EFFECT OF USING BIAS-CORRECTED ESTIMATORS IN LOGISTIC REGRESSION MODEL IN SMALL SAMPLES: PROSTATE-SPECIFIC ANTIGEN (PSA) DATA

M. A. Matin

Department of Statistics, School of Mathematics, University of Leeds, LS2 9JT
Email: matin@maths.leeds.ac.uk

ABSTRACT

This study investigates the effect of bias-corrected estimators in analyzing real-world skewed data where categorization and transformation are necessary. It also reports a small-scale simulation study to indicate factors which can influence the bias correction to be small or large. For the complete data-set, it is observed that the maximum likelihood estimates and Schaefer's bias-corrected estimates are not greatly different. However, when the original sample size is reduced by about 50%, the difference between the estimates is found to be much larger, possibly even large enough to influence the conclusions drawn. The impact of transformation and categorization is visibly present. However, the broad impression gained in categorization is the same though difference in types of categorizations can not be overlooked. A factor which seems to influence the size of the bias correction is identified.

Keywords: Bias correction, Categorization, Log transformation , Small sample, Skewed data.

1 INTRODUCTION

The popular method used to estimate the parameters of a logistic regression (LR) model is the maximum likelihood (ML) method. The ML estimates are asymptotically unbiased. However, for small samples, these estimates have substantial bias and can thereby lead to incorrect conclusions concerning the effects of individual explanatory variables. Bias correction procedures are available in the literature (e.g., Schaefer (1983)). However, these bias corrections have found little use in practice which leads to some obvious questions. How much do these corrections affect the analysis of real-world data? If it is little, what is the justification behind advocating these corrections? Moreover, researchers do choose to categorize the continuous explanatory variable. One reason for this is to see how the logistic transform of the response probability varies over the levels of categories of an inherently continuous variable. Furthermore, for variables having substantial skewness, it is conventional to transform the data (e.g., using the log or square root transformation).

This study aims to search for answers to the questions raised above. In doing so, we use a real-world skewed data set (see section 2 for description) obtained from the Tibblin *et al*. (1995) study. The explanatory variable in the data set is considered in its original continuous form, in logarithmic form and in categorized form of different types. The data is used in its full and 50% reduced size to see the effect of the sample size . We use Schaefer's bias correction and the maximum likelihood estimates for the LR model parameters. To see which factors can influence the bias to be small or large, small-scale simulation experiments are included.

The limitations of such a study should be stressed. From a single real-world application, we can never obtain general results. What we can get is information on the size of differences in a situation when we know what the parameters stand for. This makes it easier to judge whether differences between methods are large or not, something which is not always easy to do in the artificial setting of a simulation experiment. A further application of the different methods to other real-world examples in combination with theoretical and simulation results should in due course give additional information so that it will be possible to judge whether the bias-corrected estimator has to be seriously considered in applied work.

Section 3 introduces Schaefer's bias correction while section 4 is devoted to a description of the categorization of continuous explanatory variable. In section 5, we present the results with discussion. Finally, we conclude in section 6.

2 DESCRIPTION OF DATA

Tibblin *et al*. (1995) present the results of a study which was performed to elucidate the performance of prostate-specific antigen (PSA) as a screening test (PSA, a blood test, is specific for the prostate, but not for clinically significant cancer). To investigate whether an increased PSA level predicts the subsequent occurrence of a clinical cancer and how long the clinical diagnosis can be advanced by PSA testing (the lead time), they performed a case-control study nested within a population-based cohort of men with 11 years of follow-up. The study population consisted of all men born in 1913 and alive in 1980 (at the age of 67) in Gothenberg, Sweden. The actual cohort, the men of 1913, consisted of all men meeting these criteria, who were born on a date divisible by 3. Out of 921 sampled individuals, 707 men participated in the study. The cases were the men who developed cancer during 1981-1992. Sera for 36 subjects who developed cancer were considered in the study. For each case, two individually matched controls were randomly selected from those still alive at the time of diagnosis of the case without themselves having had prostate cancer prior to that date. Because serum was lacking for some subjects, the final analysis included 36 cases and 68 control subjects. For further information about the study see Tibblin *el al*. (1995) and references therein.

Table 1: PSA Data

PSA	Controls	Cases
≤ 1.2	20	1
1.3-2.5	15	3
2.6-4.0	17	3
≥ 4.1	16	29
Total	68	36

Some summary data are presented in Table 1. Further summary statistics for PSA and log PSA (LPSA) are presented in Table 2. It is clear that the log transformed data is much more symmetric than the original strongly skewed data. The standard deviation of the original PSA variable is almost three times as large as the mean value, while for LPSA the standard deviation is slightly smaller than the mean value.

Table 2: Summary Statistics for PSA and LPSA Data

	Mean	SD	Median	Min^m	Max^m	Skewness	Kurtosis
PSA	9.80	26.86	3.40	0.30	234.00	169.54	45.94
LPSA	1.29	1.21	1.22	-1.20	5.46	0.85	1.15

Table 3: Conditional and Unconditional ML Estimates (Standard Error in Parentheses)

	PSA	LPSA	TC		
	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Conditional ML	0.38 (0.12)	1.67 (0.47)	1.26 (1.29)	0.82 (1.18)	3.31 (1.12)
Unconditional ML	0.39 (0.10)	1.84 (0.41)	1.39 (1.20)	1.26 (1.20)	3.59 (1.07)

TC refers to data in categorized data as used in Tibblin *et al*. (1995)

The basic study was a matched case-control study. Tibblin *et al*. (1995) present the results of conditional maximum likelihood method (for the LR model parameters β_i); however, we show the results (Table 3) for both conditional and unconditional ML. It is clear that there is no substantial difference between the conditional and unconditional ML estimates irrespective of whether we use original, log transformed or

categorized data. Therefore, in the following, we will not retain the matching and will consider several different estimates based on the standard logistic regression model.

3 SCHAEFER'S BIAS CORRECTION

Let $Y_i \in \{0, 1\}$ denote a dichotomous dependent variable, and let x_i denote a $k + 1$ dimensional vector of explanatory variables, for the ith observation. The probability that $Y_i = 1$, given the value of x_i, is assumed to be $P(Y_i = 1) = \pi(x_i)$ and is defined by the logistic regression model as $\pi(x) = [1 + e^{-\beta^T \boldsymbol{x}}]^{-1}$ where $\beta^T = (\beta_0, \beta_1, ..., \beta_k)$ is a vector of $k + 1$ parameters of interest, $\boldsymbol{x}^T = (1, x_1, ..., x_k)$ is a vector of explanatory variables. The logit transformation in terms of $\pi(\boldsymbol{x})$ is given by

$$log \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 x_1 + ...+ \beta_k x_k. \quad (1)$$

The ML estimate of $\boldsymbol{\beta}$ can be obtained from the likelihood equation

$$\boldsymbol{X}^T(\boldsymbol{Y} - \pi) = \boldsymbol{0} \quad (2)$$

where $\boldsymbol{X}$ is an $n \times (k + 1)$ matrix of explanatory variables, $\boldsymbol{Y}$ is an $n \times 1$ vector of values of the dependent variable, π is an $n \times 1$ vector of π_i's and $\boldsymbol{0}$ is a $(k + 1) \times 1$ vector of zero's. The likelihood equation is non-linear in $\beta_0, \beta_1, ..., \beta_k$ and is solved by suitable iterative methods.

The Schaefer (1983) bias correction (SBC) formula is given by

$$\hat{\beta}_{SBC} = \hat{\beta}_{ML} - bias\ \hat{\beta}_{ML} \quad (3)$$

where $\hat{\beta}_{ML}$ is a $(k + 1) \times 1$ vector of the ML estimates and

$$bias\ \hat{\beta}_{ML} \approx -1/2(\boldsymbol{X}^T \hat{V} \boldsymbol{X})^{-1} \boldsymbol{X}^T \hat{V} \{(1-2\hat{\pi}_j) x_j^T (\boldsymbol{X}^T \hat{V} \boldsymbol{X})^{-1} x_j\}. \quad (4)$$

Here $\boldsymbol{X}$ is the $n \times (k + 1)$ matrix of explanatory variables, x_j^T is the jth row of $\boldsymbol{X}$, and $\boldsymbol{V}$ is an nxn diagonal matrix of the variances of $\hat{\pi}_j$, $\hat{\pi}_j(1-\hat{\pi}_j)$ that is $\hat{V} = diag\{\hat{\pi}_j (1-\hat{\pi}_j)\}$. The term in the brackets represents an $n \times 1$ vector with the jth element $(1-2\hat{\pi}_j) x_j^T (\boldsymbol{X}^T \hat{V} \boldsymbol{X})^{-1} x_j$.

Using ML estimate $\hat{\pi}_j$ in (4), an estimate of the bias is obtained and can be used in (3).

4 CATEGORIZATION OF CONTINUOUS EXPLANATORY VARIABLE

If we do not have any prior knowledge about the data (such as, for blood pressure data, it is roughly known that certain values represent dangerously low or dangerously high values), a common procedure is to categorize the data as quantiles.

In categorizing continuous data, the first job is to estimate the i^{th} quantile Q_i, $i = 1, 2, ..., k$. We define the design variables (Table 4) for $k = 4$ where the quartiles are based on all observations. However, the Tibblin *et al*. categorization is based on certain natural cut-offs which is also an approximate quartile categorization using only control data.

In our case, Y is the cases and controls with $Y = 1$ for cases and $Y = 0$ for controls, and X is the PSA. The PSA is also considered in its logarithmic form (i.e., X is the LPSA) and in categorized form where X_is are D_is, $i = 2, 3, 4$. The following logistic regression models are used

$$log \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 \text{ PSA}, \tag{5}$$

$$log \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 \text{ LPSA}, \tag{6}$$

$$log \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 D_2 + \beta_2 D_3 + \beta_3 D_4. \tag{7}$$

Table 4: Specification of Design Variables for $k = 4$ $Q_1 = 1.7$, $Q_2 = 3.4$, $Q_3 = 6.2$

Condition on explanatory variable (X)		Design variable			
Tibblin et al. categorization (TC)	Quartile categorization (QC)	D_0	D_2	D_3	D_4
$X_i \leq 1.2$	$X_i < Q_1$	0	0	0	0
$X_i \geq 1.3$ and ≤ 2.5	$X_i \geq Q_1$ and $< Q_2$	1	1	0	0
$X_i \geq 2.6$ and ≤ 4.0	$X_i \geq Q_2$ and $< Q_3$	2	0	1	0
$X_i \geq 4.1$	$X_i > Q_3$	3	0	0	1

The design variable D_0 is used to perform what is often called a trend-test for the variable in categorized form using

$$log \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 D_0. \tag{8}$$

5 RESULTS AND DISCUSSION

In order to analyze the data, maximum likelihood estimate (MLE) and Schaefer's bias correction estimate (SBCE) were used. To test the significance of the regression coefficients in the continuous variable case (or in its logarithmic form), the likelihood ratio test (LRT), the score test (SCT) and the Wald test (WALD) were used. However, only LRT was used to test the overall significance in categorized and linear trend cases. Results are presented in Table 5 through Table 9.

Table 5: Results for LR model with PSA and LPSA Data

	PSA		LPSA	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
MLE	-2.6271	0.3897	-3.2892	1.8423
95% CI	-3.6189 -1.6353	0.2000 0.5794	-4.5985 -1.9799	1.0373 2.6472
OR		1.48		6.31
SBCE	-2.5789	0.3736	-3.2159	1.7735
95% CI	-3.5708 -1.5872	0.1839 0.5633	-4.5252 -1.9066	0.9685 2.5784
OR		1.45		5.89

CI = Confidence Interval

For PSA the odds ratio (OR) associated with a unit change is 1.48 according to the ML method, while the corresponding figure for LPSA is 6.31 (regression coefficients estimated from the LR model are the logarithm of OR). The corresponding figures for the bias corrected method are 1.45 and 5.89. Compared with the lowest category of PSA values, doubtful and non-significant risks are seen for values less than 4.0. In contrast, a PSA value greater than 4.0 is associated with a more than 30-fold increased risk of developing cancer.

For the data set with PSA, the observed absolute difference between MLE and SBCE is found to be 0.0482 (0.0161) for $\hat{\beta}_0$ ($\hat{\beta}_1$). However, for LPSA a much larger difference, 0.0733 (0.0688) for $\hat{\beta}_0$ ($\hat{\beta}_1$), is observed. The relative difference is rather similar, however (4.1 and 3.7% for the slope). These results are not exactly comparable to the simulation study results of Schaefer (1983) and Matin (1994). However, for the sample size 100, Matin (1994) obtained an absolute difference between mean MLE and mean SBCE of 0.0407 (0.0830) for $\hat{\beta}_0$ ($\hat{\beta}_1$) for a model with β_0 = -2 and β_1 = -2.

To show the effect of sample size, the analysis (Table 7) is based on the diagnosis years 1981-86, which comprise 51 observations, and 1987-91, which comprise 53 observations. Thus we reduce the sample size by almost 50%. The difference between MLE and SBCE is now much larger than in the cases previously considered with 104 observations. This result is in agreement with Schaefer (1983) and Matin (1994, 2005) where in small samples the difference between MLE and SBCE is found to be larger. Compared with the original skewed data, a much larger difference between the estimates is obseved for LPSA. The relative difference is rather similar and markedly larger than for the complete data set (as expected).

Table 6: Results for LR Model with Categorized Data

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Tibblin *et al*. Categorization				
MLE	-2.9957	1.3863	1.2611	3.5904
95% CI	-5.0041 -0.9873	-0.9739 3.7464	-1.0926 3.6149	1.4913 5.6895
OR		4.00	3.53	36.25
SBCE	-2.9957	1.3863	1.2611	3.5764
95% CI	-5.0041 -0.9873	-0.9739 3.7464	-1.0926 3.6149	1.4773 5.6755
OR		4.00	3.53	35.74
Quartile Categorization				
MLE	-2.5257	0.0408	2.1893	4.0073
95% CI	-3.9660 -1.0854	-1.9976 2.0793	0.5361 3.8424	2.2703 5.7444
OR		1.04	8.93	56.00
SBCE	-2.5257	0.0408	2.1893	3.9301
95% CI	-3.9660 -1.0854	-1.9976 2.0793	0.5361 3.8424	2.1930 5.6671
OR		1.04	8.93	50.91

CI = Confidence Interval

Table 7: Results for LR Model with PSA and LPSA Data

	1981-86				1987-91			
	PSA		LPSA		PSA		LPSA	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
MLE	-2.7441	0.3388	-4.1773	2.1860	-2.7251	0.4816	-2.8250	1.7005
OR		1.40		8.90		1.62		5.48
SBCE	-2.6138	0.3014	-3.9162	1.9640	-2.6589	0.4535	-2.7504	1.6178
OR		1.35		7.13		1.57	1.00	5.04

Are the differences large enough to be practically meaningful? For n = 104 the ORs obtained are very similar. In small samples, the situation is somewhat different. For LPSA based on diagnosis year 1981-86, the difference between ORs (computed from MLE and SBCE) is found to be approximately 2 units (Table 7). This may be a difference of some importance in practice.

Regarding the significance of the regression coefficient (i.e., to test H_0: β_1 = 0) all the test statistics in the continuous variable case show significant results. For PSA the test statistics show the relationship SCT < WALD < LRT, but for LPSA the relationship is WALD < SCT < LRT (Table 8). Matin (1998, 2005) found the later relationship true for the mean values of the test statistics although that does not hold in every sample. If LPSA rather than PSA is used in univariate analysis, all the test statistics have larger values (Table 8). This supports the logarithmic transformation of the PSA data.

For the Tibblin *et al.* categorized data, with a sample of size 104, the difference between MLE and SBCE is virtually nil except for the fourth category where the difference is found to be only 0.0140 (Table 6). Similar comments can be drawn for the quartile categorized data. These results are not surprising at all, since Schaefer's study (1983) confirms that as the number of explanatory variables (continuous) increases with sample size, the bias reduction becomes negligible. For the quartile categorized data, the difference between ORs (computed from MLE and SBCE of the fourth category) is found to be approximately 5 units (Table 6). As the ORs are large, it is doubtful if this difference is of any practical importance. A strongly significant overall effect is obtained for both categorizations of the data.

Table 8: Test Statistic for Test of Regression Coefficient

	Continuous variable		Categorized variable		Trend	
	PSA	LPSA	TC	QC	TC	QC
WALD	16.21	20.12				
SCT	12.83	35.82				
LRT	44.12	45.22	34.42	47.33	30.97	45.11

Table 9: Results about $\hat{\pi}_j$ and 1 - $2\hat{\pi}_j$ value for Sample Sizes 51 and 104

SN		Mean	SD	Min[m]	Max[m]	Q_1	Q_3	MLE – SBCE $\hat{\beta}_0$	$\hat{\beta}_1$
Sample Size 51									
1	$\hat{\pi}_j$	0.3600	0.3945	0.0094	1.0000	0.0341	0.8930	0.3881	0.1108
	$1 - 2\hat{\pi}_j$	0.3000	0.7890	-1.0000	0.9811	-0.7860	0.9318		
2	$\hat{\pi}_j$	0.3400	0.3597	0.0303	1.0000	0.0650	0.6620	0.1696	0.0502
	$1 - 2\hat{\pi}_j$	0.3400	0.7195	-1.0000	0.9394	-0.3239	0.8701		
3	$\hat{\pi}_j$	0.4000	0.4021	0.0070	1.0000	0.0346	0.9698	0.3522	0.1111
	$1 - 2\hat{\pi}_j$	0.2200	0.8041	-1.0000	0.9860	-0.9396	0.9307		
4	$\hat{\pi}_j$	0.3333	0.3628	0.0278	1.0000	0.0615	0.6771	0.1779	0.0527
	$1 - 2\hat{\pi}_j$	0.3330	0.7257	-1.0000	0.9444	-0.3542	0.8769		
5	$\hat{\pi}_j$	0.3200	0.3692	0.0193	1.0000	0.0452	0.6481	0.1845	0.0545
	$1 - 2\hat{\pi}_j$	0.3800	0.7384	-1.0000	0.9613	-0.2963	0.9096		
Sample Size 104									
1	$\hat{\pi}_j$	0.3495	0.2912	0.0798	1.0000	0.1277	0.4408	-0.0472	0.0157
	$1 - 2\hat{\pi}_j$	0.3107	0.5825	-1.0000	0.8405	0.0814	0.7447		
2	$\hat{\pi}_j$	0.3398	0.3010	0.0652	1.0000	0.1101	0.4381	-0.0494	0.0165
	$1 - 2\hat{\pi}_j$	0.3301	0.6020	-1.0000	0.8695	0.0834	0.7798		
3	$\hat{\pi}_j$	0.2718	0.2817	0.0575	1.0000	0.0873	0.2879	-0.0487	0.0138
	$1 - 2\hat{\pi}_j$	0.4660	0.5634	-1.0000	0.8849	0.3975	0.8255		
4	$\hat{\pi}_j$	0.3398	0.2926	0.0743	1.0000	0.1196	0.4247	-0.0642	0.0156
	$1 - 2\hat{\pi}_j$	0.3301	0.5852	-1.0000	0.8514	0.1137	0.7607		
5	$\hat{\pi}_j$	0.4175	0.3320	0.0573	1.0000	0.1221	0.6661	-0.0642	0.0247
	$1 - 2\hat{\pi}_j$	0.1748	0.6639	-1.0000	0.8855	-0.3837	0.7558		

SN = Sample Number

With the Tibblin *et al*. categorization, we find higher risks (Table 6) in all other categories when compared with the first. The risk increases are not significant in the second and third categories. The very high fourth

quartile OR on the other hand is highly significant. With QC there is no difference between the risk in the first two categories. The third and fourth category risks on the other hand are larger and significant even in the third category case according to the QC. Obviously, the categorization has effect on the results obtained although the broad impression gained is the same.

Categorizing the data on the basis of all observations as in QC or on the basis of the controls only (which is what is approximately done in TC) produces quite different results in a case such as the present one with a large risk associated with the explanatory variable. The large difference between the two categorizations can be seen already from the number of observations in the different categories, a very even distribution of 25-27 in the case of QC, but a much larger variation (21,18, 20 and 45) in the case of TC. A consequence of this is the larger OR obtained in the former case for the fourth category compared with the first one.

5.1 Factors which effect the bias correction

Can we say anything about when the bias is large and when it is small given a certain sample size? The term $1 - 2\hat{\pi}_j$ in (4) is equal to 0 if $\hat{\pi}_j = 0.5$, if $\hat{\pi}_j > 0.5$ $1 - 2\hat{\pi}_j$ assumes negative value, if $\hat{\pi}_j < 0.5$ $1 - 2\hat{\pi}_j$ assumes positive value and $1 - 2\hat{\pi}_j \in (-1\ 1)$. For a particular j, when $\hat{\pi}_j = 0.5$, the term in curly brackets becomes 0 and contributes nothing to the whole term in (4). However, as the value of $1 - 2\hat{\pi}_j$ moves away from zero its contribution increases according to the sign of $1 - 2\hat{\pi}_j$. To see how this term can influence the bias correction, we conducted a small-scale simulation study. In doing so, the original 104 PSA values were considered fixed and the values of Y were generated conditional on PSA as in equation (5), whereby $Y = 1$ with probability $[1 + e^{-(\beta_0 + \beta_1 \mathrm{PSA})}]^{-1}$, and = 0 otherwise. The parameter pair $\beta_0 = -2.6271$, $\beta_1 = 0.3897$ (given in Table 5 as the ML estimates of LR model parameters) was used to compute the probability mentioned above. Then the LR model (5) parameters were estimated with these new cases and controls and the $\hat{\pi}_j$'s were computed. This experiment was repeated five times. The same procedure was applied to PSA based on the diagnosis year 1981-86 with 51 observations and the parameter pair $\beta_0 = -2.7441$, $\beta_1 = 0.3388$ (given in Table 7 as the ML estimates of LR model parameters). Results are presented in Table 9. To facilitate the comparison of these results, the difference between MLE and SBCE is included in the last two columns of Table 9.

With $n = 51$, the differences are found to be larger for the sample numbers 1 and 3, which are accompanied by a smaller mean value of $1 - 2\hat{\pi}_j$. The inter-quartile difference of $1 - 2\hat{\pi}_j$ is larger for the same sample numbers. With $n = 104$, sample number 5 provides similar results. Furthermore we know that the logistic function is essentially linear for $\hat{\pi}_j \in (0.20\ 0.80)$, but outside this interval it becomes markedly non-linear (Collett, 1991). The case where $\hat{\pi}_j \in (0.25\ 0.75)$ for 50% of the indices j, $1 \le j \le n$ is more favourable for ML estimates (Duffy and Santner, 1988). Now as a clue, if the value of $\hat{\pi}_j$ ranges on both sides of 0.5 symmetrically, then each positive contribution of the term $1 - 2\hat{\pi}_j$ is counterbalanced by a negative contribution at least for the linear components of the logistic function. Hence, almost no contribution needs to be added to the whole term in (4). Also, the closer the $\hat{\pi}_j$ values are to 0.5, the smaller the contribution to (4). Smaller contributions are more likely in large samples than in small ones, because cancellation of each positive contribution by a negative one is more likely due to the more symmetric nature expected in large samples.

6 CONCLUSION

With the explanatory variable in its original continuous form, a small difference between MLE and SBCE is observed. However, the conclusions are not affected by the method used. When the original sample size (104) is reduced by almost 50%, the estimated difference is larger. The differences are possibly so large that they may influence the conclusions drawn. The log transformed data is much more symmetric than the

original skewed one. Compared with the skewed data, a much larger absolute difference between the estimates is observed for the log transformed data, although the relative difference is not large. For the original sample size, the ORs obtained are very similar. However, in practice, for small samples the difference between ORs is found to be of some importance. The categorization has an obvious effect on the results obtained. However, no real difference is observed between the estimates of the two methods, the fourth category serving as the only exception. That is, the broad impression gained is the same although the large difference between the two types of categorizations can not be overlooked. It is observed (from the small-scale simulation study) that a smaller mean value of the term $1 - 2\hat{\pi}_j$ is accompanied by a larger difference between the estimates of the two methods used in the study. Thus, the term $1 - 2\hat{\pi}_j$ can influence the bias correction to be small or large. The closer the $\hat{\pi}_j$ values are to 0.5, the lesser will be the contribution to the bias correction.

REFERENCES

Collett, D. (1991). *Modelling binary data*, Chapman & Hall.

Duffy, D. E. and Santner, T. J. (1988). Estimating logistic regression probabilities. In S. S. Gupta and J. O. Berger (Eds), *Statistical decision theory and related topics IV*, Volume 1, pp. 177-194. New York, Springer-Verlag.

Matin, M. A. (1994). Small-sample properties of different tests and estimators of the parameters in the logistic regression model. *Research Report, 94-4, Department of Statistics, Uppsala University* (94-4).

Matin, M. A. (1998). The small-sample properties of the logistic regression model parameters for a real-world based model. *Journal of statistical Studies* **18**, 71-84.

Matin, M. A. (2005). Small-sample properties of estimators and tests in logistic regression with skew-normally distributed explaatory varaibles. *Manuscript.*

Schaefer, R. L. (1983). Bias correction in maximum likelihood logistic regression. *Statistics in Medicine,* **2**, 71-78.

Tibblin, G., Welin, L., Bergstrom, R., Ronquist, G., Norlen, B, and Adami, H. O. (1995). The value of prostate-specific antigen in early diagnosis of prostate cancer: the study of men born in 1913. *Journal of Urology* **154**, 1386-1389.