Multivariable binary logistic regression model to predict risk factors of peste des petits ruminants in goat and sheep

A.A. Sobeih¹, K.M. El-Bayoumi¹, M.S. El-Tarabany¹, A.A. Abuel-Atta² and S.A. Moawed³

¹Animal Wealth Development Department, ²Histology and Cytology Departments, College of Veterinary Medicine, Zagazig University, Zagazig, ³Animal Wealth Development Department Faculty of Veterinary Medicine, Suez Canal University, Suez, Egypt

Abstract

Peste des petits ruminants (PPR) highly contagious illness that affects domestic and wild small ruminants, causing significant economic losses. The goal of this study was to use a multivariable logistic regression model to determine risk factors for PPR. A total of 113 healthy non-vaccinated goats and sheep (63 goats and 50 sheep) more than five months (1st group from 5-12 months, 2nd group above 12 months), subject to a seroprevalence study by competitive ELISA which was used to detect antibodies against PPRV antibodies in serum sample during the period between April 2018 and March 2019. The incidence of PPR in autumn was significantly increased compared with the spring. Additionally, animals had a 4.08 more likelihood of being infected with PPR in the autumn compared with the spring season. There was a significant difference between male and female groups; The female group had 5.236 times increased odds of being infected with PPR than the male group. Moreover, the old age group had 2.771 times higher odds of being infected with PPR than the young age group. On the other hand, the test model found no evidence to support any significant differences between sheep and goat species. According to finding, PPR is common in females and mature small ruminants. Furthermore, throughout the spring season, the incidence of PPR was significantly reduced. Indeed, the current study may help plan an effective vaccination program against the PPR disease in Egypt.

Keywords: Coefficient of regression, Odds ratio, PPR, ROC curve

Correspondence: A.A. Sobeih, abdoosawd.dr@gmail.com

DOI: 10.33899/IJVS.2022.132934.2151, ©Authors, 2022, College of Veterinary Medicine, University of Mosul. This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Peste des petits ruminants (PPR), which affects domestic and wild ruminants, is a highly contagious illness that spreads rapidly irrespective of country borders. Infection is spread primarily through the secretions and excretions of infected animals (1). Goat plague, kata, syndrome of stomatitis-pneumonenteritis, and ovine rinderpest are all synonyms for PPR (2). Peste des petits ruminants were originally defined in 1942 in the West African Republic, of Côte-d'Ivoire (1). The disease occurs due to Peste des petits ruminant’s virus (PPRV) which causes acute, highly contagious, and economically critical diseases in domestic, wild small ruminants, and camels (1). PPRV, like other paramyxoviruses, is an enveloped virion with a single-stranded RNA genome with opposing polarity and a diameter of 400-500 nm. It has the second-longest genome among morbilliviruses, with a genomic length of 15,948 nucleotides (3). Furthermore, the PPRV genome is divided into six transcriptional units, each encoding for at least six structural and two nonstructural proteins (4). There is historical information available that reveals the prevalence of PPRV considerably earlier. The shortage of in advance diagnoses was possible because of the clinical similarity to ruminant plague (RP) and the use of diagnostic tests, which could not differentiate among them. Therefore, the PPR defined in
small ruminants probably became a PPRV infection (5). The
disease has been assured in most geographical locations in
West Africa, including Nigeria, Senegal, Togo, and Benin.
At the same time, in 1982, the disease was recognized in
Sudan, an eastern African country (6). Goats/sheep pox
(SGPX), peste des petits ruminants (PPR), contagious
ecthyma (CPD or ORF), Rift Valley fever (RVF), bluetongue
(BT), Foot and mouth (FMD), Nairobi sheep (NSD), and
Border (BD) diseases are the most common viral infections
of sheep and goats in Africa. (7). Most of the reported viral
diseases of small ruminants in Egypt result in significant
losses in sheep and goat husbandry (8). PPR symptoms are
analogous to rinderpest in cattle, including an abrupt rise in
body temperature to 40-41.3°C. Some infected animals have
a cough, pneumonia, oral necrosis, mucopurulent nasal and
ocular discharges, and copious catarrhal conjunctivitis with
matting of the eyelids. Furthermore, the severity of the illness
varies depending on the sheep's breed and initial immune
condition, geographical location, season, and whether the
infection is acute or chronic. Fever, along with either
diarrhea or indications of mouth discomfort, is enough to rule
out the diagnosis. Young animals have higher rates of
morbidity and mortality than adults. However, mature
animals frequently exhibit bronchopneumonia and abortion
(9). As in Border disease usually show high rates of
infertility, and production of underweighted-lambs in young
animal, but in mature animal show abortion, stillbirth and
lambs born weak, uncharacteristic body shapes and
undersized, if occurs between the 50th and 60th day of
gestation may lead to birth of immunotolerant lambs that will
continue shedding the virus for their whole lifetime and be
the most significant source of BDV among ruminants (10).
In this context, Sick animals that recover from infection have
an entire life of protective immunity, and there is no evidence
of carrier status. Moreover, the virus can spread in animals
with a weak form of the disease, causing disease outbreaks
during which young susceptible populations mix with those
infected and develop a mild form of the disease. Other host
parameters, such as age, sex, species, monthly temperature,
relative humidity, immune status, feeding pattern and
geographical area (11), and season, may also influence
disease progression (2). Because of a lack of epidemiological
data and various management strategies in the herds where
the problems emerge, infectious viral diseases are difficult to
control (12). As a result, the current study attempted to create
a multivariable statistical model to assess risk factors of PPR
disease.

Materials and methods

The procedures followed the ethical norms and
guidelines of the Institutional Animal Care and Use
Committee of the Faculty of Veterinary Medicine at Zagazig
University (ZU-IACUC-94/2021).

Study design and data description

A total of 113 healthy, non-vaccinated goats and sheep
(63 goats and 50 sheep), more than five months (which were
divided into young groups from 5-12 months and old group
above 12 months), were subjected to a seroprevalence study
by competitive ELISA (ID PPR competition kit) to detect
antibodies against PPRV during the period between April
2018 and March 2019. Animals were subdivided according
to species, age, gender, and season in the Al-Sharqia
governorate.

Data collection and the exploratory variables

The current study concerned 113 animals that have been
divided into two groups: the first group consists of 70 PPR
positive animals, during the second group of 43 Healthy
control animals. In this study, the possible predictor
variables were: species (63 goats and 50 sheep), age (41
adults (>12 months) and 72 young (5-12 months), gender (51
male and 62 female), and season (winter, spring, summer,
and autumn). All predictor variables were categorical and
coded as follow: species (goat = 1, sheep = 2, sheep was
considered as reference category), age (adult = 1, young = 2,
young age was considered as reference category), gender
(male = 1, female = 2, female was considered as a reference
category), season (winter = 1, spring = 2, summer = 3,
autumn = 4, the autumn was considered as reference
category). In addition, the outcome variable was coded as
ONE for positive PPR animals and ZERO for unrelated
healthful controls.

Data management and statistical analysis

Using the SPSS statistical software package version 22
(13), all statistical methods were carried out. To choose the
variables, the entry technique was used first, followed by
multivariable binary logistic regression models. The
following is a summary of the multivariable binary logistic
regression model: 

\[ \log \left( \frac{P}{1-P} \right) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_kX_k \]

Where: \((P/1-P)\) is the odds, \(P\) is the probability of PPR, \(\beta_0\) is
the Y-intercept of the model (y value when all \(X\) equal zero),
\(\beta_i\) is the first regression coefficient of \(X_i\), also known as the
first explanatory variable in the model, and \(\beta_k\) is the last
regression coefficient of \(X_k\), also known as the last
explanatory variable in the model. The coefficient of
regression (\(\beta\)), the coefficient's standard error (SE), the Wald
value, and the related risk factor's odds ratio (OR) at a 95%
confidence interval (CI), as well as the P-values, were used
to illustrate the results. Wald value is a statistical measure
that determines whether or not the explanatory variable is
statistically significant in differentiating between the two
categories in each binary logistic comparison depending on
the relation between Wald’s statistics and the standard error
of the logistic coefficient of predictors. It is analogous to the
t-test in linear regression and distributed as a Chi-square
distribution (14): \( W_k = (\hat{\beta}_k / \text{St error} \ (\hat{\beta}_k))^2 \). Where \( \hat{\beta}_k \) is the k-th estimated regression coefficient, each Wald statistic is compared to a chi-square distribution with one degree of freedom, and P-value less than 0.05 indicates that x has a significant effect on y. Independent variables in logistic regression models are commonly represented as odds ratios (ORs), which reflect the magnitude of the independent variable's ability to contribute to the outcome. ORs are defined as the probabilities of the outcome occurring (p) vs. not occurring (1-p) for each independent variable. It is also the antilogarithm of the coefficient for the independent variable and is equal to exp(\( \beta \)) (exponentiation of logit coefficients). This equation's odds ratio is (15): \( \text{OR} = e^\beta \).

Where: e = 2.718 and \( \beta \) is the regression coefficient. Moreover, imply that a one-unit change in the independent variable multiplies the outcome odds by the amount contained in \( e^\beta \). The absence of multicollinearity between the explanatory variables was ensured before the multivariable analysis. Multicollinearity is a statistical phenomenon in which two or more predictor variables in a regression model are highly correlated (r), with r values ranging from -1 to 1 (16). Multicollinearity exists for every (r) greater than 0.7, implying highly correlated independent variables. This issue can be investigated by looking at the tolerance and variance inflation factors (VIF). Because we have categorical predictors, we performed spearman’s rank correlations and the Chi-square test to determine multicollinearity. Multicollinearity is also considered when the VIF value reaches 10 (17) and the tolerance value equals or less than 0.1 (18). The overall goodness of fit of the final selected model was tested using the Receiver Operating Characteristic curve (ROC curve) via the area under the curve (AUC). The model is considered fit to the current dataset if the AUC is more than 0.5 and has a significant probability value (\( P<0.05 \)).

Results

To begin, test assumptions were made, and the results of the correlation matrix for independent variables were calculated. The spearman’s rank correlations (r) were less than 0.20, indicating no significant correlations between the variables. In addition, the variance inflation factor (VIF) and tolerance values were calculated (Table 1). There is no multicollinearity between independent variables because none of the variables in the model have a VIF value over 2.0 or a tolerance value close to 0.95. The overall predicted percentage of correctly classified cases was 75.2 % (Table 2). The outputs of the Omnibus test and Hosmer and Lemshow goodness-of-fit test (Table 3). There is a significant difference between the Log-likelihoods (specifically the -2LL). The -2LL value equal 126.44 and the chi-square value was highly significant \( (\chi^2 = 23.7, P = 0.001) \). Regarding the Hosmer and Lemshow goodness-of-fit test, the chi-square value was non-significant \( (\chi^2 = 7.009, P = 0.536) \). The area under the curve (ROC curve) was also used to assess the model with used predictors (AUC = 0.77, \( P < 0.001 \)), which is significantly considerable acceptable value for discrimination ability between diseased and non-diseased animals (Figure 1).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Tolerance</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>0.998</td>
<td>1.002</td>
</tr>
<tr>
<td>Age</td>
<td>0.972</td>
<td>1.029</td>
</tr>
<tr>
<td>Season</td>
<td>0.986</td>
<td>1.014</td>
</tr>
</tbody>
</table>

Table 2: Validity and predictive ability of the incorporated binary logistic regression model

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Value</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2log likelihood</td>
<td>126.437</td>
<td></td>
</tr>
<tr>
<td>Nagelkerke’s R²</td>
<td>0.257</td>
<td></td>
</tr>
<tr>
<td>Hosmer and Lemshow Chi square</td>
<td>7.009</td>
<td>0.536</td>
</tr>
<tr>
<td>Omnibus Test Chi square</td>
<td>23.7</td>
<td>0.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>62.8%</td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>82.9%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>75.2%</td>
<td></td>
</tr>
<tr>
<td>Area under curve</td>
<td>AUC</td>
<td>0.77 &lt;0.001</td>
</tr>
</tbody>
</table>

Figure 1: The ROC curve using the estimated probability denoted by the multivariable binary logistic regression model (AUC = 0.77, \( P < 0.001 \)).
The associations between risk factors and the outcome variable (incidence of PPR) are summarized in Table 4. For the season variable, autumn was considered as the reference category. The summer and winter seasons showed a non-significant effect on the incidence of PPR when compared with the autumn season (OR = 1.073 and 1.398, respectively). Meanwhile, the incidence of PPR in the spring season was significantly reduced compared with the autumn season (OR = 0.245). Additionally, animals had a 4.08 more likelihood of being infected with PPR in the autumn season when compared with the spring season.

Table 4: Multivariable binary logistic regression showing the association between risk factors and PPR disease

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>B</th>
<th>Odds ratio</th>
<th>95% C.I. (OR)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species (goat), ref. = sheep</td>
<td>0.189</td>
<td>1.208</td>
<td>0.52-2.83</td>
<td>0.664</td>
</tr>
<tr>
<td>Sex (male), ref. = female</td>
<td>-1.655</td>
<td>0.191</td>
<td>0.08-0.48</td>
<td>0.001**</td>
</tr>
<tr>
<td>Age (old), ref. = young</td>
<td>1.019</td>
<td>2.771</td>
<td>1.07-7.19</td>
<td>0.036*</td>
</tr>
<tr>
<td>Season autumn</td>
<td></td>
<td></td>
<td></td>
<td>0.077</td>
</tr>
<tr>
<td>Winter</td>
<td>0.335</td>
<td>1.398</td>
<td>0.42-4.61</td>
<td>0.582</td>
</tr>
<tr>
<td>Spring</td>
<td>-1.406</td>
<td>0.245</td>
<td>0.07-0.90</td>
<td>0.034*</td>
</tr>
<tr>
<td>Summer</td>
<td>0.071</td>
<td>1.073</td>
<td>0.37-3.16</td>
<td>0.898</td>
</tr>
</tbody>
</table>

*Significant at P<0.05. **Significant at P<0.01.

Discussion

The evaluation criteria of the model were reported as -2log likelihood with a chi-square value which reflects how much information remains unexplained after the model has been fitted and utilized to check the overall relationship between an explanatory variable and the outcome by comparing the likelihood of getting data when the parameter is zero to the likelihood of obtaining data evaluated at the maximum likelihood estimate of the parameter (19). The Omnibus test of model coefficients determines whether the new model (including explanatory variables) is superior to the baseline model. It applies chi-square tests to determine whether there is a statistically significant difference between the Log-likelihood value (specifically the -2LL). Its chi-square value is highly significant, indicating that the current model is significantly better at explaining more of the variance in the outcome. The other evaluation criteria are Hosmer and Lemshow goodness-of-fit tests, where small values of $\chi^2$ with significant P-values indicate an excellent fit to the data, while large $\chi^2$ values with P-values less than 0.05 indicate a poor fit (20). Meanwhile, the low Nagelkerke’s R² value 0.257 may indicate that the model’s efficiency is questionable as the logistic model explains only a 25.7% probability of the change in outcome.

There is no multicollinearity between independent variables because none of the variables in the model have a VIF value greater than 2.0 or a tolerance value less than 0.90 (18,21). A tolerance value close to 1 indicates little multicollinearity between predictors, whereas a value close to zero indicates much multicollinearity. A VIF value that is equal to 1 means no multicollinearity, while a VIF value >1 means moderate multicollinearity, and VIF between 5 and 10 means high multicollinearity, and if VIF value > 10 indicates serious multicollinearity (17).

In the current study, females were more likely affected by PPR, with an odds ratio than males. Consistent with these findings, the infection rate of PPR in females 74.7% was more significant than in males 54.1% (22). Because the sex of the animals did not affect the development of PPRV antibodies, this significance has no biological plausibility. This observation could be explained by the fact that producers of small ruminants keep more females for breeding purposes. As a result, females are more likely than males to be exposed to PPRV throughout their lives (23). On the other hand, Males are the most affected by PPR, which might be due to genetic variation of goats in Raj Shahi, Bangladesh (24).

For the sex variable, old animals more than 12 months were more likely affected by PPR, with an odds ratio than young animals 5-12 months. This agrees with the report’s findings that showed a higher prevalence of PPR in animals older than two years of age (25). On the contrary, the prevalence of PPR was higher in young animals 61.8% than in adult animals 49.2% (26). It is possible. Sometimes, young goats require additional nutritional supplements to reach sexual maturity and gain body weight. As a result, they suffer from long-term malnutrition, making them susceptible to disease. Increased susceptibility of young goats may be due to malnutrition, poor immunity, and a poor management.
system. Saadullah (27), in disagreement with our result, showed that the age of animals was analyzed into three categories, adult (>1 year) and young (between 4 to 12 months). Sucklers (between 1 to 3 months) were 10.15%, 31.06%, and 13.14%, respectively. This is due to the loss of maternal (colostral) immunity in young animals beyond four months and the absence of PPR antibodies in the serum of older animals who have never been exposed to or vaccinated against the disease (24).

In the current trial, goats exhibited a higher incidence of PPR 56.5% than sheep 35.1%. However, the differences were non-significant. In another context, higher seropositivity of PPRV in goats 34.5% compared to sheep 11.2% (28). In northeastern Egypt, the prevalence of PPR was higher in goats 69.3% (104/150) than in sheep 48.5% (197/450) (26). The differences in PPRV seropositivity depend on species, sex, age, season, and geographical location. Although both goats and sheep are susceptible to infection and may show disease, they are not always affected concurrently; for example, in Africa, PPR is most commonly seen in goats, whereas sheep are usually among the most in western and south Asia visible victims (29).

The incidence of PPR in the spring season was significantly reduced compared with the autumn season. Additionally, animals had more likelihood of being infected with PPR in the autumn season when compared with the spring season. The dry season has a greater prevalence of PPR than the wet/rainy season (30). Others have suggested that seasons with solid wind speeds support PPR aerosol transmission significantly more than seasons with low wind speeds (31). Moreover, a significant association between the PPR prevalence and the winter season was noted due to the dusty and dry wind (22).

Conclusion

It could be concluded that PRP is more prevalent in female and adult Sheep and Goat. Moreover, the disease is relatively more prevalent in goats compared with sheep. The infection rate of PPR was significantly reduced in the spring season. Uncontrolled animal movements and across borders pose a high risk for PPRV spread and warrant further studies to adopt standard prevention and control programs. Considering the current results, the present study may allow the planning of an effective vaccination program against PPR disease in Egypt. In addition, the study showed the applicability and effectiveness of the multivariable binary logistic regression to deal with veterinary datasets, denoting broad information as compared with the traditional Biostatistical methods.

Acknowledgments

Author wishes to thank Knowledge his colleges in both the Department of animal wealth development and Department of infectious Diseases for supporting and providing facilities

Conflict of interest

There is no conflict of interest.

References

الخلاصة

طاعون المجترات الصغيرة هو مرض شديد العدوى قد يؤثر على كل من المجترات الصغيرة المستأنسة والبرية، مما قد يسبب خسائر اقتصادية كبيرة. كان هدف هذه الدراسة هو استخدام نموذج الانحدار اللوجستي متعدد المتغيرات لتحديد عوامل الخطر بالنسبة لتوسيع أداء البرنامج، خضم 13 من المعايير الجانبية، 21% من الجماعات و 10% من الذكور. على ذلك، انخفض معدل الإصابة بشكل ملحوظ طوال فصل الربيع. والواقع أن الدرا

27. Sayllah M. Research and development activities and needs on small ruminants in Bangladesh. Res Devel Small Ruminant Asia. 1999:12-20. [available at]