

## 动物群体诊断检测的原理

Principles of diagnostic testing of animal populations

作者: Peter Davies

译自: American Association Of Swine Veterinarians, 2006, pp345-350

译者: 伍少钦

同任何其它行业一样,养猪业的兽医们应用于解释诊断检测的方法已经发生系统化的变化。诊断结果的合理解释需要综合临床的、流行病学和实验观察结果,以及考虑潜在的误差来源。除了传统的疾病爆发调查和疾病诊断应用之外,诊断检测日益被用于部分群体(全体)中疾病的量化,用于支持涉及群体(全体)健康管理的决策。群体检测的最佳策略,包括方法的选择和动物数目的抽样,是检测意图的体现(表1)。基本目的是确定群体(组别、生产线、种群、小群、公司、区域、国家)中的疾病状态(阳性或阴性),或者量化疾病在时间或空间上的趋向。种群统计学的变化(尤其是随着畜群规模的增加)和持续伴有PRRS的问题已经改变了群体诊断的需要和可行性,需要更加精细地应用群体水平的诊断。兽医必须在理论上充分地理解诊断检测的性能、取样理论和疾病流行率之间的关系,加强群体(全体)检测的可靠性。本文综述一些对兽医们来说是基本的且经常会碰到的核心原则。猪群诊断检测和猪群中疾病监测方法方面更加详细的论述可参考其它文章。

**表1 兽医诊断检测的目的 (改编自Gardner 和 Blanchett, 2005)**

检测疾病爆发或非最佳生产的病原因子
评价个体猪只感染/接触的状态
确定某一畜群中是否感染某一病原
确定年龄或生产组是否受影响
评估畜群或猪只对某一感染性因子产生抗体的百分比率
监测畜群对疫苗的血清学反应
监测疾病控制和净化计划的进程与终止

### 疾病诊断时群体采样的基本原理

当打算对一个群体进行疾病检测时,弄明白以下意图是很重要的,通常要么检测疾病,要么估计流行率或发病率。限定这些目标哪个是你首要目标是重要的,因为:

- 适当的样本数目可能会不同的,
- 估计流行率,动物的无偏向选择是必要的;但是,假如是针对动物的高危险疾病检测,有偏向取样是可以理解的。

动物数目的选择受必要条件(精确度和置信区间,成本限制)和假设条件(预期流行率)的影响。检测疾病,决定所需样本大小的因素包括:群体大小、最小的预期疾病流行率(假如发病)和在此流行率上检测疾病的置信区间。执行这些

计算且界面友好的用户软件可以在互联网上免费获的（比如：WinEpiScope 见 <http://www.clive.ed.ac.uk/winepiscope>）。虽然本程序假定检测方法性能是理想的，但是对快速查看样本数量与检测概率之间的关系以及对于评估检测结果都是有用的。

许多兽医已经习惯于熟悉的 30 个样本数（假如流行率  $> 10\%$ ，置信区间在 95%，30 个样本数足以检测出至少 1 头阳性）。但是，重要的是我们要记住：当流行率比较低时，用 30 个样本来诊断疾病的可靠性会不够（图 1）。

注意：当流行率是 5%，从一个大的群体中挑选 30 个样本，其中不包含 1 头单一发病动物的可能性超过 20%。另外考虑阴性检测结果的最大可能的流行率。在 1000 头群体中选取 30 个阴性样本，假定理想的检测敏感度，95% 的置信区间；群体的流行率可以在 9.5% 以下。换句话说，假如你有 10 个动物样本，畜群中的流行率在 25.8% 以下，那么检测结果只有 95% 的可信度。这么小的样本数量只有当预期流行率高 ( $> 25\%$ ) 时，才能提供有意义的检测信息。

用于评估疾病流行率的取样，在试验研究中比临床中用得更加频繁。类似检测疾病的上述例子，关键的决定因子是所需的样本数和群体（全体）大小，所需的置信区间和预期流行率。第 4 个因子是所需的精密度 (precision)，以估计允许出现的误差（比如在 5% 或 10% 以内）来表示。由于这部分的方差最高可达 50%，当流行率非常高 ( $> 90\%$ ) 或者非常低时，估计流行率所需的样本数量相对较少（图 2）。

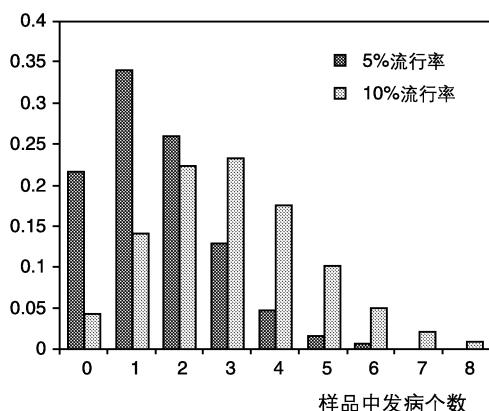


图1 从5%或10%流行群体中挑选30个样本中的预期感染数

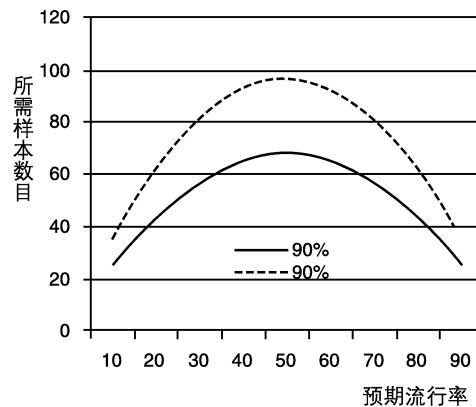


图2 估计疾病流行率在10%以内，置信区间为90%或95%（假定理想的方法），预期流行率对样本数量的影响

缺少预期流行率的资料，一个保守的方法是使用表示 50% 流行率的样本数目。再者，这些指导方针是假定理想的检测性能，依据检测性能的参数可能需要加以修改。

### 个体水平 (individual level) 诊断检测的特性

诊断是一个不完美的过程。有效群体检测一个必不可少的基础是要在个体水平上理解诊断检测的原理。这意味着最低限度地理解检测性能参数(敏感性和特异性)和疾病流行率(或疾病的“验前概率(Pretest probability)”)之间的数量关系，用于确定检测结果的预测值(即动物个体或群体中疾病状况与检测结果一致性的概率)。对猪兽医们来说，这些概念必须十分熟悉，因而简短地回顾一下做为背景知识，用于总体(畜群或群体)水平检测中的讨论。虽然检测数据可能呈二分歧(比如培养结果是阳性或阴性)、离散性(血清学滴度)或连续性(比如OD值或S/P比值)，大部分检测结果可最终被解释为阳性(表明发病或接触过)或阴性(表明无病或没有接触过)。检测试验的敏感性(Se)和特异性(Sp)被定义为分别正确检测出患病和健康个体的概率(图3)。也就是，已知动物的状况(发病或没有)，检测结果是正确的概率是多少？虽然这些对评价检测性能是重要的参数，兽医想要解答的典型问题是检测结果的预测值——即已知检测结果(阳性或阴性)，出现(或没有)疾病的概率有多大？因此阳性预测值(PPV)指检测为阳性动物患病的概率，阴性预测值(NPV)是检测为阴性动物无病的概率。

		疾病状态	
		出现	不出现
检测结果	阳性	真阳性 A	假阳性 B
	阴性	假阴性 C	真阴性 D

敏感性=患病动物检出阳性结果的概率= $A/(A+C)$   
特异性=无病动物检出阴性结果的概率= $D/(D+B)$

假阴性概率=患病动物检出阴性结果的概率= $C/(A+C)$   
假阳性概率=无病动物检出阳性结果的概率= $B/(B+D)$

阳性预测值=检测阳性结果是患病的概率= $A/A+B$   
阴性预测值=检测阴性结果是无病的概率= $D/C+D$

图3 个体诊断检测性能的基本指标

虽然敏感性、特异性和来源于它们二者的其它概念(比如受试者工作特征曲线(ROC)和似然比)，基本上独立于疾病流行率(disease prevalence)，但试验预期值强烈地受到流行率的影响(图4)。更通俗地说，“疾病的验前概率”(本例中指流行率)影响已知检测结果的解释。当疾病的验前概率低(即罕有或不太可能的疾病)，阳性检测的预测值将会低，除非检测方法的特异极高。

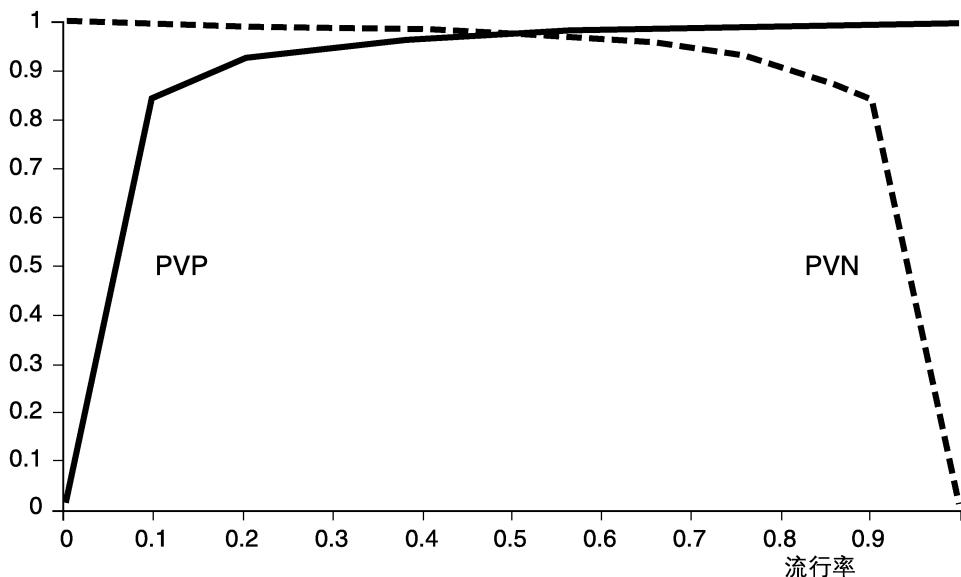


图4 预测值(PVP, PVN)和流行率(敏感性和特异性均为0.95)之间的关系

虽然这些关系在理论上很简单，但它们在实际中的应用却不尽然。几乎总是存在相当多的关于疾病验前概率的不确定性（兽医们清醒地认识到）。然而，评价检测方法的敏感性和特异性也是不确定的。我们往往是查看公布的（或规定的）Se值和Sp值作为修正过且精确的数值。然而，在选定目标群体应用诊断方法时，往往难以获得该方法可靠的Se和Sp估计值。来源于实验疾病模型的估计值，将典型过高地估计检测方法的性能，获得的数据可能并不代表选定目标群体（即应用于检测方法的群体）。获得的估计值很少包括表明关于不确定性的置信区间。而且，通常检测方法的性能在疾病过程的不同阶段是不同的，然而这种变异性很少在检测方法性能指标中表现。理论上，兽医们用于调查研究中的群体诊断方法性能，需要精确和无偏的估计。大多数情况下，可获得的数据达不到这个标准，我们必须在所有诊断检测中承认和处理检测性能方面的不确定因素。当采纳估计检测参数时，我们必须调查这些估计值的来源。

#### 整体（畜群）检测性能的决定因素

Se、Sp和预测值的概念直接从个体移植到整体（群体）规模。畜群诊断检测的敏感性指当感染群体被检测时，该检测方法产生阳性结果的概率；畜群检测的特异性指当无病群体被检测时，该检测方法产生阴性结果的概率。畜群PPV是检测结果为阳性群体中出现疾病的概率；畜群NPV是检测结果为阴性群体中不出现疾病的概率。个体检测性能指标（Se, Sp）明显是畜群水平检测性能的关键性决定因素。然而，考虑畜群检测性能其它的决定因素以及它们间的相互关系也是重要的。同样地，选择个体动物检测的取舍点（比如S/P比值）涉及Se和Sp二者之间的折衷，群体检测规程的设计面临同样的进退两难。表2显示各因子结合数

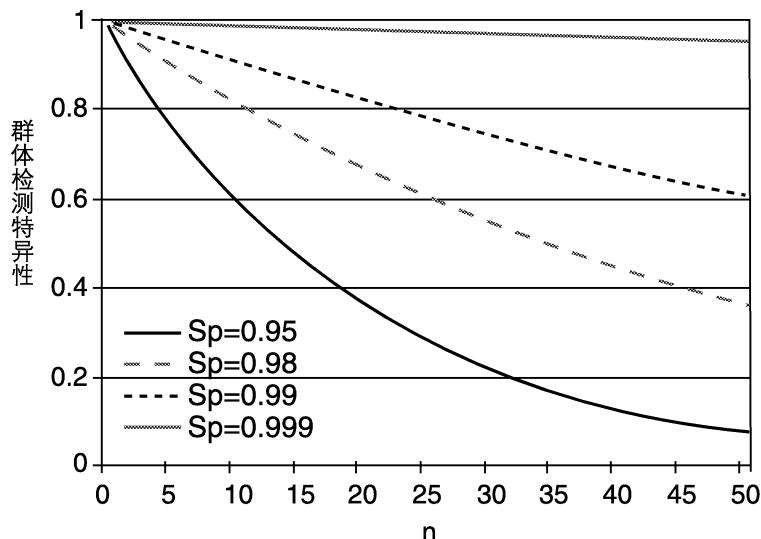
学分析确定全体检测方法的性能。

**表2 群体检测中畜群敏感性 (herd Se) 和畜群特异性 (herd Sp) 的决定因素**

	在畜群流行率之内(P)	n个被检测	个体检测Se	个体检测Sp
畜群Se	+	+	+	+
畜群Sp	- (=0)	+	-	+

#### 畜群检测方法的特异性 (HSp)

根据定义，畜群特异性仅应用于无疾病的群体，是检测方法能正确区分畜群为阴性的概率。特异性的补集（即  $1-HSp$ ），是假阳性检测结果的概率。因为在无病群体中的所有动物是阴性的，检测的敏感性是不相关的； $HSp$  仅取决于2个因素：被检数目(n)和个体检测特异性(Sp)。所有样本为阴性的概率是简单的(Sp) $n$ 。注意  $HSp$  不是按指数 n 规律地增加（图 5）。



**图5 在不同个体检测特异性值(0.95, 0.98, 0.99, 0.999)下，样本数目(n)与畜牧特异性之间的关系**

由此，除非  $Sp$  值极高 ( $>0.99$ )，检测大量的动物通常产生一些假阳性结果，必须对它们进行预计与处理。当兽医们在一个预期低流行率的群体中检测疾病时，这更成问题。在这些情况下，直接检测病原（比如病原分离或PCR）或许比检测抗体更加有效。

#### 畜群检测方法的敏感性 (HSe)

最简单的设想，假如任何被检动物返回阳性的结果，那么这个畜群的检测结果被认为是阳性的。然而，假如一个患病的畜群被检测，观测到阳性结果（“表观流行率 (apparent prevalence)”）的比例可能包括真阳性和假阳性的动物。表观

(不是真实的)流行率是指从一个群体中取样的个体动物被检出阳性结果的概率。在数学上, 真实流行率 (TP) 和表观流行率 (AP) 之间的关系是:

$$AP = TP * Se + (1-TP)(1-Sp)$$

因此, 疾病的真实流行率同个体的 Se 和 Sp 一样, 也影响阳性畜群检测的概率。首先, 让我们看一个简化的情况 (比如细菌培养), 检测方法的 Sp 当作是理想的 (即没有假阳性个体结果)。在这种情况下, 上面公式中的第二项等于 0 (因为  $1-Sp=0$ ), 因此:

$$AP = TP * Se, \text{ 动物检测中阴性的概率是 } 1 - (TP * Se)。$$

为得到一个阳性畜群的检测, 我们只需要检出 1 头感染的个体。假如我们从 n 头动物中抽检 1 个样本, 所有检测结果都是阴性的概率是  $(1-AP)^n$ 。相应地, 畜群检测的敏感性 (大于等于 1 头阳性检测结果) 是  $(1-(1-AP)^n)$ 。图 6 表明对一给定真实流行率范围, 样本数量对畜群 Se 和畜群 Sp 的影响。

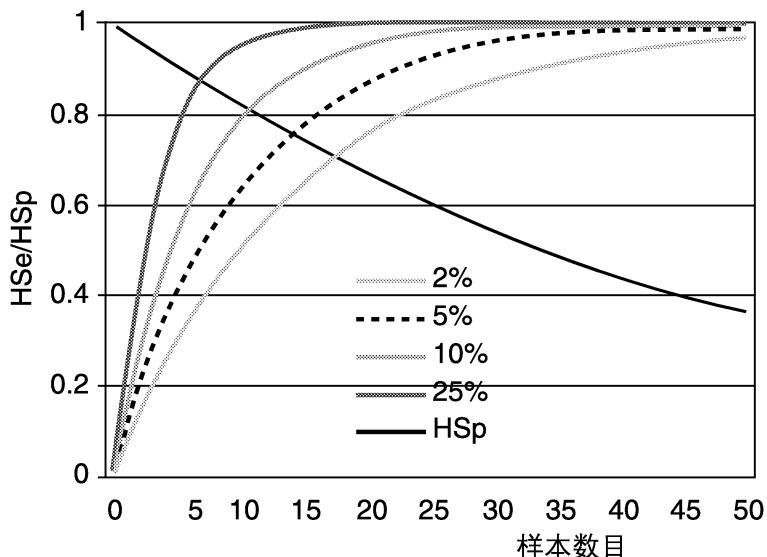


图6 不同真实流行率 (2%, 5%, 10%, 25%) 下, 样本数目与敏感性之间的关系。同时显示个体检测敏感性和特异性为0.95和0.98时, 畜群检测特异性与样本数目之间的变化关系

另一个在评价检测结果中有用的概念是似然比。似然比是患病动物(或畜群)检测阳性概率与阴性动物(或畜群)检测阳性概率的比值。

$$\text{似然比} = \frac{\text{患病个体的检测概率}}{\text{无病个体的检测概率}}$$

比值为 1 表明来源于患病(真阳性)和无病(假阳性动物)的阳性检测结果是一样的, 因而这个检测结果提供无效的信息。通常涉及个体检测评估中的似然比, 指检测敏感性与假阳性率的比值。相应地在畜群检测中, 当流行率和敏感性高时, 会出现高的似然比; 当流行率和检测特异性低时, 会出现低的似然比。在

疾病真实流行率为 2%、Se 为 0.95 和 Sp 为 0.98 时，在无病的畜群中，比有病的畜群中更容易出现假阳性检测结果，但是随着流行率的增加似然比也随之增加（表 3）。

**表3 2%,5%,10%,25%的真实流行率、个体检测敏感性和特异性为95%和98%时，畜群检测的阳性似然比**

真实流行率	真阳性流行率	假阳性流行率	似然比
2%	0.019	0.0196	0.97
5%	0.0475	0.019	2.5
10%	0.095	0.018	5.3
25%	0.2375	0.015	15.8

近年来，出现相当多用于兽医诊断检测解释方面的理论进展，尤其在群体诊断方面 4–11。这些理论引发一些实用工具的发展，以促进设计和阐明群体检测规则的量化方法。比如，Survey Toolbox ([http://www.ausvet.com.au/content.php?page=res\\_software#st](http://www.ausvet.com.au/content.php?page=res_software#st)) 是一套程序软件，它设计用来支持计划和分析动物健康调查。这套软件的一个组成部分是流行病学概率的计算 (FreeCalc)，FreeCalc 支持检测结果用于评估健康状况。FreeCalc 有两种模型，可以计算样本数和分析检测结果的健康状况。它解释诊断检测和样本数目的不完美性，需要输入检测方法的敏感性和特异性、群体大小、最小疾病（假如出现）预期流行率，以及允许使用者设定可接受的误差水平。更加复杂的工具已经可以获得，它延伸本方法合并检测性能特征和预期流行率中的不确定因素。8,9 后者公布说明免费获得的软件 (BayesFreeCalc) 是用于评估健康无病的概率，允许输入参数的不确定性。

#### 校正阳性结果的取舍点 (k) 数目

同样地，ELISA 方法的取舍点涉及内在个体敏感性和特异性的折衷，在畜群检测中可以任意改变动物取舍点数目（即被视为阳性的畜群检测中所需要的阳性数目）。明显地，假如  $k=1$ （即假如 1 个动物被检出阳性，畜群被认为阳性）将得到最大的敏感性，但象上面提到的特异性可能是一个问题。由于成本因素，限制样本数目和使用  $k=1$  作为取舍点有一定的争议，然而对低流行率疾病方法的限制必须清楚。如同个体检测一样，希望高特异性和高敏感性是疾病的验前概率和诊断误差代价（术语叫“误分类代价”）的一个要求。当假阴性结果的代价超过假阳性结果的代价时，高敏感性（即低的取舍点）是必需的。相反，当假阳性检测结果的代价更高，更高的特异性是理想的。这些理论关系表现在误分类代价术语 (MCT) 中：

（下接第 37 页）

(上接第 32 页)

$$MCT = (1-P)(1-Sp) + rP(1-Se)$$

这里  $r$  是假阳性试验代价与假阴性试验代价的比值， $P$  是疾病的验前概率。最佳取舍点被定义为使  $MCT$  最小化的  $Se$  值和  $Sp$  值。再者，在实际水平上，涉及疾病验前概率的不确定因素，HSe 和 HSp 束缚这些量化方法的应用。然而，在大型的生产体系中，常规群体检测的应用规则增加时；更加准确的误分类代价的估计，对诊断进程中的优化是重要的。

#### 畜群检测结果的解释

总之，畜群检测的合理解释依赖于充分理解畜群检测性能的决定性因素。畜群检测的解释比个体检测解释更加复杂。畜群内的疾病流行率类似于个体检测中“疾病的阶段”，主要影响畜群检测的性能。当大量的个体被检测时，个体检测不理想的特异性将被放大。兽医们需要在畜群检测程序中量化假阳性结果的概率，制定处理它们的规程（比如：重新检测规程）。不理想的敏感性在某种程度上问题不大（除非目标要完全确保健康无病），并且可通过增加样本数目加以弥补。畜群检测中的两个判定阈（个体检测的取舍点，阳性动物的取舍点  $k$ ），通过不同的  $n$  和  $k$ （假定可获得个体  $Se$  和  $Sp$  的估计值），在选择 HSe 和 HSp 时引入一些弹性。然而当流行率低（且要求足够的置信区间）时，不大可能可靠地区分感染畜群与未感染畜群。