
用挑选出的“病例”和“非病例”算出的ROC曲线可靠么？

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/1507.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！

一、一起来YY

小明发现最新的文献中提出了一个新的指标“神奇标记物”(x marker, xM)，通过定量测量xM，有可能能够帮助诊断疾病“x”(x disease, xD)。对于这种定量指标，当我们试图评价它对区分某一疾病或是预后状态的能力时，往往会使用诊断性研究中常用的统计学方法——ROC曲线，并计算曲线下面积AUC。

小明手头刚好收集了一批临床患者和他们的生物样本。其中总共收集了待诊断是否有xD的患者500例，都保留了生物样本，可供定量测量xM水平。500例患者中，最终通过随访确诊为xD的患者总共有50例。

二、问题来了！

小明想通过绘制ROC曲线来评价xM对xD的诊断价值，下面哪种做法更可取呢：

1、小明必须把这500例患者的生物样本都拿来测一遍，这样才能保证ROC曲线是正确的，有价值的。如果破坏了真阳性率，就有可能导致ROC曲线估计系统性的偏离原始位置，AUC也不可靠了。对么？

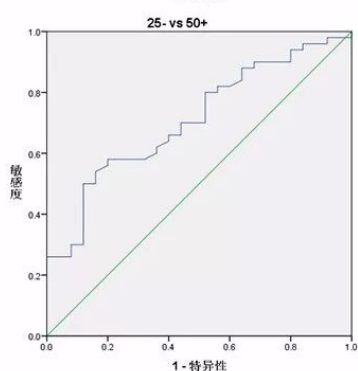
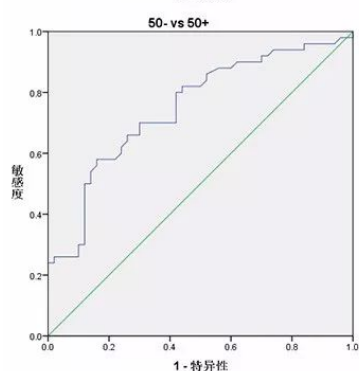
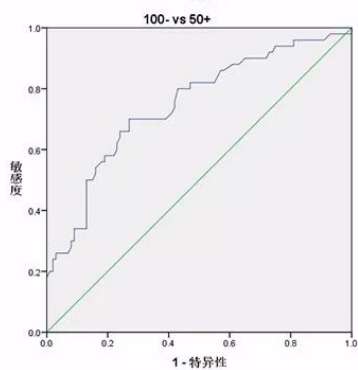
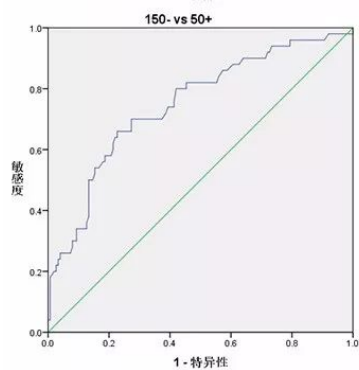
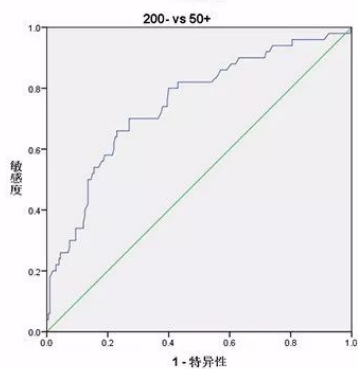
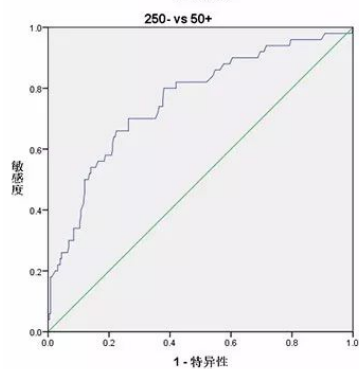
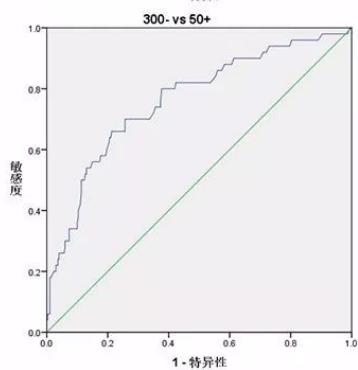
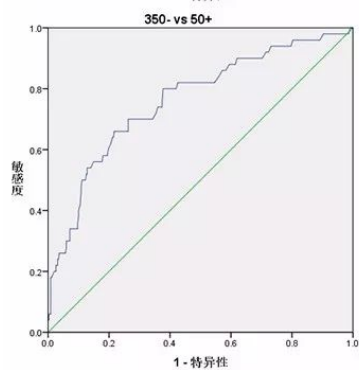
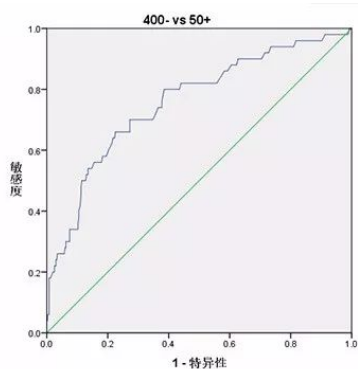
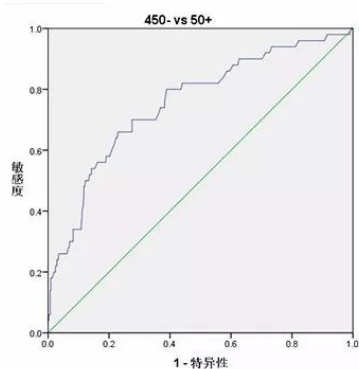
2、把500例生物样本都做了实在太浪费，我们完全可以挑出确诊xD的50例真阳性患者，再从450例肯定没有xD的真阴性患者中随机抽取50例出来。这样只用对100例样本测量xM的水平并且绘制ROC曲线就行了。对么？

三、不看公式看结果

如果您是统计学大拿，或者以前诊断试验相关的知识足够扎实，想必已经在心中得出结论了。如果还没有得出结论，那就先跟着我们看一组模拟数据的结果，相信无需进行公式推导您也能大致了解一二了。

让我们模拟一下，保留全部确诊xD的50例真阳性患者，同时逐渐减少真阴性患者的数量，看看ROC曲线和AUC到底会产生什么变化。

先来看看几个ROC曲线的形状：



看图明显差不多啊!真阴性无论是保留原始的450例，还是逐渐进行随机抽样到最后仅剩25例，画出来的ROC曲线似乎都是一个模样的!所以如果从曲线上找分界点的话，当然也不会差十万八千里啊。

我们当然看到了他们有相似的曲线下面积(AUC)，但是到底有多相似呢?让我们来看看汇总的结果。

是否全部样本	真阴性 (n)	真阳性 (n)	AUC	AUC 95%可信区间		AUC SE	p
				下限	上限		
是	450	50	0.755	0.680	0.830	0.038	0.000
否	400	50	0.758	0.683	0.834	0.038	0.000
否	350	50	0.762	0.687	0.838	0.038	0.000
否	300	50	0.763	0.687	0.838	0.038	0.000
否	250	50	0.760	0.684	0.836	0.039	0.000
否	200	50	0.749	0.671	0.828	0.040	0.000
否	150	50	0.748	0.667	0.829	0.041	0.000
否	100	50	0.746	0.660	0.831	0.044	0.000
否	50	50	0.749	0.653	0.845	0.049	0.000
否	25	50	0.710	0.591	0.828	0.061	0.003

我们看看，好像无论是否对真阴性的患者进行抽样，以及抽样比例如何，AUC的点估计值都非常相近。换句话说，是否保留原始的真阳性、真阴性患者比例，并不太影响AUC的准确性。受影响的只有AUC估计的精度，也就是标准误(SE)和95%可信区间的宽度。

因此，即便是一个破坏了原始比例的横断面研究，或是在真实待诊断人群中随机、不等比例的抽取了部分真阴性、真阳性的患者。ROC曲线和AUC都还是可靠的。总之一句话，如果要用ROC曲线评价某一连续指标的诊断价值时，未必非得测量所有的待诊断人群才行。分别抽取一部分阳性、一部分阴性的患者，得到的结论也是对的。对于上面的两个问题，当然我个人还是倾向于选择2的。不过有时候如果真阳性率不算太高或太低，而且指标的测量不会成本太高，还是鼓励都测测的，毕竟对AUC的估计精度还是有些贡献的。

至于原理嘛，其实很简单。ROC曲线的本质是对特异度和灵敏度的连续计算和展示，而特异度和灵敏度对真阴性和真阳性的构成比并没有什么特殊要求，无论怎么调整都不会发生变化。

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

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