

Repeated measures
logistic regression
health and social care
essay



Dose-response studies in arthropod research usually involve repeated measures over time on groups of organisms subjected to different doses. When data is collected over time from the same experimental unit, the observations are not expected to be independent. Highly correlated dose-response data cannot be estimated using Probit model which is the usual model for dose-response analysis. The objective of the study was to illustrate how to estimate LT50 using repeated measures logistic regression using Generalized Estimating Equations (GEE). The study used insect mortality data collected from dose-response studies on thrips. Repeated measures logistic regression using GEE was used to estimate LT50. Results of this study suggest that different extracts were significantly different from each other. Logistic regression showed that dose and the extracts had significant effect on insect mortality. The results indicated that dose 500 was best / effective. GEE for repeated measures logistic regression is a very promising appropriate and elegant method for repeated measures analysis in dose response studies for estimating LT50 and should be employed alongside other existing methods for estimating LT50. The method reaches far beyond traditional methods of statistical analysis that are not well suited to estimating LT50 in repeated measures data. Future studies may involve wider exploration of GEE techniques and further testing and refinement are needed to fully develop its promising capabilities. Keywords: Lethal Time, Repeated measures, Dose, Survival Analysis, GEE, Probit Analysis

INTRODUCTION

Dose-response studies in arthropod research usually involve repeated measures over time on groups of organisms subjected to different doses.

When data is collected over time from the same experimental unit, the observations are not expected to be independent. Highly correlated dose-response data cannot be estimated using Probit model (Finney, 1971) which is the usual model for dose-response analysis. However, researchers resort to analysis at each time point which is not very efficient especially when the interest is also in the speed of kill (LT50). In many arthropod studies the assumptions for Probit analysis such as the independence of variables and their normal distribution are hardly met yet it is being used in the analysis and estimating LT50 (Robertson and Preisler, 1992). According to Ziegler (1998), neglecting dependencies in these situations can lead to false conclusions. Repeated measures logistic regression using Generalized Estimating Equations (GEE) is a possible models for repeated dose-response data. Generalized Estimating Equations (GEE) were introduced by Liang and Zeger (1986) as an extension of Generalized Linear Model (GLM) method (McCullagh and Nelder, 1983; McCullagh and Nelder, 1989) to handle correlated discrete data. This approach accommodates dichotomous outcomes and counts for which the correlation among observations. With GEE correlated data can be modeled with output that looks similar to generalized linear models (GLMs) with independent observations. The primary difference is their ability to account for the within-subject covariance structure for the various types of response data (Ziegler et al., 1998; Zeger and Liang, 1986). The available covariance structures specify how observations within a subject or cluster are correlated with each other. Correlated data are modeled with the same link functions and linear predictor equation (systematic component) as found with independent data.

The random component of GEEs is also described by the same variance
<https://assignbuster.com/repeated-measures-logistic-regression-health-and-social-care-essay/>

functions, but now the covariance structure of the correlated measurements must also be modeled (Zuur et al., 2009). Logistic regression is a GLM method for analyzing binary outcome but ignores the correlated nature of the data. The standard errors may be incorrectly estimated and thus certain covariates may be incorrectly identified as significant predictors in a model. Since the data has a binary repeated measures response, generalized estimating equations (GEE) in a logistic regression setting is a good way to model the data. In this paper the use of repeated measures logistic regression using GEE method is considered as complementary approach to LT50 estimation to address the limitation of Probit Analysis in estimating LT50 for correlated dose response data. Repeated measures logistic regression using GEE was used because the data was binary and correlation due to time was to be taken into account.

METHODS

Repeated measures data sources

The dose response data on insect mortality used in this paper was from a laboratory experiment on Thrips project at ICIPE. A populations of 50 insects per batch restricted in a container were subjected to different kinds of insecticides (plant extracts) of different concentration levels and their responses to dose (time to death) were recorded at 6 time intervals of 12 hours in terms of the number dead.

Definition of parameters of interest

The factors of interest under study are the different treatments (plant extracts), their doses (concentration level), and time to death. Lethal time

(LT) can be defined as the period of time required for a proportion of a large group of organisms to respond after exposure to a specific dose of an injurious agent, such as a drug or radiation or pathogen at a given concentration under a defined set of conditions. Two parameters in dose response data: concentration and time. Concentration is the dose levels while time are the different time interval by which the effects of the different concentration are being monitored.

Data organization and layout

The data set for GEE was sorted by the extract and for each dose level. A variable 'IDD' which identifies the time clusters was created at it coincided with the distinct time points. The data set was created for each extract at each dose level as shown in Appendix 1. Other variables in the column were replicate, total number of insect subjects and the number of insects responding to the exposed dose levels. There were no missing values were reported.

STATISTICAL ANALYSIS

Estimating Lethal Time using GEE for Repeated Measures Logistic Regression

Logistic (logit) regression is a type of regression analysis used for predicting the outcome of a categorical dependent variable based on one or more predictor variables (McCullagh and Nelder, 1983; McCullagh and Nelder, 1989). For outcome variable y , and a set of predictor variables, X . Consider a binary response variable with a logistic transformation or logit function, then logistic regression is $\text{logit}(p) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ Where α is the intercept, β_j is the regression coefficient for each corresponding predictor variable, X_j , and ϵ is the error of the prediction. The <https://assignbuster.com/repeated-measures-logistic-regression-health-and-social-care-essay/>

logit of a probability is simply the log of the odds of the response taking the value one. The above Equation can be rewritten as The logistic regression model indirectly models the response variable based on probabilities associated with the values of μ . The logit function can take any real value, but the associated probability always lies in the required interval. The parameters of the logistic regression model (the vector of regression coefficients) are estimated by maximum likelihood. Logistic regression which is GLM was used but since it cannot account for the correlation the GLM was extended using GEE to estimate the parameters (β and α) by specifying the correlation structure (Zurr et al. 2009; Zeger and Liang, 1986; Liang and Zeger, 1986). GEE is an implementing method for non-normal data where observations are correlated and is used to estimate the parameters and α . The parameters are estimated using different correlation structures and QIC is used to choose which correlation structure is giving the least QIC using GEE and then fit it to the logistic regression model. LT50 was estimated using repeated measures logistic regression which uses GEE

Expressing the lethal time

Consider the above equation having time as an explanatory variable and the response variable as the proportion of mortality The LT50 is, by definition, the time at which equals 0.5. (Preisler and Robertson, 1989; Finney, 1971) by substituting with 0.5 in the above equation we get To account for correlation effect due to time (repeated measures) we use Generalized Estimating Equations to estimate the parameters and

Generalized Estimating Equations

To use GEE in estimating, we have three-part specification; the conditional expectation of each response, the conditional variance of each given the covariates and the covariance (correlation) matrix (Zuur et al. 2009; Liang and Zeger, 1986). Let's define the marginal regression model to be: Where is a vector of covariates, consists of the p regression parameters of interest, is the link function, and denotes the outcome for the subject and is the correlation matrix. Common choices for the link function include [identity link], [for count data], and [logit link for binary data]The regression model (score) is given byWhere is a GLM dispersion parameter - allows for over dispersion, is a diagonal matrix of variance functions, is the matrix of derivatives, is the "working" covariance matrix of, and is the working correlation matrix of. Given a mean model, , and variance structure, , ("working" covariance matrix of), the parameter estimates will be given by solvingThere are two classical ways of estimating the covariance. The model-based estimator of the covariance matrix of and the empirical or robust estimator of the covariance matrix of. Let be an "working" correlation matrix that is fully specified by the vectors of parameter. The covariance matrix of can be modeled asWhere is a diagonal matrix with as the diagonal element. If is the true correlation matrix of, then will be the true variance matrix of. The working correlation is usually unknown and must be estimated. It is estimated in the iterative fitting process by using the current value of the parameter vector. Common choices for the correlation structure within GEE include Independent, exchangeable, autoregressive (AR(1)), unstructured and M-dependent (Crowder, 1995)Unstructured correlation matrix was used to extract the variances and the covariances and <https://assignbuster.com/repeated-measures-logistic-regression-health-and-social-care-essay/>

since the data was binary it belongs to the binomial family and the link function used was the logit. Taking the estimate of Intercept as alpha and the estimate of time as the beta, the extracted variance covariance (empirical) matrix from the regression will have parameters as variance of alpha, covariance of alpha and beta, and variance of beta. These parameter estimates are then used to construct the confidence interval using delta method. Delta Method was used to estimate the variance so as to get the confidence intervals of the LT50 estimates. This is a method of finding approximations based on Taylor series expansions to the variance of functions of random variables (whose variances are known). The Delta method gives the variances of α and hence an approximate 95% confidence interval for the LT50 is given by

RESULTS

Table 2 gives the summary of the analysis for repeated measures logistic regression using GEE. The lethal time (LT50) ranged between 10.3 hrs to 52.1 hrs for extract B; 7.2 hrs to 74.2 hrs for extract C and between 10.3 hrs to 55 hrs for extract E. The LT50 values for the different concentration levels ranged between 52.1 hrs to 74.2 hrs for concentration 12.5; 16.6 hrs to 43.4 hrs for concentration 50; 12.2 hrs to 42.9 hrs for concentration 250; and between 7.2 hrs to 10.3 hrs for concentration 500. From Table 2 dose 500 was the effective dose since it has the lower LT50 value across all the extracts.

DISCUSSION

This paper reviews the application of repeated measures logistic regression using GEE approaches in estimating LT50 in repeated measures dose

<https://assignbuster.com/repeated-measures-logistic-regression-health-and-social-care-essay/>

response studies. The repeated measures logistic regression used GEE as implementing tool. Time was a correlation factor which had to be included in the analysis and therefore estimating LD50 at each time point will end up giving wrong estimates since interest is in the speed of kill. The estimated LT50 corresponded to specific extracts and the different concentration levels for the doses. Dose 500 was the effective dose since it took shorter time to kill half of the insects' population. There is a strong body of knowledge regarding the lethal effects of doses in that the stronger the concentration level the more effective the dose is and this might have been reflected in the estimated LT50 for the different doses. Further research should be done to ascertain the claim of the estimated LT50 to rule out if the estimates may have been affected by some other factor. GEE for repeated measures logistic regression method estimated LT50 just as the time-dose-mortality model and logistic regression in terms of using parameters such as time, dose and proportion of dead insects. However, the issue of time as a correlated factor was not taken into account in other methods. This showed that the methods presented in this paper estimated LT50 by taking into account the correlation aspect while the other methods estimated LT50 at each time points. The same results also showed that higher concentration doses are more effective than lower concentration doses. The results of this paper were compared with results from the same methods but applied in a different setting to show that the methods were versatile for analyzing repeated measures dose response data. The LT50 and the confidence intervals of the estimates in this paper were similar to those given by Bugeme et al. (2009) who also used repeated measure logistic regression using GEE. The confidence intervals were bounded and appeared not to be

<https://assignbuster.com/repeated-measures-logistic-regression-health-and-social-care-essay/>

wider This research note is unique in that there is a step-wise procedure on how to analyze repeated measures using R free statistical software especially in estimating LT50. The implications of these findings are that there is need to improve on the way repeated measures data is being analyzed by adopting or using the appropriate methods in analysis such as repeated measures logistic regression which uses GEE. In addition, the kind of the information given will be important in guiding how to analyze and interpret results from repeated measures data. Although repeated measures logistic regression using GEE the analysis takes into account the correlation aspects of the data which is brought about by the time, the exact times of kill is not known since time is used cumulatively to estimate if the insect has been killed at a particular time point. To address this concern effective data collection methods and use of existing methods of estimating LT50 should be used in a complementary fashion. The unstructured correlation matrix was the only one used in repeated measures logistic regression using GEE to estimate LT50. Wider comparisons should be considered to make the research more representative. Repeated measures logistic regression is a very promising appropriate and elegant method for analyzing repeated measures data from dose response studies especially in estimating LT50 and should be employed alongside other existing methods for estimating LT50. The method reaches far beyond traditional methods of statistical analysis that are not well suited to estimating LT50 in repeated measures data. Future studies may involve wider exploration of GEE techniques and further testing and refinement are needed to fully develop its promising capabilities.

Competing interests

NONE

Contributions

GO, GAW and DS analyzed the data. GO produced the first draft of the manuscript. All authors critically reviewed the paper and approved the final version.

Author's information

GO is B. Ed Sc, MSc and a Graduate Research Intern within ICIPE in Nairobi.

AGW is Bsc, MSc, PhD and Senior Lecturer Department of Statistics and Actuarial Science JKUAT. DS is Bsc, MSc, PhD and a Biostatistician within ICIPE in Nairobi.

Acknowledgements

The authors would like to thank all members of the African Insect Science for Food and Health (ICIPE) for their contribution and support. We also thank the members of staff from Faculty of Agriculture and department of Statistics and Actuarial Science at Jomo Kenyatta University of Agriculture and Technology (JKUAT). GO, GAW and DS acknowledge the support of the RUFORUM for the funding the study at JKUAT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The paper is published with the permission of the Director of African Insect Science for Food and Health.