Citation:

Greiner, M., Wegscheider, K., Böhning, D. & Dahms, S. (2003). Influential factors for sensitivity and specificity for serodiagnosis of human and porcine trichinellosis. In R. Schulze, H. Holling & D. Böhning (Eds.), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 165-178). Hogrefe & Huber.

11

Influential Factors for Sensitivity and Specificity for Serodiagnosis of Human and Porcine Trichinellosis

Matthias Greiner
International EpiLab
Danish Veterinary Institute, Copenhagen

Karl Wegscheider
Institute for Statistics and Econometrics, Department of Economy
University of Hamburg

Dankmar Böhning
Working Group: Biometry and Epidemiology
Institute for International Health, Joint Center for Humanities and Health Sciences
Humboldt-University in Berlin / Free University Berlin

Susanne Dahms
Institute for Biometrics and Information Processing
Free University Berlin

Summary

The diagnostic sensitivity and specificity of enzyme-linked immunosorbent assays (ELISAs) for the detection of *Trichinella* antibodies in humans and swine was assessed by a systematic, quantitative literature review. The objective was to identify influential factors for specificity and sensitivity covering a wide range of technical and study design characteristics. Nine out of 12 publications selected for analysis reported more than one

pair of sensitivity/specificity. We suggest an explorative-analytical approach that accounts for these "multiple-study type" publications. Two mixed logistic regression models that included study specific explanatory variables, an adjustment for the cut-off value (both fixed effects) and a random effects term (publication) were established for analysis of specificity and sensitivity. The use of an elaborated test antigen was associated with perfect (100%) specificity. In studies that used crude antigen preparations, a positive effect on specificity was associated with publication after 1991, application of the test for humans (versus swine), single-point (versus titration) assays and testing of healthy or non-target (versus other) populations. A positive effect on sensitivity was associated with application of the test for swine (versus humans), testing of other populations than experimentally infected swine or advanced human cases, testing after 26 (versus less than 26) days post infection and medium (versus small, $n < 16$) sample sizes. The impact of the sample size and the status of the positive reference population is obscure and may be due to uncontrolled confounding. The other effects are plausible and show that this form of "exploratory meta-analysis" of diagnostic tests is of practical concern.

11.1 INTRODUCTION

The evidence for the accuracy of a diagnostic test is usually based on multiple primary validation studies rather than on a single study. Multiple studies cover a wider range of marginal conditions such as reference populations, study design and laboratory proficiency and, therefore, are thought of yielding more reliable test performance parameters. The planned multi-centre validation study and the systematic review of published studies are important realizations of a multiple-study based test validation and differ in the extent to which the involved primary studies can be controlled for marginal conditions. Various methods are described for a quantitative summary of multiple validation studies which is here referred to as meta-analysis of diagnostic tests (MADT). These methods include the summary receiver operating characteristic (sROC) analysis (Hurblut III, Littenberg, & Diagnostic Technology Assessment Consortium, 1991; Moses, Shapiro, & Littenberg, 1993), weighted mean values of sensitivity and specificity (Carlson, Skates, & Singer, 1994), relative risk (Mantha et al., 1994), and standardized mean difference (Hasselblad & Hedges, 1995). Irwig et al. (1994) pointed out that a simple pooling of sensitivity (Se) and specificity (Sp) estimates across primary studies is not appropriate because this would underestimate the overall accuracy. The methods for MADT usually presuppose that each primary study contributes exactly one pair of Se and Sp, that is, one data point in the ROC space. We refer this to as *single study type*. The data points are further assumed to be independent. In practice we are concerned with deviations from this ideal situation since published evaluation studies often provide more than one estimate of Se and Sp. We refer such studies to as *multiple-study type* and distinguish three cases. Firstly, more than one test entity (i.e., different tests or technical modifications or application of one test to different host species) is described in

a single publication (*multiple-study type I*). Secondly, a set of different cut-off values is used for evaluation (*multiple-study type II*). Thirdly, multiple reference populations are used (*multiple-study type III*). We further distinguish between an enrollment of distinct (mutually exclusive and independent) reference populations (*multiple-study type IIIa*) and repeated measurements on the same reference population (*multiple-study type IIIb*). Combinations of various multiple-study types may occur. We do not consider the case of multiple reference methods and argue for the selection of the most reliable (in terms of accuracy) method as gold standard instead. The scope of a MADT is usually restricted to a single test entity but situations may occur in which a comparison of the test performance between different test entities is relevant (Irwig, Macaskill, Glasziou, & Fahey, 1995). In analogy with the general meta-analytic terminology we shall refer the estimate of the diagnostic test performance to as effect size.

In this chapter, we describe a meta-analytic approach for the validation of diagnostic tests when the source data include multiple-study type publications. Our data derive from a systematic review of published studies on the validation of ELISAs for the detection of *Trichinella* antibodies in humans and swine. Trichinellosis is a zoonosis with severe medical implications if untreated. The ELISA is recommended for diagnosis of both human (Ljungstrom, 1983) and porcine infection (Gamble, 1997). Furthermore, ELISA testing may become mandatory for certification of "*Trichinella* free" pig production within the framework of an anticipated modified trichinellosis control scheme in countries of the European Union (Borowka & Ring, 1997). The emphasis of our application is to identify influential factors for the diagnostic accuracy covering a broad range of marginal conditions rather than validation of a single test entity.

11.2 MATERIALS AND METHODS

11.2.1 Literature Retrieval

The databases Medline ™, VetCD ™, BeastCD ™, and CAB Helminthological Abstracts™ were used as searching frame as described elsewhere (Greiner, Böhning, & Dahms, 1997). The list of retrieved publications was cross-checked and supplemented by experts (Dr. K. Nöckler and Dr. W. P. Voigt, Federal Institute for Health Protection of Consumers and Veterinary Medicine, Berlin, and Dr. K. Wacker, Institute for Epidemiological Diagnosis, Federal Institute for Virus Diseases of Animals, Wusterhausen). Included for analysis were studies on trichinellosis antibody ELISAs in humans and farm pigs published from 1990 to 1995, where the number of true positive, false positive, false negative and true negative test results was either indicated or derivable from the published data. Furthermore, inclusion required a minimum of 5 subjects for each reference sample. Positive subpopulations were not considered if sampled before day 10 post infection or, in the case of repeated measurements (multiple-study type IIIb), earlier than 10 days after the preceding sampling

date. Generally, language was not an exclusion criterion except Chinese without translation. A list of publications excluded from this study can be obtained from the first author.

11.2.2 Data Transcription

One data base was constructed that included all available estimates of specificity. We considered here (where applicable) different test entities (i.e., distinct technical procedures) described in one publication as well as different applied cut-off values, different negative subpopulations tested and different time points at which the negative subpopulation was sampled for one test entity, respectively. Basic outcome variables were the number of true negative observations (TN) and the sample size of the respective negative reference population (NNEG), respectively (we omit the index for the unit of observation). We considered the sensitivity associated with one unit of observation as the weighted (using the sample size) average of all available sensitivity estimates with the respective test entity. In case of repeated measurements, we selected the first sampling date following the 35th day after infection as base for sensitivity estimation. A second data base was constructed analogously and comprised all available estimates of sensitivity. Basic outcome variables were the number of true positive observations (TP) and the sample size of the respective positive reference population (NPOS). We considered the specificity associated with one unit of observation as the weighted (using the sample size) average of all available specificity estimates with the respective test entity.

A set of variables was recorded as covariate information for each study. The publication YEAR (0 = 1990, 1991, 1 = 1992+) was recorded from the bibliographic data. The variable SPECIES (0 = human, 1 = swine) denotes the species tested. Variables describing technical aspects were AGPREP (coating antigen; 0 = crude preparation or extract of larval antigen, 1 = excretory/secretory (E/S) antigen or purified preparations), CONJUG (specificity of the anti species-enzyme conjugate; 0 = anti whole-Ig fraction, 1 = anti IgM, IgG, or IgE fraction), TITER (dilution of serum samples; 0 = single-point determination, 1 = titration). The selection of a cut-off value in favor of specificity (e.g., the confidence limit of the negative reference population) was coded with SPW (specificity optimized; 0 = no, 1 = yes). Design characteristics were recorded by the variables STATN (status of the negative reference population; 0 = healthy controls, samples from a non-target population or unrelated diseases, e.g., atopic conditions, 1 = any other selection), STATP (status of the positive reference population; 0 = experimental infection or extreme cases, 1 = any other selection), DPI (days post infection at which the positive reference population was sampled; 0 = 10–25, 1 = 26+, 2 = no information), NNEG and NPOS (categorized sample size for the positive and negative reference population, respectively; 0 = 5–15, 1 = 16–50, 2 = 51+) and RESUBST (identity of the reference population for cut-off selection and test validation; 0 = no, 1 = yes).

11.2.3 Analysis of Influential Factors for Specificity and Sensitivity

The analysis of influential factors on specificity and sensitivity was addressed by two mixed logistic-binomial regression models for distinguishable data that take into account the correlation within-publications. The fixed effects term includes the intercept (α_F), the logit transformation of the associated "counter-parameter" (x_i) and the row-vector of explanatory factors ($\mathbf{z}_i$). The random effects term consists of the random effects parameter (α_R). The data were matched on the publication. The models have the general form

$$\text{logit}(p_i) = \alpha_F + \alpha_R + \beta x_i + \boldsymbol{\gamma}^T \mathbf{z}_i.$$

For analysis of specificity, p_i denotes the simple proportion of $\text{TN}_i/\text{NNEG}_i$, X_i denotes the logit transformation of the associated sensitivity (with $1/2$ correction) and $\mathbf{z}_i$ denotes a row-vector of explanatory factors. The variables YEAR, SPECIES, CONJUG, TITER, SPW, STATN, NNEG and RESUBST were selected as candidates for explanatory variables in the analysis of specificity. Multilevel variables were used after dummy coding. The inclusion of the counter-parameter takes care for the part of the variance that can be explained by the applied cut-off value. $\hat{\alpha}_F$, $\hat{\alpha}_R$, $\hat{\beta}$ and the vector $\hat{\boldsymbol{\gamma}}^T$ are empirical coefficients and were found with standard algorithms (logistic-binomial model for distinguishable data with 6 support points; EGRET LBDD(6) module; Statistics and Epidemiology Research Corporation (SERC), 1988). A stepwise backwards fitting strategy was used whereby the variable with the highest p-value of the likelihood ratio statistic (LRS = deviance without/with variable, referred to the chi-square distribution with degrees of freedom, df = number of levels of the variables minus 1) was excluded. This procedure was repeated until the LRS was significant ($p < .05$) for all variables. The goodness-of-fit of the final model was assessed by the LRS of the final model. The candidate variables were also analyzed in univariate mixed logistic regression models. The analysis of sensitivity was accomplished in a complementary manner. Candidates for explanatory variables were YEAR, SPECIES, AGPREP, CONJUG, TITER, STATP, DPI, NPOS, and RESUBST.

11.2.4 Further Analyses

The effect sizes in terms of sensitivity and specificity of the trichinellosis ELISAs were displayed in the ROC space to visualize the scatter of the estimates (Figure 11.1). All possible combinations of a sensitivity and a specificity estimate were considered in case of multiple-type studies. A summary ROC function was established as described by Moses et al. (1993). A chi-square test on homogeneity ($\alpha = .05$; df= number of estimates minus 1) of sensitivity and specificity estimates was done using TP and TN as observed frequencies and $\text{NPOS}\times\widehat{Se}_p$ and $\text{NNEG}\times\widehat{Sp}_p$ as expected frequencies, respectively (Stata macro "chitest" by Nick Cox, personal communication, StataCorp, 1997). Here, $\widehat{Se}_p$ and $\widehat{Sp}_p$ denote the pooled sensitivity and specificity, respec-

tively. Using a total of r estimates of sensitivity,

$$\widehat{Se}_p = \frac{\sum_{i=1}^{r} (\mathrm{TP}_i)}{\sum_{i=1}^{r} (\mathrm{NPOS}_i)}$$

and using s estimates of specificity,

$$\widehat{Sp}_p = \frac{\sum_{i=1}^{s} (\mathrm{TN}_i)}{\sum_{i=1}^{s} (\mathrm{NNEG}_i)},$$

respectively.

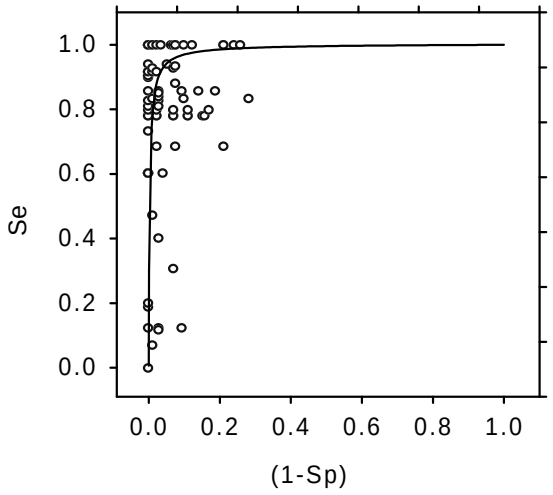

Figure 11.1 Summary ROC plot of the diagnostic specificity (Sp) and sensitivity (Se) of *Trichinella* antibody ELISAs (meta-analysis of 12 studies published between 1990 and 1995). The points represent the reported pairs $(\widehat{Se}, \widehat{Sp})$ in case of a single-type publication and all possible combinations of the two estimates reported for one test entity in case of multiple-type publication (refer to the text for further explanation). The summary ROC function is displayed as solid line.

11.3 RESULTS

11.3.1 Data Transcription

The data from twelve publications (7 on human and 5 on porcine trichinellosis) were included in this meta-analysis (Arriaga, Yepez–Mulia, Morilla, & Ortega–Pierres, 1995; Bruschi, Tassi, & Pozio, 1990; Chan & Ko, 1990; Dzeben-

ski, Bitkowska, & Plonka, 1994; Gamble, 1995; Lind et al., 1991; Mahannop, Chaicumpa, Setasuban, Morakote, & Tapchaisri, 1992; Mahannop, Setasuban, Morakote, Tapchaisri, & Chaicumpa, 1995; Morakote et al., 1991; Morakote, Sukhavat, Siriprasert, Suphawitayanukul, & Thamasonthi, 1992; Nöckler, Voigt, Protz, Miko, & Ziedler, 1995; Serrano, Perez, Reina, & Navarrete, 1992). Three studies belonged to the single study type, nine studies belonged to one of the multiple-study types (Table 11.1). The null hypothesis of homogeneity of the specificity estimates could not be rejected ($\chi^2 = 3.87;\ df = 33,\ p = 1.0$). The null hypothesis of homogeneity of the sensitivity estimates was rejected ($\chi^2 = 132.53;\ df = 55,\ p < .001$). The distribution of study characteristics (here referred to as covariate factors) is described elsewhere (Greiner et al., 1997).

Table 11.1 Types of Evaluation Studies of *Trichinella* Antibody ELISAs Published Between 1990 and 1995 and Number of Analytical Units They Contribute to the Analysis of Specificity and Sensitivity[a]

PUBNR[b]	Study Type	m	c	n	p	t	Specificity	Sensitivity
1	I/IIIa	3	1	4	1	1	12	3
2	IIIa	1	1	2	1	1	2	1
3	IIIa	1	1	2	1	1	2	1
4	I/IIIb	3	1	1	1	7	3	21
5	I/IIIa/IIIb	2	1	2	1	2	4	4
6	single	1	1	1	1	1	1	1
7	I	3	1	1	1	1	3	3
8	IIIa	1	1	2	1	1	2	1
9	single	1	1	1	1	1	1	1
10	II/IIIb	1	2	1	1	5	2	10
11	IIIb	1	1	1	1	9	1	9
12	single	1	1	1	1	1	1	1
Total							34	56

Note. [a]A published primary study that reports only one estimate of sensitivity and specificity is referred to as single study type. A publication for which one or more of the values m, c, n, p and t is greater than 1 is referred to as a multiple study, where m is the number of test entities ($m > 1$ for multiple-study type I), c is the number of cut-off values ($c > 1$ for multiple-study type II), n and p is the number of negative and positive reference populations considered, respectively ($n + p > 2$ for multiple-study type IIIa), and t is the number of time points at which the positive reference populations was evaluated ($t > 1$ for multiple-study type IIIb).
[b]PUBNR=publication number.

11.3.2 Influential Factors for Sensitivity and Specificity

We found a wide range of estimates for specificity and particularly for sensitivity. The variability was not only due to different cut-off values as suggested by the deviations of data points from the summary ROC function (Figure 11.1). The interest was to identify the main reasons for this variability. An overwhelming positive effect of AGPREP on specificity was observed (Table 11.2). Therefore, sub-studies that used an elaborated antigen (AGPREP = 1) were excluded from the analysis of further explanatory factors for specificity.

Table 11.2 Impact of the Type of Antigen Preparation (Crude and Elaborated) Used in *Trichinella* Antibody ELISAs on the Test Specificity ($\widehat{Sp}$) Based on 12 Studies Published Between 1990 and 1995

	Crude (AGPREP = 0)	Elaborated (AGPREP = 1)
$\widehat{Sp} < 1$	19	0
$\widehat{Sp} = 1$	8	7

Using a stepwise backward fitting procedure of a mixed logistic regression model, we identified four variables as potential factors for specificity. Four other candidate variables (CONJ, SPW, NNEG, RESUBST) were excluded due to non-significant LRSs. The extension of the base model that included the fixed effects intercept, the counter parameter, and the random effects term by the four explanatory variables was significant (LRS ($df = 4$) $= 31.4$, $p < .001$). According to the (Wald test significant) effects in the final model, the specificity appeared to be better in studies published after 1991 (YEAR, $p < .001$), better in humans than in swine (SPECIES, $p = .010$), better in single-point assays than in titration assays (TITER, $p < .001$), and better in healthy, non-target or unrelated reference controls than in any other control samples (STATN, $p = .001$). The counter-parameter (logit Se) had a significant ($p < .001$) negative effect (Table 11.3).

The inferences from univariate analyses were consistent for two (TITER, STATN) variables. The effects of YEAR and SPECIES were not discovered whereas NNEG = 1 and RESUBST were associated with a positive and negative univariate effect. For the analysis of potential factors for sensitivity, we selected (stepwise backward fitting procedure) four variables as potential factors for specificity. Five other candidates for sensitivity analysis (YEAR, ANTIG, CONJ, TITER, RESUBST) were excluded due to non-significant LRSs. The extension of the base model by the four explanatory variables was significant (LRS($df = 6$) $= 321.6$, $p < .001$). According to the (Wald test significant) effects in the final model, the sensitivity appeared to be better in swine than in humans (SPECIES, $p = .005$), better in any other than extreme cases or experimental infections (STATP, $p = .005$), better when samples where taken 26 days or more after infection (DPI = 1, $p < .001$), and better in sub-studies that used a sample size between 16 and 50 in than smaller sub-studies (NPOS = 1, $p = .006$). The counter-parameter (logit Sp) had a significant ($p < .001$) neg-

Table 11.3 Coefficients (and Wald Test p-values) From Univariate and Multivariate Mixed Effects Logistic Regression Models for Analysis of Explanatory Variables for the Diagnostic Specificity and Sensitivity of *Trichinella* Antibody ELISAs (Meta-Analysis of 12 Studies Published Between 1990 and 1995)[a]

Variable[b]	Specificity (Sp)				Sensitivity (Se)			
	Univariate		Multivariate		Univariate		Multivariate	
YEAR	0.31	(.439)	1.17	(< .001)	0.08	(.610)	n.i.	
SPECIES	0.71	(.258)	−1.67	(.010)	0.22	(.119)	4.25	(.005)
AGPREP	n.i.		n.i.		0.86	(.001)	n.i.	
CONJUG	−17.40	(.999)	n.i.		−1.58	(.061)	n.i.	
TITER	−1.08	(.018)	−1.70	(< .001)	0.02	(.936)	n.i.	
SPW	0.21	(.637)	n.i.		n.a.		n.a.	
STATN	−2.07	(.006)	−2.43	(.001)	n.a.		n.a.	
STATP	n.a.		n.a.		−0.19	(.162)	4.25	(.005)
DPI	n.a.		n.a.		21.3	(.995)	20.38	(.995)
					3.5	(<.001)	3.65	(< .001)
NPOS	n.a.		n.a.		−0.15	(.620)	0.23	(.455)
					0	(.994)	0.96	(.006)
RESUBST	−0.56	(.046)	n.i.		0.32	(.461)	n.i.	
X	n.a.		−0.57	(< .001)	n.a.		−0.27	(< .001)

Note. n.i. = variable not included; n.a. = variable not applicable.

[a]Final multivariate models (Sp: $n = 27$; Se: $n = 56$) obtained by stepwise backwards fitting starting. Base models included intercept term, counter parameter (logit($\widehat{Se}$) and logit ($\widehat{Sp}$) for analysis of Sp and Se analysis, respectively), and a random effects (RE) term. The coefficient for the RE term was 0.23×10^{-14}, and 0.95 for the Sp and Se model, respectively.

[b]YEAR, publication year (0 = 1990, 1991, 1 = 1992+); SPECIES (0 = human, 1 = swine); AGPREP, coating antigen (0 = crude preparation or extract of larval antigen, 1 = excretory/secretory antigen or purified preparations); CONJUG, specificity of the anti species-enzyme conjugate (0 = anti whole-Ig fraction, 1 = anti IgM, IgG, or IgE); TITER, dilution of serum samples (0 = no titration, 1 = titration); SPW, specificity optimized (0 = no, 1 = yes); STATN, status of the negative reference population (RP) (0 = healthy controls, samples from a non-target population or unrelated diseases, 1 = any other selection); STATP, status of the positive RP (0 = experimental infection or extreme cases, 1 = any other selection); DPI, days post infection (0 = 10–25, 1 = 26+, 2 = no information); NNEG and NPOS, categorized sample size for the positive and negative RP, respectively (0 = 5–15, 1 = 16–50, 2 = 51+); RESUBST, identity of the RP for cut-off selection and validation (0 = no, 1 = yes); X = counter parameter. The base category is 0 for all variables. For variables with more than two categories, the coefficients are shown for categories in decreasing order.

ative effect (Table 11.3). By univariate analysis we found a consistent effect of DPI = 1 whereas discrepant results were obtained for SPECIES, STATP and NPOS = 1 (no effects) and for AGPREP (positive effect).

11.4 DISCUSSION

11.4.1 Parameter Heterogeneity and the Impact of Influential Covariate Factors

The objective of the study was to analyze influential factors for specificity and sensitivity of trichinellosis antibody ELISAs based on a systematic review of the literature. Our intention was not to estimate the "global diagnostic accuracy" of trichinellosis serology. Such an enterprise was strictly invalid due to the inclusion of different test systems in our review. The wide range of marginal conditions in the published studies a priori justified the assumption of parameter heterogeneity (i.e., differences in the diagnostic accuracy between studies). In fact, homogeneity could be rejected for sensitivity but not for specificity. However, since the power of homogeneity tests is generally limited, we continued to investigate explanatory factors for sensitivity and specificity using two separate logistic regression models.

11.4.2 Problem of Multiple Sub-Studies per Publication

Typically, test validation studies are pre-stratified in their design. That means, sample sizes and population characteristics are pre-determined and, consequently, Se and Sp are stochastically independent random variables. A problem arises when multiple estimates of Se and/or Sp are reported in primary studies, for example, when more than two reference samples (multiple-study type IIIa) or repeated measurements (multiple-study type IIIb) are encountered. Obviously, the pooling of type III study data results in a loss of information that may be useful to study influential factors for test accuracy. On the other hand, since meta-analysis generally deals with summary measures of sensitivity and specificity, a complete analysis should involve *all possible combinations* of the reported Se and Sp estimates, which results in a data augmentation. A preliminary analysis of this data set treated these combinations as if they were independent (Greiner et al., 1997). This approach was associated, however, with (i) a substantial violation of the independence assumption, (ii) a severe bias towards significant effects (due to artificially increased sample sizes) and (iii) the risk of bias (with unpredictable direction) due to inappropriate weights. More stringent inclusion criteria and complete avoidance of repeated measurements at the same population were a solution to the problem if the overall goal was to estimate the summary effect size. The separate analysis of specificity and sensitivity may provide a solution to the problem. We have included the counter-parameter into the explanatory part of the models in order to account for the inherent impact of the cut-off value. The lack of independence within publications was considered when we chose a mixed effects

model with a random effects term and the publication as matching variable. This approach allows an estimate of effect sizes in the presence of overdispersion (as caused by correlation within publications).

11.4.3 Interpretation of the Multivariate Analyses

We started by postulating that certain covariate factors may be influential for either sensitivity or specificity. Multivariate models that use summary measures (e.g., Moses et al., 1993; Hasselblad & Hedges, 1995) are not suitable to discover such factors. Our analysis overcomes this problem but is limited through the number of published, eligible studies. Petitti (1994, p. 126) argued that a small number of studies should not preclude the application of regression methods, but the number of explanatory variables should be kept small. Using stepwise backwards fitting, eight and nine variables could be investigated simultaneously for analysis of specificity and sensitivity, respectively. The final mixed effects logistic regression models included four explanatory variables (in each case) besides the counter-parameter.

The test specificity was better in tests that used an elaborated antigen as shown by cross-tabulation (Table 11.2). The results of the multivariate analysis of specificity pertain to studies that used a crude or extract antigen preparation (AGPREP = 0). In these studies, the publication year was positively associated with an increase in specificity. Unobserved changes in technical or other factors (as expressed by the surrogate variable YEAR) may have led to a better specificity. Interestingly, a better specificity and worse sensitivity in humans than in swine was found. This finding might reflect a different medical decision making situation. Trichinellosis serology in medicine usually is a confirmatory instrument (with emphasis on specificity) whereas screening applications (with emphasis on sensitivity) are dominating in veterinary applications. This effect cannot be explained by the choice of the cut-off value because the analysis was adjusted for the counter-parameter. The data also suggest that titration methods do not improve the test properties. In fact, according to our results, titration was associated with worse specificity. Single-point ELISAs - preferred for practical and economic reasons - have been recommended for veterinary seroepidemiologic applications (Wright, Nilsson, van Rooij, Lelenta, & Jeggo, 1993). The selection of reference populations is a critical factor in the evaluation process as pointed out elsewhere (e.g., Knottnerus & Leffers, 1992). It is also well recognized that likelihood ratios of diagnostic tests (i.e., combined expressions of Se and Sp used to establish post-test probabilities) are not invariant to changes in the source population (e.g., Miettinen & Caro, 1994). Our results confirm that the specificity may be overestimated when using healthy or non-target populations or patients with unrelated diseases as negative reference population. Experimental infections (in swine) and clinically advanced cases (in humans) were unexpectedly associated with worse sensitivity than other selections of positive reference populations. The opposite seems to be a common finding in laboratory sciences according to Gerhardt and Keller (1986). The duration of infection prior to sampling is re-

lated to the degree to which specific antibodies have been produced and, thus, can be considered a true factor for sensitivity. The positive effect of medium sample sizes on the sensitivity is obscure and may be due to uncontrolled confounding. Finally, the negative weights of the counter-parameters included in the models underline the inherent effect of the cut-off value. We had expected that other variables such as the type of immunoglobulin detected with the test would contribute to the explanation of the observed variability of sensitivity and specificity as well. We cannot rule out any of those factors since the number of studies included in our analysis was fairly small. Some of the above mentioned effects were also detected by univariate analysis. However, eight discrepant results show that the lack of adjustment for confounding and interaction may lead to invalid inferences.

11.4.4 Limitations

Some potentially important design factors were not included in the analysis because of their distribution. Blinding, for example, has been suggested as a standard for validation studies (Mulrow, Linn, Gaul, & Pugh, 1989). The knowledge of the true disease status might result in biased (too optimistic) accuracy parameters ("test review bias"; Begg, 1987). Only one (human trichinellosis; PUBNR 1) of the reviewed studies indicated that samples were coded prior to analysis. Furthermore, the diagnostic accuracy will be enhanced if test results within an intermediate range ("grey zone") were excluded from sensitivity and specificity calculations. Two studies (one on human and one on porcine trichinellosis; PUBNR 6, 9) used intermediate ranges.

11.5 CONCLUSION

The mixed logistic regression models described above have been found suitable to investigate influential factors for specificity and sensitivity of a diagnostic test based on a quantitative, systematic review of the literature. The approach allows the inclusion of studies that contribute more than one pair of parameters.

REFERENCES

Arriaga, C., Yepez–Mulia, L., Morilla, A., & Ortega–Pierres, G. (1995). Detection of circulating trichinella spiralis muscle larva antigens in serum samples of experimentally and naturally infected swine. *Veterinary Parasitology*, *58*, 319–326.

Begg, C. B. (1987). Biases in the assessment of diagnostic tests. *Statistics in Medicine*, *6*, 411–423.

Borowka, H. J., & Ring, C. (1997). Trichinenfreie Region – Eine realistische Prämisse für den Verbraucherschutz [Trichinella-free region – A realistic premise for consumer protection]. *Fleischwirtschaft*, *73*, 1362–1365.

Bruschi, F., Tassi, C., & Pozio, E. (1990). Parasite-specific antibody responses in trichinella sp. human infection: A one year follow-up. *American Journal of Tropical Medicine and Hygiene*, *43*, 186–193.

Carlson, K. J., Skates, S. J., & Singer, D. E. (1994). Screening for ovarian cancer. *Annals of Internal Medicine*, *121*, 124–132.

Chan, S. W., & Ko, R. C. (1990). Serodiagnosis of human trichinosis using a gel filtration antigen and indirect IgG-Elisa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *84*, 721–722.

Dzebenski, T. H., Bitkowska, E., & Plonka, W. (1994). Detection of a circulating parasitic antigen in acute infections with trichinella spiralis: Diagnostic significance of findings. *Zentralblatt Bakteriologie*, *281*, 519–525.

Gamble, H. R. (1995). Detection of trichinellosis in pigs by artificial digestion and enzyme immunoassay. *Journal of Food Protection*, *59*, 295–298.

Gamble, H. R. (1997). Trichinellosis. In Office International des Epizooties (OIE) (Ed.), *Manual of standards for diagnostic tests and vaccines* (pp. 477–480). Paris: OIE.

Gerhardt, W., & Keller, H. (1986). Evaluation of test data from clinical studies. *Scandinavian Journal of Clinical Laboratory Investigations*, *181*, 1–74.

Greiner, M., Böhning, D., & Dahms, S. (1997). Meta-analytic review of ELISA test for the diagnosis of human and porcine trichinellosis: Which factors are involved in diagnostic accuracy? In E. A. Goodall & M. V. Thrusfield (Eds.), *Proceedings of a meeting held at the University College, Chester, 1997* (pp. 12–22). Chester: Society for Veterinary Epidemiology and Preventive Medicine.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, *117*, 167–178.

Hurblut III, T. A., Littenberg, B., & Diagnostic Technology Assessment Consortium. (1991). The diagnostic accuracy of rapid dipstick tests to predict urinary tract infection. *American Journal of Clinical Pathology*, *96*, 582–588.

Irwig, L., Macaskill, P., Glasziou, P., & Fahey, M. (1995). Meta-analytic methods for diagnostic test accuracy. *Journal of Clinical Epidemiology*, *48*, 119–130.

Irwig, L., Tosteson, A. N. A., Gatsonis, C., Lau, J., Colditz, G., Chalmers, T. C., & Mosteller, F. (1994). Guidelines for meta-analysis evaluating diagnostic tests. *Annals of Internal Medicine*, *120*, 667–676.

Knottnerus, J. A., & Leffers, P. (1992). The influence of referral patterns on the characteristics of diagnostic tests. *Journal of Clinical Epidemiology*, *45*, 1143–1154.

Lind, P., Eriksen, L., Henriksen, S. A., Homan, W. L., van Knapen, F., Nansen, P., & Stahl Skov, P. (1991). Diagnostic tests for trichinella spiralis infection in pigs. a comparative study of ELISA for specific antibody and histamine release from blood cells in experimental infections. *Veterinary Parasitology*, *39*, 241–252.

Ljungstrom, I. (1983). Immunodiagnosis in man. In W. C. Campbell (Ed.), *Trichinella and Trichinosis* (pp. 403–424). New York: Plenum Press.

Mahannop, P., Chaicumpa, W., Setasuban, P., Morakote, N., & Tapchaisri, P. (1992). Immunodiagnosis of human trichinosis using excretory-scretory (ES) antigen. *Journal of Helminthology*, *66*, 297–304.

Mahannop, P., Setasuban, P., Morakote, N., Tapchaisri, P., & Chaicumpa, W. (1995). Immunodiagnosis of human trichinosis and identification of specific antigen for Trichinella spiralis. *International Journal of Parasitology*, *25*, 87–94.

Mantha, S., Roizen, M. F., Barnard, J., Thisted, R. A., Ellis, J. E., & Foss, J. (1994). Relative effectiveness of four preoperative tests for predicting adverse cardiac outcomes after vascular surgery: A meta-analysis. *Anesthesia & Analgesia*, *79*, 422–433.

Miettinen, O. S., & Caro, J. J. (1994). Foundations of medical diagnosis – What actually are the parameters involved in Bayes theorem. *Statistics in Medicine*, *13*, 201–209.

Morakote, N., Khamboonruang, C., Siriprasert, V., Suphawitayanukul, S., Marcanantachoti, S., & Thamasonti, W. (1991). The value of enzyme-linked immunosorbent assay (ELISA) for diagnosis of human trichinosis. *Tropical Medicine and Parasitology*, *42*, 172–174.

Morakote, N., Sukhavat, K., Siriprasert, V., Suphawitayanukul, S., & Thamasonthi, W. (1992). Persistence of IgG, IgM, and IgE antibodies in human trichinosis. *Tropical Medicine and Parasitology*, *43*, 167–169.

Moses, L. E., Shapiro, D., & Littenberg, B. (1993). Combining independent studies of a diagnosis test into a summary ROC curve – Data analytic approaches and some additional considerations. *Statistics in Medicine*, *12*, 1293–1316.

Mulrow, C. D., Linn, W. D., Gaul, M. K., & Pugh, J. A. (1989). Assessing quality of a diagnostic test evaluation. *Journal of General Internal Medicine*, *4*, 288–295.

Nöckler, K., Voigt, W. P., Protz, D., Miko, A., & Ziedler, K. (1995). Indirect ELISA for the diagnosis of trichinellosis in living pigs. *Berliner und Münchener Tierärztliche Wochenschrift*, *108*, 167–174.

Petitti, D. B. (1994). *Meta-analysis, decision-analysis, and cost-effectiveness analysis: Methods for quantitative synthesis in medicine.* Oxford: Oxford University Press.

Serrano, F., Perez, E., Reina, D., & Navarrete, I. (1992). Trichinella strain, pig race and other parasitic infections as factors in the reliability of ELISA for the detection of swine trichinellosis. *Parasitology*, *105*, 111–115.

StataCorp. (1997). *Stata statistical software (version 5.0) [Computer software].* College Station, TX: Stata Corporation.

Statistics and Epidemiology Research Corporation (SERC). (1988). *Epidemiological graphics, estimation, and testing package (EGRET). Analysis module (PECAN; version 25.6) [Computer software].* Washington, DC.

Wright, P. F., Nilsson, E., van Rooij, E. M. A., Lelenta, M., & Jeggo, M. H. (1993). Standardisation and validation of enzyme-linked immunosorbent assay techniques for the detection of antibody in infectious disease diagnosis. *Revues Scientifiques et Techniques des Office Internationale des Epizooties*, *12*, 435–450.

Acknowledgments

We thank Dr. K. Nöckler, Dr. W. P. Voigt (Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin, BgVV, Berlin, FRG) and Dr. K. Wacker (Institut für epidemiologische Diagnostik, Wusterhausen, FRG) for cross-checking and supplementation of our literature retrieval. We also thank Dr. N.J. Cox (University of Durham) for making the chi-square test available.