MIXED EFFECTS LOGISTIC REGRESSION MODEL FOR CROSS SECTIONAL BINARY RESPONSE DATA: SEROPOSITIVITY AND RISK FACTORS ASSOCIATED WITH WITHIN-FLOCK TRANSMISSION OF *Leptospira interrogans* ON TRANSHUMANT FARMING SYSTEMS IN MEXICO

Gabriel Arteaga Troncoso, DVM., MSc., Ph D.
National Institute of Perinatology
Mexico City, Mexico

Background

Some reports emphasize the risk of zoonotic diseases and the high degree of prevalence of asymptomatic animals with *Leptospira interrogans*. This report sought to evaluate the prevalence of antibodies to certain serovars of *L. interrogans*, and to describe the association between seropositivity and risk factors associated with within-flock transmission in a mountainous region of Mexico.

Overview

✓The purpose of this presentation is to illustrate clearly the application of the methodological techniques, using source of data that were collected from Sheep flocks (Primary Sampling Units) and ewes (Second Sampling Units) in a mountainous region of Estado de Mexico, Mexico to approximate variance estimates or standard errors by Stata (version 13.0).

✓ The ovine census in 2000, provided by the Local Sheep-Farmers Association, recorded 3762 ewes and 3818 rams and lambs in 75 flocks. The ewes were mainly Hampshire breed or a cross with Suffolk and Pelibuey.

Sampling method

✓Sampling is concerned with the selection of a subset of individuals from within a statistical population to estimate characteristics of the whole population.

✓The sampling process comprises different stages:

Defining the population of concern

Specifying a sampling frame, a set of items or events possible to measure

Specifying a sampling method for selecting items or events from the frame

Determining the sample size

Implementing the sampling plan

Sampling and data collecting

Data which can be selected

Assumptions in logistic regression

✓As with linear regression, two important assumptions are independence and linearity.

$$Y_i\{^1_0 P(Y_I = 1) = p_i = 1 - p(Y_i = 0)$$

Independence: If animals are maintained in groups or, if multiple measurements are being made on the same individual, this assumption has probably been violated.

Linearity: Any predictor that is measured on a continuous scale is assumed to have a linear (straight-line) relationship with the outcome.

Sample Design

✓A cross-sectional study was carried out to enroll a random sample of unvaccinated ewes from November 2008 until March 2010. Stratified random sampling with proportional allocation was the sampling scheme utilized.

✓Flock size was the variable upon which stratification was based, and the flock-size strata were (A) to be <50 animals; (B) 51-140 animals, and (C) to be >141 animals. The number of flocks sampled considered a 27.8% within-flock frequency for the stratum A, 39.4% for the stratum B, and 44.4% for the stratum C, respectively**.**

Sample Design

✓Assuming a 95% level of confidence and setting error limits of 5%, approximately 10 per cent of the animals (or all ewes in flock <10 animals) were randomly sampled using a random-number calculator in each flock.

✓For providing accurate estimates, design effect (DEFF) was used to determine the difference of variances between the sample design actually used to obtain the data and a simple random sample of animals. Thirty-five flocks included in the sample were distributed uniformly throughout the area being studied, and blood samples were collected from 367 animals in selected flocks.

Source of data (1)

✓The aim of the study and its confidentiality; and, a letter of invitation along with a request for the sampling schedules.

✓Interview form which recorded information on each animal and flock management data.

✓This study was performed in strict accordance with the recommendations in the Guide Technical Specifications for the Production, Care and Use of Laboratory Animals (NOM-041-ZOO-1995). The protocol was approved by the Committee Institutional of Research and Advanced Studies in Health Animal Center at the UAEM, Toluca, Mexico for which protocol number was 2230/2006U.

Survey of leptospirosis in ovine transhumant farming systems 2008-2010

This form lists a number of questions about leptospirosis and some risk factors which be associated with within-flock transmission. We place a mark only one in the appropriate box.

1. SHEEP FLOCK DATA:

Owner´s name:_____

Municipality: □Capulhuac □Chapultepec □Santiago-Tianguistenco □Xalatlaco

Sheep production Type: □Reproduction □Meat only

Did you make additions to the sheep flock in the last 6 months?: □No □Yes

Place of lambing: □Pasture lambing □Shed lambing

Grazing time: □Permanent □Occasional

What type of pasture does sheep eat when they are mobilized from one place to another?

□Alpine herbage □Stalks of maize □Stalks of oats

Supply of water: □Tap water □lake □Irrigation canals

2. SHEEP YARD DESIGN AND CONSTRUCTION:

Site selection: □Valley □Intermountain □Mountain

Holding area (m^2): _____ Number of animals in paddock: _____

Number of animals housed at night: _____

Materials of building design:

□Brick, timber and steel □Building design with easily removable materials

Materials of bedding:

□Straw or hay □Dried corn stalks □Wood chips and leaves □Pine shavings □Sand

3. HEALTH AND MANAGEMENT:

Do you segregate those animals by production stage? □No □Yes

Frequency of cleaning of lambing paddock: □Never □Twice a week

Cleaning of bedding: □No □Yes

Disposition of excrement: □Spread as fertilizer □Accumulation of excrement in a place

Removal of aborted fetuses and fetal membranes (fomites):

□In meadows □Collection for trash

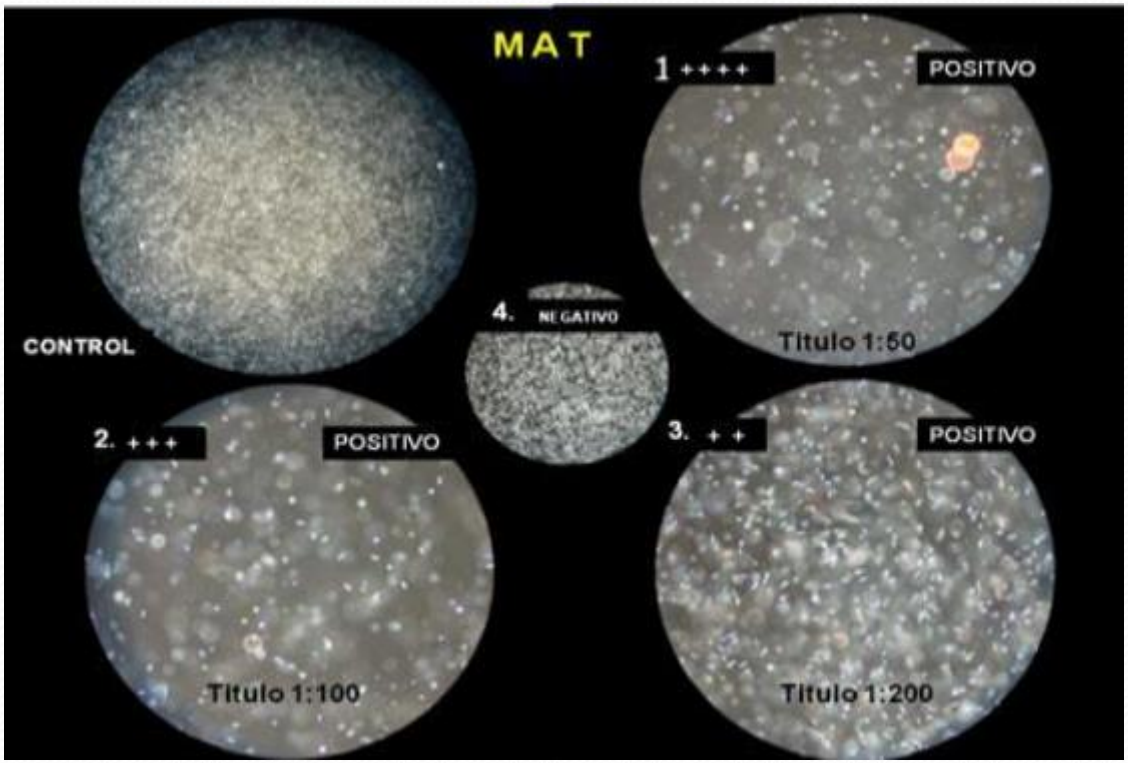
Source of data (2)

✓ Antibodies were detected in sera by Microscopic Agglutination Test (MAT). The antigens used were live cultures of reference strains *Leptospira interrogans* serovars Bratislava, Pyrogenes, Grippotyphosa, Pomona, Wolffi, Tarassovi; and, the field strains Icterohaemorrhagiae, Hardjo, and Canicola (Restriction endonuclease analysis of DNA).

✓The endpoint reading of the microagglutination reaction was reported as the serum dilution at which 50% of the leptospires were agglutinated by direct observation using dark-field microscopy. Sen: 98.2% and Sp: 96.4%

Source of data (2)

Antibodies were detected in sera by Microscopic Agglutination Test.



The minimum serum dilution was 1:50, and titers ≥1:100 were considered as positive samples.

Statistical analysis

Overall prevalence *L. interrogans* antibodies and 95% confidence interval (CI):

$$\hat{P} = \sum_{h=1}^{L} \frac{Nh}{N} \hat{P}_h$$

Where: L: The number of strata in the population; N: The number of observations in the population; Nh: The number of observations in stratum h of the population; Ph: The true proportion in stratum h of the population; nh: The number of observations in stratum h of the sample; and, Wh: The sampling fraction, Nh/N.

The variance for the estimated population total is given by:

$$\hat{V}ar\left[\hat{P}\right] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{\hat{P}_h\left(1 - \hat{P}_h\right)}{nh - 1}$$

Statistical analysis

✓Univariate odds ratios (OR) with 95% CIs were estimated for selected factors that could be relevant for L. interrogans seropositivity.

✓Factors related to the response variable were identified during the reduction process; factors with a P value <0.25, estimated by the Wald test, were included in the entry model.

✓A multilevel mixed-effects logistic regression (MMELR) was used to model the seropositivity of Leptospira and possible risk factors associated with within-flock transmission.

Multilevel mixed-effects logistic regression (MMELR)

✓The ovine population was considered to have a two-level hierarchical structure, with lower level units at level 2 (animals), nested within the groups at level 1 (flocks).

✓The percentage variance explained by the higher-level hierarchy was estimated by the variance partition coefficient.

✓To control the flock as a random effect on the response variable (seropositivity) in the absence of other explanatory variables, the MMELR was adjusted by considering a variance component of zero according to the likelihood ratio test. The random effects assumption is that the individual specific effects are uncorrelated with the independent variables of the flocks.

Multilevel mixed-effects logistic regression (MMELR)-2

✓The procedure for selecting variables was similar to a stepwise elimination of covariates.

✓The model was built and applied in four steps: (1) The variable ‘municipality’ was introduced as the first variable of integration; (2) A new variable in the model, ‘deviance’, was assessed by comparing the current model with the previous one. The additional variable was maintained if the P value of the Wald test was <0.10 based on the Schwarz Bayesian (BIC) and Akaike (AIC) information criteria. (3) If any variable in the previous model did not show statistical significance in the presence of a new factor, it was removed from the new model. (4) The procedure was repeated until none of the variables entered or left the final model.

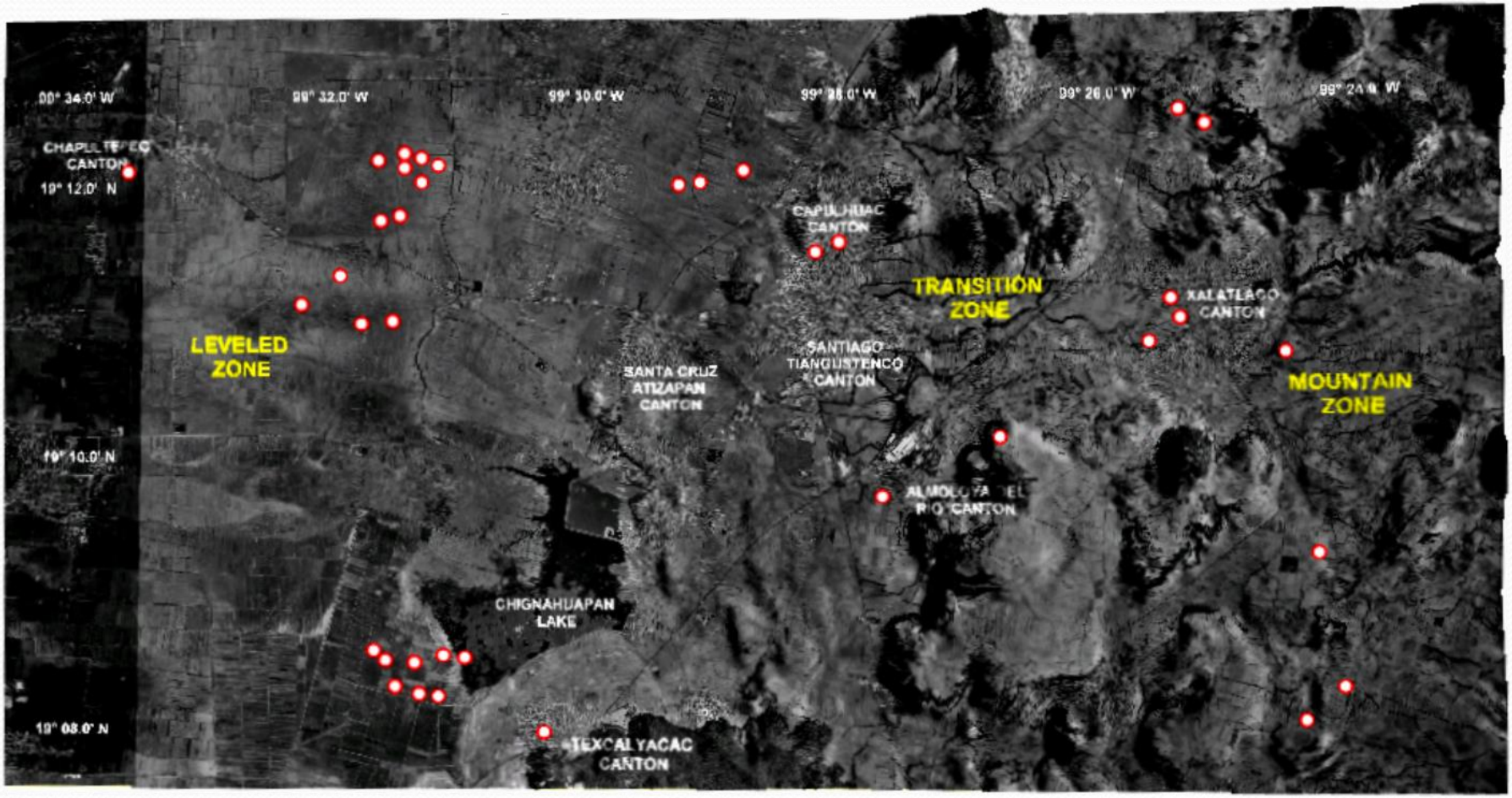
Multilevel mixed-effects logistic regression (MMELR)-3

✓Variables that exhibited multicollinearity were excluded from the model.

✓The confounding effects of the variables that were not included in the final model were evaluated by successively replacing each variable in the model and assessing the percentage change in the odds ratio of the factors retained.

✓A variable was considered as a confounder if there was a change greater than 20% in the estimated odds ratio.

Spacial-time flock location



Results

Overall seroprevalence of Leptospira was 54.5% (95% CI 48.3-60.7; DEFF 1.36)

The sampling and probability weight were established to expand the sample to the population level represented by the sample.

In our two-stage design, the probability weight was calculated as the product of two probabilities; the inverse of the sampling fraction (number of elements in the population and number of animals in the sample) for the first stage was multiplied by the inverse of the sampling fraction for the second stage.

Ewes were selected with a probability of 0.04, and the animal sampled represented 25 animals of the entire population.

Prevalence of agglutinins against *Leptospira interrogans* serovars

Serovar	Prevalence (95% CI)	DEFF
Icterohaemorragiae	54.5 (48.3-60.7)	1.36
Bratislava	40 (33.5-46.6)	1.57
Pyrogenes	3.3 (0.9-5.5)	1.45
Grippotyphosa	3.5 (1.9-5.2)	0.7
Canicola	19.1 (14.4-23.8)	1.26
Pomona	4.9 (2.6-7.2)	0.98
Tarassovi	15.8 (11.7-19.9)	1.1
Hardjo	2.2 (0.8-3.5)	0.74
Portland vere	3.8 (1.6-6)	1.18

CI, Confidence interval.
DEFF, Design effect.

Leptospira interrogans antibodies (%) in ewes´ population and univariate analysis of factors associated with seropositivity.

Variable	No. (%) of ewes		OR (95% CI)	P value
	Seropositive (n=79)	Seronegative (n=288)		
Municipality[&]				**0.04**
Chapultepec	9 (50)	9 (50)	1.1 (0.3-4.7)	
Santiago-Tianguistenco	18 (45)	22 (55)	0.7 (0.2-1.7)	
Capulhuac	35 (46.1)	41 (53.9)	0.4 (0.2-0.8)	
Xalatlaco	138 (59.2)	95 (40.8)	2.2 (1.3-3.9)	
Production-type				**0.13**
Meat only	33 (46.5)	38 (53.5)	Reference	
Reproduction	167 (56.4)	129 (43.6)	0.7 (0.4-1.1)	
Grazing time				**0.04**
Permanent	190 (56)	149 (44)	Reference	
Occasional	10 (35.7)	18 (64.3)	2.3 (1.03-5.1)	
Supply of water				**0.11**
Tap water	121 (58.2)	87 (41.8)	Reference	
Fresh drinking water from lake	79 (49.7)	80 (50.3)	1.4 (0.9-2.1)	
Place of lambing				**0.05**
Pasture lambing	27 (69.2)	12 (30.8)	Reference	
Shed lambing	173 (52.7)	155 (47.3)	2 (1-4.1)	

Leptospira interrogans antibodies (%) in ewes´ population and univariate analysis of factors associated with seropositivity.

Variable	No. (%) of ewes		OR (95% CI)	P value
	Seropositive (n=79)	Seronegative (n=288)		
Number of ewes in paddocks[&]				**0.03**
<5 animals	116 (57.1)	87 (42.9)	1.7 (0.4-7.3)	
6 to 15 animals	36 (66.7)	18 (33.3)	1.8 (0.9-3.4)	
>16 animals	32 (71.1)	13 (28.9)	2.2 (1.1-4.5)	
Building design of paddocks				**0.05**
Brick, timber and steel	144 (58.1)	104 (41.9)	Reference	
With easily removable materials	56 (47.1)	63 (52.9)	1.6 (1-2.4)	
Number of ewes housed at night[&]				**0.09**
<10 animals	51 (52.6)	46 (47.4)	0.9 (0.6-1.4)	
11 to15 animals	100 (50.8)	97 (49.2)	1.1 (0.7-1.8)	
>16 animals	49 (67.1)	24 (32.9)	1.9 (1.1-3.3)	
Cleaning of bed				**0.02**
No	181 (52.8)	162 (47.2)	Reference	
Yes	19 (79.2)	5 (20.8)	0.3 (0.1-0.8)	
Frequency of cleaning of lambing paddock Never	64 (47.4)	71 (52.6)	Reference	**0.04**
Twice a week	136 (58.6)	96 (41.4)	0.6 (0.4-0.9)	

Leptospira interrogans antibodies (%) in ewes´ population and univariate analysis of factors associated with seropositivity.

Variable	No. (%) of ewes		OR (95% CI)	P value
	Seropositive (n=79)	Seronegative (n=288)		
Disposition of excrement				**0.09**
Spread as fertilizer	**162 (52.6)**	**146 (47.4)**	**Reference**	
Accumulation of excrement in a place	**38 (64.4)**	**21 (35.6)**	**0.6 (0.3-1.1)**	
Disposition of fomites				**0.03**
In meadows	**136 (79.1)**	**36 (20.9)**	**Reference**	
Collection for trash	**64 (32.8)**	**131 (67.2)**	**1.3 (1.03-1.7)**	

*Univariate multilevel mixed-effects logistic regression with leptospiral seroreactivity as the outcome, and flock as a random effect.
[&]Dummy explanatory variables.
OR, Odds ratio; CI, confidence interval.

The random effects assumption is that the individual specific effects are uncorrelated with the independent variables within flocks.

Adjusted model:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + u_j$$

$$u_j \approx N\left(0, \sigma_u^2\right)$$

Πij positive ewe (i) in flock (j) and u (j) random effect with flock (j)
Command xtmelogit from Stata

Fixed effect

	Coefficient	SE	Z	P>[Z]	90%CI
β_0	0.1563239	0.1323877	1.18	0.0852	0.1031512-0.4157989

Random effect

	Coefficient	SE	X^2	P>[X^2]	90%CI
σ_u^2	0.1797179	0.1617517	2.4	0.0608	0.041-0.789

Schwarz and Akaike Bayesian Information Criterion for 14 mixed-effect logistic regression models of the serologic prevalence of *Leptospira interrogans* in 367 ewes from 35 flocks.

Mixed-effect logistic regression models[&]	BIC[ŧ]	AIC[ŧŧ]	p-value
MUNIC*	513.3	505.5	0.04
MUNIC*+ PROD	516.3	504.6	0.03
MUNIC*+ PROD+GRAZING	304	290.6	0.06
MUNIC*+ PROD+ PADDOCK	304	290.6	0.06
MUNIC**+ PROD+ LAMBING	302	288.6	0.03
MUNIC*+ LAMBING+ WATER	303.2	289.7	0.05
MUNIC*+ LAMBING+ ANICOR	303	289.2	0.04
MUNIC*+ LAMBING+ ENCIERRO	302.6	289.1	0.04
MUNIC*+ LAMBING+ CLEANING	302.9	289.6	0.05
MUNIC**+ LAMBING+ FLAMBING	302.4	289	0.03
MUNIC**+ LAMBING+ EXCREMENT	301	287.5	0.02
MUNIC**+EXCREMENT+ FOMITES*	301.1	287.7	0.02
MUNIC*+ EXCREMENT	298.9	288.8	0.03
MUNIC**+ FOMITES*	296.3	286.2	0.01

* <0.05; ** <0.01 p-value

[&]With the flock as the random effect.

[ŧ]Schwarz Bayesian Information Criterion

[ŧŧ]Akaike Information Criterion

MUNIC: Municipalities; PROD: Sheep production; GRAZING: Grazing time; PADDOCK: Design of lambing paddock; LAMBING: Place of lambing; WATER: Supply of water; ANICOR: Ewes in paddock; ENCIERRO: Number of ewes housed at night; CLEANING: Cleaning of bed; FLAMBING: Frequency of cleaning of lambing paddock; EXCREMENT: Disposition of excrement; FOMITES: Disposition of placentas and fetus remains.

✓The final MMELR appeared to fit the data adequately (Pearson goodness-of-fit test statistic = 2.41, *P* = 0.49). A significant (deviance) statistic would show that the MMELR is inappropriate for the data.

✓The area under the ROC curve (0.64) was significantly different from 0.5, since the *P* value was 0.000, indicating that the MMELR classified the group significantly better than chance.

✓Our MMELR had an acceptable predictive ability; 213 (58.04%) of the ewes sampled were correctly classified by sensitivity (80%) and specificity (31.8%).

```
. estimates stats

Akaike's information criterion and Bayesian information criterion

-------------------------------------------------------------------------------
     Model |    Obs    ll(null)   ll(model)     df        AIC        BIC
-----------+-------------------------------------------------------------------
         . |    212           .  -141.4101      3     288.8203   298.8901
-------------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

```
. estat gof
Logistic model for Leptospira, goodness-of-fit test
 number of observations =      367
 number of covariate patterns =        7
Pearson chi2(3) =        1.91
Prob > chi2 =         0.5920
```

```
. lroc
Logistic model for Leptospira
number of observations =      367
area under ROC curve   =   0.64
```

Final multilevel mixed-effect logistic regression model for risk factors associated with seropositivity to *Leptospira interrogans* in ewes from transhumant farming systems in Mexico.

Variable	Std. Error	OR (95% CI)	P-value
Xalatlaco municipality	0.39	1.8 (1.2-2.7)	0.01
Accumulation of placentas and fetuses remains at a place close to lambing paddocks	0.16	2.4 (1.2-4.7)	0.02
Intercept	0.14	0.5 (0.3-0.9)	0.02

OR, Odds ratio; CI confidence interval.

Key results:

✓Overall seroprevalence to *L. interrogans* was 54.5%.

✓Accumulation of placentas and fetuses can be a significant risk factor for within-flock transmission of L. interrogans.

✓The high prevalence of *L. interrogans* antibodies supports the hypothesis that natural foci of this zoonosis are present in sheep flocks.

Acknowledgements:

✓The authors thanks Dr. V. Banda, Dr. E.A. Barragan, Ing. A. Arteaga, and C. Vidali for technical advice, and Dr. C. Bolin from Ames, Iowa, USA for classifying field strains. We are grateful to the owners of the affected properties for their cooperation and assistance with the investigations.

✓Financial support:

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

✓Declaration Conflict of interest: None

Members of team:

G. ARTEAGA-TRONCOSO[1,2*], J. M. JIMÉNEZ-ESTRADA[3], R. MONTES DE OCA-JIMENEZ[3], M. LÓPEZ-HURTADO[4,5], M. LUNA-ALVAREZ[6], L. HERNANDEZ-ANDRADE[6], A. MORENO-ALFARO[2], J. F. GALAN-HERRERA[7], F. M. GUERRA-INFANTE[4,5]

[1]Department of Immunobiochemistry, National Institute of Perinatology, Mexico City, Mexico; [2]Applied Statistics Postgraduate Program, Applied Mathematics and Systems Research Institute, UNAM, Mexico City, Mexico; [3]Faculty of Veterinary Medicine, Universidad Autonoma del Estado de Mexico, UAEM, Toluca, Mexico; [4]Department of Infectology and Immunology, National Institute of Perinatology, Mexico City, Mexico; [5]Department of Veterinary Microbiology, Escuela Superior de Ciencias Biológicas, IPN, Mexico City, Mexico; [6]CENID-Microbiology INIFAP-SAGARPA, Mexico City, Mexico; [7]Mexican Clinical Research Network for Emerging Infectious Diseases, National Institute of Medical Sciences and Nutrition, Mexico City, Mexico.