Binomial distribution and estimation of the true prevalence of infected animals when pooled samples are used

Dear Sir,

A binomial experiment is one that possesses the following properties:

1. The experiment consists of several identical trials.
2. Each trial results in 1 of 2 outcomes; one, a success (negative for contamination), the other, a failure (positive for contamination).
3. The probability of success on a single trial is equal to p and remains the same from trial to trial. The probability of a failure is equal to $q = 1 - p$.
4. The trials are independent.

The binomial theorem can be most easily seen by observing the pattern obtained in calculating $(p + q)^n$ for $n = 1, 2, 3, \ldots$. In a general terms, the following formula is called the binomial theorem:

$$(p + q)^n = p^n + n\,p^{n-1}q + [n(n - 1)]/n!\;p^{n-2}q^2 + \ldots + q^n$$

Example: The published data of Sorensen et al (1). Feces were cultured for the presence of *Mycobacterium paratuberculosis.* Culture was not carried out on individual fecal samples. Approximately 2 g from each of 3 individual 60-mL fecal samples (a total of approximately 6 g) were pooled and mixed. Samples were pooled in the order of sampling. Five hundred fecal pools were tested for *M. paratuberculosis.* Thirty (6.0%) of the 500 fecal pools became overgrown with fungus. No determination could be made with these cultures and they were removed from the subsequent analysis. Sixteen of the 470 readable cultures (3.4%) showed growth of *M. paratuberculosis* (positive) and 454 cultures did not show growth of *M. paratuberculosis* (negative). If we assume that the prevalence of positive and negative fecal pools was q and p, respectively, by using binomial distribution, it could be concluded that:

- p^3 is equal to frequency of fecal pools that 3 animals were not infected with *M. paratuberculosis* (negative).
- $3p^2q$ is equal to frequency of fecal pools that 2 animals were negative and 1 was positive.
- $3pq^2$ is equal to frequency of fecal pools that 1 animal was negative and 2 were positive.
- q^3 is equal to frequency of fecal pools that 3 animals were positive.

As mentioned above, the frequencies of 1, 2, and all 3 individual fecal samples in the positive fecal pool are not equal to each other.

By using the frequency of negative fecal pools (p^3), we can estimate the p, and then the prevalence of infected cows ($q = 1 - p$).

$p = (p^3)^{1/3}$ = (frequency of negative fecal pools)$^{1/3}$ = $(0.9659)^{1/3} = 0.9885$ or 98.85%.

q = frequency of infected cows = $1 - p = 0.0115$ or 1.15%.

The true cow-level prevalence should be estimated by using the sensitivity and specificity of the diagnostic test.

In summary, by using data from pooled samples of n animals (where n = 2, 3, 4, …), we can estimate the true prevalence of infected animals and its standard deviation (*s*) as follows:

True prevalence of infected animals =
$[1 - (\text{Frequency of negative samples})^{1/n}]$
(Specificity − 1)/[Sensitivity + (Specificity − 1)]

s of true prevalence = [(True prevalence)
(1 − True prevalence)/Sample size]$^{1/2}$

In conclusion, Sorensen et al over-estimated the true prevalence of infection at cow-level.

Maryam Ansari-Lari
Department of Food Hygiene and Public Health
College of Veterinary Medicine
Shiraz University, Shiraz, Iran

Reference

1. Sorensen O, Rawluk S, Wu J, Manninen K, Ollis G. *Mycobacterium paratuberculosis* in dairy herds in Alberta. Can Vet J 2003;44:221–226.

Letters to the Editor *on topics of general veterinary interest are solicited, and ongoing debate on controversial topics is encouraged through this feature. Also welcomed are letters which challenge, support, or add to articles appearing in the CVJ in the previous two months. Authors will be allowed one month for reply, so that their reply may appear with the relevant letter in the same issue, usually within two months of receipt. Letters must be signed by all authors, should not exceed 500 words (two double-spaced typewritten pages), and may be abridged and edited as necessary. Financial associations or other possible conflicts of interest should always be disclosed.*

Les **Lettres à la rédaction** sont en principe des textes portant sur des intérêts généraux pour les vétérinaires ou qui ajoutent au débat sur des sujets controversés. Ce sont des textes rédigés à notre demande, mais nous acceptons aussi les lettres qui remettent en question ou appuient les articles qui ont paru dans la Revue dans les deux mois précédents ou y ajoutent quelque chose. Les auteurs auront un mois pour répondre, de façon que leur réponse puisse figurer avec la lettre pertinente dans le même numéro, habituellement dans les deux mois suivant leur réception. Les lettres doivent être signées par tous les auteurs, ne pas dépasser 500 mots (deux pages de texte à double interligne) et peuvent être abrégées et révisées au besoin. Toute implication financière ou autres conflits d'intérêts potentiels devraient être indiqués.