ELSEVIER

Preventive Veterinary Medicine 59 (2003) 181–183

PREVENTIVE VETERINARY MEDICINE

www.elsevier.com/locate/prevetmed

Letter to the Editor

Comments on "Pouillot, R., Gerbier, G., Gardner, I.A., 2002. "TAGS", a program for the evaluation of test accuracy in the absence of a gold standard. Prev. Vet. Med. 53, 67–81"

Recently, Pouillot et al. (2002) developed and made available for public use a method to assess conditional independence of diagnostic tests under a variety of conditions. Two remarks need to be made.

First, the data of Saegerman et al. (1999) are quoted incorrectly. Pouillot et al. (2002) use the data as shown in Table 1, whereas the correct data are shown in Table 2. Apart from an apparent confusion over the various tests, the authors increased the number of reference animals reacting negatively to all three test by adding 562 animals, referred to in the original paper as negative reactors to the skin test and classic tests: results for iELISA were not reported in the original paper and these animals should thus not be considered in the current context. The TAGS output of the analysis of the data in Table 2 is presented in

Table 1
Saegerman et al. (1999) data as quoted by Pouillot et al. (2002)

	Test1, CT	Test2, ST	Test3, iELISA	Pop1	RefInd
1	0	0	0	275	1144
2	1	0	0	6	0
3	0	1	0	33	22
4	1	1	0	11	3
5	0	0	1	0	2
6	1	0	1	0	0
7	0	1	1	6	0
8	1	1	1	32	0

Table 2
Correct Saegerman et al. (1999) data

	Test1, CT	Test2, ST	Test3, iELISA	Pop1	RefInd
1	0	0	0	275	**582**
2	1	0	0	**0**	0
3	0	1	0	**6**	**2**
4	1	1	0	**0**	**0**
5	0	0	1	**33**	**22**
6	1	0	1	**6**	**3**
7	0	1	1	**11**	0
8	1	1	1	32	0

0167-5877/03/$ – see front matter
doi:10.1016/S0167-5877(03)00005-9

```
❶ DATA SUMMARY
1 Population(s);  3 Tests;  1 Reference Population(s)
df: 14 ; parameters: 7
  test1 test2 test3 pop1 RefInd RefInf
1     0     0     0  275    582      0
2     1     0     0    0      0      0
3     0     1     0    6      2      0
4     1     1     0    0      0      0
5     0     0     1   33     22      0
6     1     0     1    6      3      0
7     0     1     1   11      0      0
8     1     1     1   32      0      0

❷           pre1 Sp1 Sp2 Sp3  Se1  Se2 Se3
Best Guess 0.95 0.8 0.8 0.8 0.95 0.95 0.8

❸ EXPECTATION MAXIMISATION
$Iterations
[1] 45
$Likelihood
[1] -467.2169
$Estimations
      pre1    Sp1    Sp2    Sp3    Se1    Se2 Se3
Est 0.1491 0.9967 0.9911 0.9424 0.7014 0.7921   1

❹ NEWTON-RAPHSON
$Iterations
[1] 32
$Likelihood
[1] -467.2169
$Estimations
       pre1    Sp1    Sp2    Sp3    Se1    Se2 Se3
Est   0.1491 0.9967 0.9911 0.9424 0.7013 0.7921   1
CIinf 0.1117 0.9898 0.9824 0.9229 0.5276 0.6006   0
CIsup 0.1962 0.9989 0.9956 0.9572 0.8316 0.9061 NaN

WARNING 1: test results are assumed to be independent conditional on infection
or disease status
WARNING 2: tests are supposed to have constant sensitivity and specificity in
all populations

❻ Expected Results (NR) and Goodness-of-fit test
$Expected
  test1 test2 test3 pop1 RefInd RefInf ExpPop1 ExpRefInd ExpRefInf
1     0     0     0  275    582      0  287.55    566.94         0
2     1     0     0    0      0      0    0.96      1.89         0
3     0     1     0    6      2      0    2.57      5.07         0
4     1     1     0    0      0      0    0.01      0.02         0
5     0     0     1   33     22      0   20.94     34.66         0
6     1     0     1    6      3      0    7.95      0.12         0
7     0     1     1   11      0      0   12.96      0.31         0
8     1     1     1   32      0      0   30.06      0.00         0

❼ $Test
 Max Likelihood: Achievable  Obtained Deviance d.f.      p value
                  -447.7164 -467.2169 39.00098    7 1.953636e-06

$Commentary
[1] "The model does not fit: Assumptions may be not justified"

❽ Residual correlations between test
$ResCor
           Corr1-2     Corr1-3     Corr2-3
pop 1 : 0.03168613 -0.04327117 -0.09322111

The residuals should be randomly distributed around 0
```

Fig. 1. TAGS output for Saegerman et al. (1999) example.

Fig. 1, from which it can be seen that the correlation residuals are markedly lower than those reported by Pouillot et al. (2002) and that apparent lack of specificity (20.94 expected versus 33 observed) is for the iELISA and not the brucellin test. This latter result is to be expected when animals with known false-positive serological reactions (FPSR in Saegerman et al., 1999) are used as the disease-free reference group. The lack of fit of the model is also in line with the differences in epidemiological status of, respectively, the FPSR group and the infected herds, making equality of test sensitivity and specificity in both populations very unlikely.

The second remark is a more fundamental one. Not only does the goodness-of-fit-test pertain to a rather complex null hypothesis (conditional independence *and* constancy of *all* diagnostic test characteristics) and its rejection does not allow a statement about why it was rejected, but it is in principle also the wrong null hypothesis. The null hypothesis should be the safe option, in the present case: there is conditional dependence and/or lack of constancy of test characteristics and, thus, no estimate of prevalence or test characteristics is possible. Reaching unwarranted conclusions because a null hypothesis is not rejected (possibly due to lack of power) should be avoided.

In our opinion, Pouillot et al. (2002) might create the impression that "TAGS" provides an objective check of the independence and constancy assumptions, but the creation of the additional degrees of freedom is based entirely on a subjective appraisal of the situation. Recommending a Bayesian approach, trying to quantify (expert) opinion may be more useful.

References

Pouillot, R., Gerbier, G., Gardner, I.A., 2002. "TAGS", a program for the evaluation of test accuracy in the absence of a gold standard. Prev. Vet. Med. 53, 67–81.

Saegerman, C., Vo, T.-K.O., De Waele, L., Gilson, D., Bastin, A., Dubray, G., Flanagan, P., Limet, J.N., Letesson, J.-J., Godfroid, J., 1999. Diagnosis of bovine brucellosis by skin test: conditions for the test and evaluation of its performance. Vet. Rec. 145, 214–218.

D.L. Berkvens*
N. Speybroeck
Department of Animal Health, Institute of Tropical Medicine
155 Nationalestraat, Antwerp, Belgium

E. Lesaffre
Biomedical Sciences, Catholic University of Leuven
Biostatistical Centre, Leuven, Belgium

C. Saegerman
Ministry of Small Enterprises, Traders and Agriculture
Veterinary Inspection, Brussels, Belgium
* Corresponding author. Tel.: +32-3-2476393; fax: +32-3-2476268
E-mail address: dberkvens@itg.be (D.L. Berkvens)
5 July 2002