Journal of Modern Applied Statistical Methods

Volume 4 | Issue 2 Article 16

11-1-2005

Comparison Of Statistical Tests In Logistic Regression: The Case Of Hypernatreamia

Stylianos Katsaragakis
University of Athens

Christos Koukouvinos
National Technical University of Athens, ckoukouv@math.ntua.gr

Stella Stylianou
National Technical University of Athens

Eleni-Maria Theodoraki
National Technical University of Athens

Eleni-Maria Theodoraki
National Technical University of Athens

 Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
Katsaragakis, Stylianos; Koukouvinos, Christos; Stylianou, Stella; Theodoraki, Eleni-Maria; and Theodoraki, Eleni-Maria (2005) "Comparison Of Statistical Tests In Logistic Regression: The Case Of Hypernatreamia," *Journal of Modern Applied Statistical Methods*: Vol. 4 : Iss. 2 , Article 16.
DOI: 10.22237/jmasm/1130804100

Journal of Modern Applied Statistical Methods
November, 2005, Vol. 4, No. 2, 514-521
Copyright © 2005 JMASM, Inc.
1538 – 9472/05/$95.00

Comparison Of Statistical Tests In Logistic Regression: The Case Of Hypernatreamia

Stylianos Katsaragakis
University of Athens

Christos Koukouvinos Stella Stylianou Eleni-Maria Theodoraki
National Technical University of Athens

The logistic regression has become an integral component of any medical data analysis concerning binary responses. The main issue rising after the adaptation of the final model is its goodness-of-fit. The fit of the model is assessed via the overall measures and summary statistics and comparing them in the case of hypernateamia.

Key words: Logistic regression, goodness-of-fit, covariates

Introduction

The use of overall summary measures of goodness-of-fit has become an important and easily performed step in building logistic regression models. Pearson chi-square sum-of-squares statistics and the Score test are recommended due to their superior power in the simulations, but one must keep in mind that in small sample cases there is lack of detecting subtle deviations from the model (Hosmer, 1997). When it comes to sparse data, a non-significant result of a goodness-of-fit test does not tell that the model is correct, it just tells that the lack-of-fit is not large enough for the model to be rejected (Kuss, 2002).

In general, there are two different approaches to assessing goodness-of-fit in logistic regression models (e.g., Cook, 1979; Pregibon, 1981). The first one, residual analysis, investigates the model on the level of individuals and looks for those observations which are not adequately described by the model or which are highly influential on the model fit. The second approach seeks to combine the information on the amount of lack-of-fit in a single number. Statistical tests, so-called goodness-of-fit tests, are then calculated to judge if this lack-of-fit is significant or due to random chance and can be distinguished to specific and global. Global tests do not evaluate specific alternatives, rather test unspecific hypotheses of the form 'the model fits' versus the alternative 'the model does not fit'.

The goal is to investigate the choice of statistic test for assessing the coefficients of parameters as well as the goodness of fit by examining the medical disorder called hypernatreamia. For this purpose, three well known statistic tests will be used: the Likelihood Ratio statistic (LR), the Wald test (W) and the Score test (Scr) (Hosmer, 1989), although some authors warn that for large coefficients, standard error is inflated, lowering the Wald statistic (chi-square) value (Hosmer, 1989) and the likelihood-ratio test is more reliable for small sample sizes than the Wald test (Argesti, 1996). Methods for checking goodness-of-fit, are less developed, which may be due to the relative youth and enhanced mathematical complexity of the logistic regression model compared to, for example, the linear regression model (e.g., Bendel, 1977; Cook, 1977).

The study includes 314 patients treated at the Surgery Intensive Care Unit of a central hospital in Athens during 1996 - 2003. All data have been extracted from the Central Data Base of the Unit in which are recorded all demographic information (ID, age, sex, disease,

Stylianos Katsaragakis is an Associate Professor in the First Propedeutic Clinic of Surgery, Ippokratio Hospital, Athens, Greece. Christos Koukouvinos is a Professor in the Department of Mathematics, Zografou 15773, Athens, Greece. Email: ckoukouv@math.ntua.gr. Dr. Stella Stylianou is at the Department of Mathematics, Zografou 15773, Athens, Greece. Eleni-Maria Theodoraki is a postgraduate student in Biostatistics.

APACHE II score), daily biochemical indication and medical treatment and mortality. These patients have been chosen, excluding some from the 364 recorded, due to their staying in the ICU less than 3 days, which is thought to be a cutpoint for the ones who enter only for after surgery treatment. In addition, the patients under examination have not been transported to other hospital in order to be aware of the final condition of their health.

To compare the groups of patients having expressed the disorder hypernatreamia, with a control group, there were 35 patients from the first one with at least one indication of the electrolyte Na >147mmol/l during their staying in the ICU and 279 from the second group. With the aim of studying their behaviour, possible risk factors, sepsis criteria, Apache II score, medical treatment and mortality were examined.

In this article, the case of hypernatreamia with a multiple logistic regression model is considered.

The Logistic Regression Model

Logistic regression is part of generalized linear models (McCullagh, 1983), which allows one to predict a discrete outcome, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Dichotomous (binary) outcome is the most common situation in biology and epidemiology, standing for the presence or absence of a disease, success or failure etc. Although discriminant analysis may also predict group membership (e.g., Costanza, 1979; Efron, 1975), it can be used only with two groups, so in the cases of categorical, or a mix of continuous and categorical covariates, logistic regression is preferred (e.g., Cook, 1979; Fleiss, 1979; Furnival, 1974; Mickey, 1989).

What seems to distinguish logistic regression to linear is conditional mean $E(Y/x)$, the mean value of the outcome variable, given the value of the independent variable. In linear regression, it is assumed that this mean may be expressed as an equation linear in x, which implies that $E(Y/x)$ may take any value as x ranges between $-\infty$ and $+\infty$, but with dichotomous data conditional mean must be greater than or equal to zero and less than or greater to one. The second important difference concerns the conditional distribution of the outcome variable. In the linear regression model, it is assumed that an observation of the outcome variable may be expressed as $y = E(Y/x) + \varepsilon$, where the error ε follows a normal distribution [$\varepsilon \sim N(\mu, \sigma^2)$], whereas in logistic ε follows the binomial one.

Logistic regression makes no assumption about the distribution of the independent or predictor variables, that is they do not have to be normally distributed (Lawless, 1978), linearly related or of equal variance within each group so the relationship between the predictor and response variables is not a linear function.

Let $f(x) = P(Y = 1/\vec{x})$, where the vector

$$\vec{x} = (x_1, x_2, \ldots x_p)$$

denotes a collection of p covariates. Then the logistic regression function, in form of the logit transformation

$$g(\vec{x}) = \ln\left[\frac{f(x)}{1-f(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_p x_p$$

is:

$$f(x) = \frac{e^{g(\vec{x})}}{1+e^{g(\vec{x})}}.$$

During model creation, variables can be entered into the model in the order specified by the researcher or logistic regression can test the fit of the model after each coefficient is added or deleted, called stepwise regression. Stepwise regression is used in the exploratory phase of research but it is not recommended for theory testing. Forward variable selection enters the variables in the block one at a time based on entry criteria and backward stepwise regression appears to be a preferred method of exploratory analysis, where the analysis begins with a full or saturated model and variables are eliminated from the model in an iterative process.

Backward selection is sometimes less successful than forward or stepwise selection because the full model fit in the first step is the

model most likely to result in a complete or quasi-complete separation of response values. The fit of the model is tested after the elimination of each variable to ensure that the model still adequately fits the data. When no more variables can be eliminated from the model, the analysis has been completed. The process by which coefficients are tested for significance for inclusion or elimination from the model involves several different techniques (e.g., Bendel, 1977; Costanza, 1979). Some of these tests are described in the next section.

Assessment of the Coefficients of the Model

A Wald test is used to test the statistical significance of each coefficient β_i in the model. A Wald test calculates a *z* statistic, which is:

$$z = \frac{\beta_i}{SE(\beta_i)}.$$

This z value is then squared, yielding a Wald statistic with a chi-square distribution with p+1 degrees of freedom, where p is the number of covariates. The likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the saturated model (L_1) over the maximized value of the likelihood function for the current model (L_0). The likelihood-ratio test statistic equals:

$$-2\log\left(\frac{L_0}{L_1}\right) = -2[\log(L_0) - \log(L_1)] = -2(L_0 - L_1).$$

This log transformation of the likelihood functions yields a chi-squared statistic with p degrees of freedom equal to the number of covariates of the model. This appears to be the recommended test statistic to use, when building a model through backward stepwise elimination.

The score statistic is a quadratic form based on the vector of partial derivatives of the log-likelihood function with respect to the parameters of interest, evaluated at the values postulated by the null hypothesis.

Let

$$L(\beta|Y) = \prod_{i\in S} P_i^{w_i Y_i}(1-P_i)^{w_i(1-Y_i)} - \prod_{i\in S}\left(\frac{P_i}{1-P_i}\right)^{w_i Y_i}(1-P_i)^{w_i}$$

be the weighted likelihood function and

$$\log_e L(\beta|Y) = \sum_{i\in S}\left\{w_i \log_e\left(\frac{P_i}{1-P_i}\right) + w_i \log_e(1-P_i)\right\}$$
$$= \sum_{i\in S} w_i Y_i X_i^T \beta - \sum_{i\in} w_i \log_e(1+e^{X_i^T\beta})$$

be the log likelihood function. Then, the (p + 1) x 1 score vector, S(β), is given by

$$S(\beta) = \frac{\partial}{\partial\beta}\log_e L(\beta|Y) = \sum_{i\in S} w_i X_i^T (Y_i - P_i)$$

Testing the Fit of the Model

For a particular covariate pattern, the Pearson residual is defined as follows:

$$r\left(y_i, \hat{\pi}_j\right) = \frac{\left(y_j - m_j\hat{\pi}_j\right)}{\sqrt{m_j\hat{\pi}_j\left(1-\hat{\pi}_j\right)}}$$

The summary statistic based on these residuals is the Pearson chi-square statistic

$$X^2 = \sum_{j=1}^{J} r\left(y_j, \hat{\pi}_j\right)^2$$

and the deviance residual:

$$d\left(y_j, \hat{\pi}_j\right) = \pm\left\{2\left[y_j \ln\left(\frac{y_j}{m_j\hat{\pi}_j}\right) + \left(m_j - y_j\right)\ln\left(\frac{\left(m_j - y_j\right)}{m_j\left(1-\hat{\pi}_j\right)}\right)\right]\right\}^{1/2}$$

The distribution of the statistics X^2 and D under the assumption that the fitted model is correct in all aspects is supposed to be chi-square with degrees of freedom equal to J-p-1.

The Hosmer-Lemeshow goodness-of-fit statistic is obtained by calculating the Pearson chi-square statistic from the $2\times g$ table of observed and expected frequencies, where g is the number of groups. The statistic is written as:

$$\chi^2_{HL} = \sum_{i=1}^{g} \frac{\left(O_i - N_i\overline{\pi}_i\right)^2}{N_i\overline{\pi}_i\left(1-\overline{\pi}_i\right)}$$

where N_i is the total frequency of subjects in the ith group, O_i is the total frequency of event outcomes in the ith group, and $\overline{\pi}_i$ is the average estimated probability of an event outcome for the ith group. The Hosmer-Lemeshow statistic is then compared to a chi-square distribution with (g-n) degrees of freedom, where the value of n can be specified in the lackfit option in the model statement. The default is n=2. Large values of X^2_{HL} (and small p-values) indicate a lack of fit of the model.

Comparison of the Coefficients-Results

The data set used to compare the statistical tests contains 24 covariates for each of the two groups of patients under examination (hypernatreamic-control patients). At a brief description it is observed that both groups have statistically comparable ages ($t_{290,\ 0.025}$=-0.753, p=0.452), the sepsis score ($X^2_4(0.05)$=6.979, p=0.137) as well as the **A**cute **P**hysiology **A**nd **C**hronic **H**ealth **E**valuation, ($X^2_1(0.05)_{Kruskall\ \ Wailes}$ = 1.174, p = 0.279), which both estimate the condition of health of each patient at his entrance in the ICU, does not seem to differentiate between two groups.

It is of interest now to explore the relationship between the covariates and the presence or absence of hypernatreamia. Using a univariate model containing the intercept and every time the variable of interest, it seems to exist strong relationships with the binary outcome indicating that patients with high values of Na differentiate from the control group. But can this univariate result be used to confirm, for example, that hypernateamia is associated with mortality - taking under consideration all possible risk factors? That is one of the questions generated and concerns a set of covariates that can be partly answered with a multivariable logistic regression analysis.

For this purpose, variables are included in the model that has been shown to be associated with hypernatreamia. Covariates of interest included age, gender, evaluation of the stage of the patients condition (APACHE, sepsis score), resuscitation fluids and antibiotics containing Na. The multivariate logistic regression model also included the interactions of plasma (FFP) with the antibiotics containing furosemide, teicoplanin and humanxlasix to examine if their combination is mischievous, that is they lead to hypernatreamia.

The analysis was conducted with the SAS program and the method used for the binary model was the full one. 31 observations were deleted due to missing values for the explanatory variables so the number of observations that finally contributed to the analysis was 283 (30 patients who expressed the disorder and 253 control patients). The importance of a variable is defined in terms of a measure of the statistical significance of the coefficient of the model ($p<0.05$), which denotes the fixed decision rule for the inclusion of variables at the procedure used. However there seems to be an indication of the influential role for some covariates ($p<0.10$) that needs to be taken under consideration and are therefore illustrated.

The results for the logistic regression model to be assessed are presented in table 1. Initially the model contained all the possible interaction factors, which have already been discussed, with no statistically significant results; therefore only the main effects were used. With the exception of the design variable sepsis, there is clear evidence that each of the variables has some association with the outcome. This observation is based on an inspection of the 95% Wald confidence interval estimates which, either do not contain 1 or just barely do. At this point, a decision concerning the variable age had to be made, as it is known to be a biologically important variable, yet is not statistically significant in this model. For this reason the covariate's estimate and the Wald test's value at the Analysis of Maximum Likelihood Estimates table were included. In search of a confounding effect, it was found that

Table 1: Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq
Intercept	1	-52.186	353.700	0.022	0.883
APACHE	1	0.121	0.073	2.748	0.097
daysofst	1	2.356	0.624	14.245	0.000
age	1	0.035	0.037	0.884	0.347
qfurosemide	1	-0.145	0.050	8.462	0.004
qffp	1	-0.590	0.253	5.427	0.020
qimipeneme	1	0.844	0.292	8.386	0.004
qteicoplanin	1	1.024	0.527	3.776	0.052
qsod. hlopideamp 15%	1	-0.389	0.109	12.877	0.000
sex (0)	1	1.177	0.597	3.887	0.049
death (0)	1	-3.782	1.068	12.549	0.000
sepsis (0)	1	15.483	8.240	3.531	0.060
sepsis (1)	1	14.758	8.298	3.163	0.075
sepsis (2)	1	12.958	7.949	2.658	0.103
sepsis (3)	1	15.469	8.276	3.494	0.062
ffp (0)	1	-1.099	0.630	3.043	0.081
imipeneme (0)	1	-3.514	1.646	4.559	0.033
teicoplanin (0)	1	-16.705	6.381	6.854	0.000

Table 2: Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
APACHE	0.886	0.767	1.022
daysofst	0.095	0.028	0.322
age	1.035	0.963	1.114
qfurosemide	1.156	1.049	1.275
qffp	1.804	1.098	2.963
qimipeneme	0.430	0.243	0.761
qsod. Chlopideamp 15%	0.359	0.128	1.009
sex (0 vs 1)	0.095	0.009	0.986
death (0 vs 1)	>999.999	29.340	>999.999
sepsis (0 vs 4)	<0.001	<0.001	290.589
sepsis (1 vs 4)	<0.001	<0.001	689.112
sepsis (2 vs 4)	<0.001	<0.001	>999.999
sepsis (3 vs 4)	<0.001	<0.001	337.138
ffp (0 vs 1)	9.006	0.762	106.412
imipeneme (0 vs 1)	<0.001	<0.001	0.562
teicoplanin (0 vs 1)	<0.001	<0.001	<0.001

the absence of age indeed acts as a confounder changing remarkably the significance status of the model. Assessing the reduced model for that case, the LR and Score Tests

$$(X^2_{26}(0.05)_{(LR)(f-age)} = 126.486,$$

$$X^2_{26}(0.05)_{(Scr)(f-age)} = 123.824,\ p<0.0001)$$

agrees with the saturated one

$$(X^2_{27}(0.05)_{(LR)f} = 141.465,\ X^2_{27}(0.05)_{(Scr)f} = 120.634,\ p<0.0001)$$

and there is a small change

$$(X^2_{277}(0.05)_{(Pearson)} = 217.715\ (p=0.997),$$

$$X^2_8(0.05)(HL) = 3.322,\ (p=0.913)$$

in the Pearson and Hosmer-Lemeshow goodness-of-fit tests

$$(X^2_{255}(0.05)_{(Pearson)} = 128.107\ (p=1.000),$$

$$X^2_8(0.05)(HL) = 2.333,\ p=0.969)$$

reflecting the reduction of effectiveness in describing the outcome due to the absence of age.

Examining the results, it was also observed that the estimated coefficients for a set of variables in the model changed significantly when gender was deleted. Hence, there is clear evidence of a confounding effect due to gender describing that it is associated with both the outcome variable of interest, hypernatreamia, and the risk factors. Comparing the LR and Score tests of that model with the full one, it was found that although the LR and Score tests don't seem to denote that the absence of the variable produces an alteration in the model

$$(X^2_{26}(0.05)_{(LR)(f-gender)} = 136.777,$$

$$X^2_{26}(0.05)_{(Scr)\,gender} = 120.05,$$

$$p<0.0001,\ X^2_{27}(0.05)_{(LR)f} = 141.465,$$

$$X^2_{27}(0.05)_{(Scr)f} = 120.634, p<0.0001),$$

the goodness-of-fit statistics seem to ascertain a small one

$$(X^2_{256}(0.05)_{(Pearson)(f-gender)} = 194.389$$

$$(p=0.998),\ X^2_8(0.05)_{(HL)(f-gender)} = 2.127$$

$$(p=0.977),\ X^2_{255}(0.05)_{(Pearson)f} = 128.107$$

$$(p=1.000),\ X^2_8(0.05)_{(HL)f} = 2.334 = 0.969).$$

The confounding status of sepsis score has also been examined, confirming that it is interactively associated with both the disorder and the covariates. The results of the comparison are very interesting since the absence of the polytomous covariate sepsis score produces remarkable changes to the model fit. In specific, although the saturated model seems to fit well, the null hypothesis for the reduced model is rejected

$$(X^2_{259}(0.05)_{(Pearson)(f-sepsis)} = 591.935$$

$$(p<0.001),\ X^2_8(H-L)_f = 20.167\ (p=0.0097)).$$

Considering that the overall goal is to obtain the best fitting model while minimizing the number of parameters, the next step is to fit a reduced model containing only those variables thought to be significant, and compare it to the full model containing all the variables. The results fitting a model with intercepts only and for fitting a model with intercepts and explanatory variables, show that the overall statistic tests reject the global null hypothesis

BETA=0 in the case of both the reduced and the full model.

$(X^2_7(0.05)_{(LR)_r}$ =65.395, $X^2_7(0.05)_{(Scr)_r}$ =94.37

7,p<0.0001) $X^2_{27}(0.05)_{(LR)_f}$ =141.465,

$X^2_{27}(0.05)_{(Scr)_f}$ =120.634, p<0.0001).

However examining the Pearson and Hosmer-Lemeshow statistics

$(X^2_8(0.05)_{(HL)}$ =17.756 (p=0.023),

$X^2_{278}(0.05)_{(Pearson)}$ =1316.375 (p<0.0001)

a remarkable change demonstrating a better fit of the full model is observed

$(X^2_8(0.05)_{(HL)}$ =128.107 p=1.000,

$X^2_{278}(0.05)_{(Pearson)}$ =2.333, p=0.969)).

During model assessment, it was observed that deviance does not seem to alter

$(X^2_{255}(0.05)_{(Deviance)_f} = 49.891$

(p=1.000), $X^2_{277}(0.05)_{(Deviance)_{(f-age)}} = 78.103$(p

=1.000), $X^2_{256}(0.05)_{(Deviance)_{(f-gender)}} = 54.58$

(p=1.000)),

placing all models containing confounders or other reduced models in the same goodness-of-fit status with the full model. That happens even in the last case of the confounding of sepsis score when Pearson and Hosmer-Lemeshow tests agree in rejecting the goodness-of-fit but deviance fails to identify such alteration

$(X^2_{255}(0.05)_{(Deviance)_{(f-sepsis)}} = 88.531$, p=1.000).

The estimated coefficients and odds ratio show that women are 10.6 times more likely to express the disorder (p<0.05) than men, mortality increases to hypernatriemic patients (p<0.01) and the ones with sepsis score 4 are much less likely to get hypernatreamic compared to any of the other 3 sepsis levels (0, 1, 2, 3). In the case of the design variables of sepsis, although between levels 2 and 4 there seems to be a marginal relationship at the 10% level **(**p=0.103), the variable was included because the W statistics for all relative coefficients exceed 2 (Hosmer & Lemeshow, 1989).

There is great interest to the influential part that the antibiotics and resuscitation fluids containing Na, play during patients treatment in ICU. Especially, patients that were treated intravenously with furosemide increased the risk of getting hypernatriemic 15% every time they accepted 20mg as long as getting FFP they increased the risk 9 times from those who didn't (an increase of 1 point led to a 80% increase of risk).

Conclusion

During or after model creation, there seems to be efficiency and applicability of the proposed Wald Test, Likelihood Ratio Test, and Score test, because they agree in refining the significance of the coefficients. Our comparison of the proposed goodness-of-fit statistics Pearson chi-square and Hosmer-Lemeshow, showed small deviations between them at the omission of important confounders, but both are much more powerful from deviance in detecting the fit of the model. That leads to an important association between the behaviour of the logistic regression model through the application of different assessment statistics, in representing best the biological mechanism, hence correctly logistic regression is a significant tool in any medical data analysis of an ordinal response model with both categorical and continuous covariates.

References

Argesti, A. (1996). *An introduction to categorical data analysis*. Wiley.

Bendel, R. B., & Afifi, A. (1977). Comparison of stopping rules in forward regression. *Journal of the American Statistical Association, 72,* 46-53.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, *19,* 15-18.

Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, *74,* 169-174.

Costanza, M. C., & Afifi, A. (1979). Comparison of stopping rules in forward stepwise discriminant analysis. *Journal of the American Statistical Association*, *74,* 777-785.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant function analysis. *Journal of the American Statistical Association, 70,* 892-898.

Fleiss, J. (1979). Confidence intervals for the odds ratio in case control studies: State of the art. *Journal of Chronic Diseases*, *32,* 69-77.

Furnival, G. M., & Wilson, R. W. (1974). Regression by leaps and bounds. *Technometrics*, *16,* 499-511.

Hauck, W. W. (1985). A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *Journal of Chronic Diseases*, *38,* 125-126.

Hosmer, D. W., Hosmer, T., Cessie , S. Le, & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for logistic regression model. *Statistics in Medicine*, *16,* 965-980.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*, Wiley.

Kuss, O. (2002). Global goodness-of-fit tests in logistic regression with sparse data. *Statistics in Medicine*, *21,* 3789-3801.

Lawless, J. F., & Singhal, K. (1978). Efficient screening of non-normal regression models. *Biometrics*, *34,* 318-327.

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. Chapman Hall: London.

Mickey, J., & Greenland, S. (1989). A study of the impact of confounder - selection criteria on effect estimation. *American Journal of Epidemiology*, *129,* 125-137.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, *9,* 705-724.

Pulkstenis, E., & Robinson, T. J. (2002). Two goodness of fit tests for logistic regression models with continuous variables. *Statistics in Medicine*, *21,* 79-83.