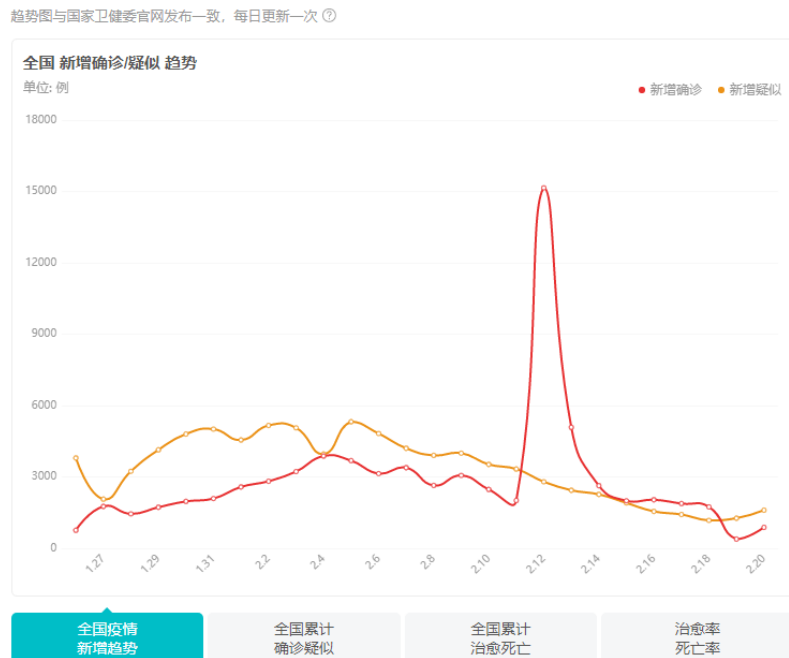
网址：

https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_pc_1

(2020年2月21日的网页内容)

一、图标刻度的变化

在疫情统计数据中，每日新增确诊病例和疑似病例是从一开始就公布的数据（如下图），能非常直观地反映疫情的现状和对后续变化趋势的大概预期。

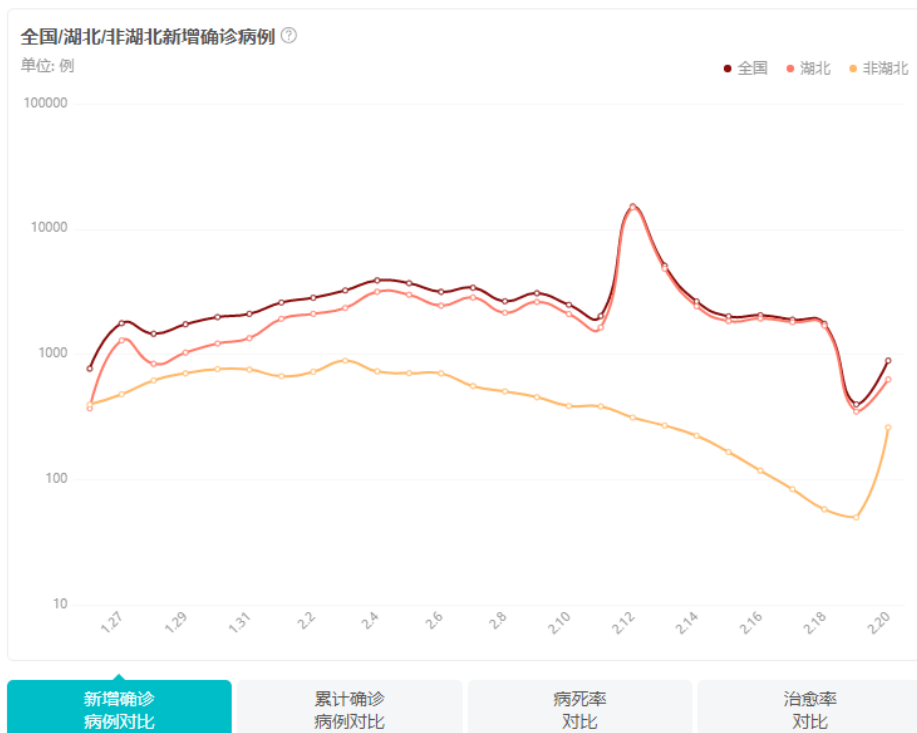


在本次突发新疫情应对的过程中，因为“诊断标准”也是逐渐完善起来的。由于“标准”的修改，统计结果中出现了某一天（上图中的2月12日）的病例

数远大于其他日子的情况。这时，其他日子的病例数及相对变化就较难从这张图中看出来。

为了解决这个问题，百度“新型冠状病毒肺炎一疫情实时大数据报告”在这两天新增了第二张图（如下）：这里的刻度改成了对数坐标。对照前后两张图，大家不难发现：第二张图中，“峰”不那么突出了，数值较小部分“曲线”的变化更容易看清楚了，譬如，我们很容易看出每一天的数值都是在哪个区间；但除了几个特殊的恰好位于刻度线上的几个点，大多数点的准确数值还是比较难以读出的。

对比两张图，我们可以看出不同的刻度标注的特点和适用范围：前一张图，是我们常用的线性刻度，适用于数据变化范围不太大的情形；第二张图，纵坐标是对数刻度，其最大特点是：非常适合于变化范围很大的数据序列的展示。将一个轴的刻度取为对数的展示方式，在两个量之间的依赖关系是指数关系时，特别有效，特别直观；它的另一个常用应用场景是质谱、光谱等的检测领域。以质谱为例，我们经常遇到某个样品中所含的各个成分的比例差别非常大的情形，在样品深度谱的展示时，将纵轴设为对数坐标，既能展示含量很高的元素的变化规律，也能展示含量很少的元素的变化过程。

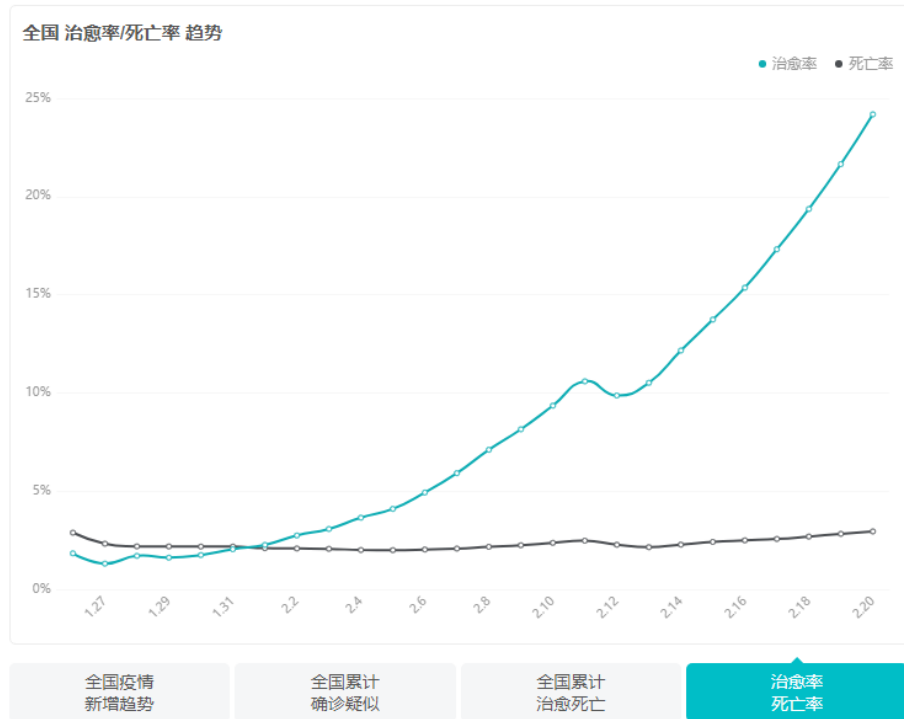


从这个例子可以看出：选择合适的展示方式，对结果展示效果有比较大的影响。平常学习更多的展示技能，在需要的时候就会有用武之地。

二、科学合理地表达

对于下面这张图，不知道大家有何看法？

趋势图与国家卫健委官网发布一致，每日更新一次 ①



作为一个正规的结果，我看到这张图，第一个疑问是：两个比例之和怎么不是 100%？而且还一直在变！

当然，我们都知道这个问题的答案：疫情还没有结束，确诊病例的大多数，现在还在医院救治中。

查阅“百度百科”：治愈率，是指某种疾病平均每百名患者中可治愈的人数，反应其可治愈的概率。

治愈率，反映的是某种疾病可治愈的概率，与其对应的量是“不可治愈率”，或者“病死率/死亡率”。由此，我们不难明白，上图中的两个量之和应该是 100%。因此，这两个量都应该是在一次疫情结束之后，没有了正在治疗的病人，或者正在治疗的病例占总病例的比例很低了，才可以来计算。

当然，为了反映病毒的“毒性”，我们也可以计算“即时/当下”病死率或者死亡率，但前面的定语肯定不应省略。相应地，要反应救治工作的效果，我们应

该用“已治愈比例”一类的概念。

对于这张图，总体感觉：如果表达的不严谨，至少不能反映实际的情形，甚至可能会产生误导作用。衷心希望，媒体，甚至是有关的统计部门，在公布数据的时候，都能越来越严谨，做到科学、合理地表达。