CME Article

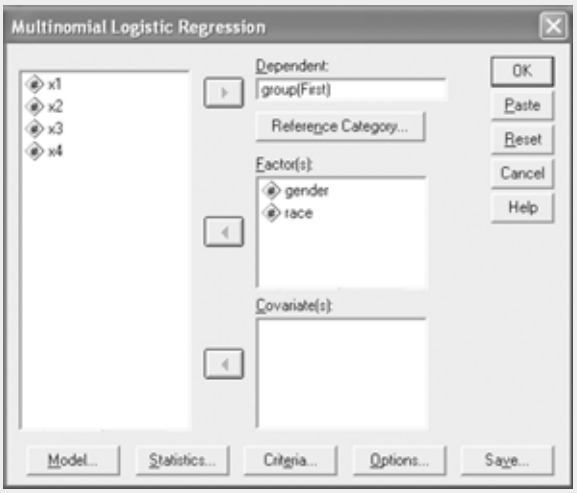
Biostatistics 305. Multinomial logistic regression

Y H Chan

Multinomial logistic regression is the extension for the (binary) logistic regression[1] when the categorical dependent outcome has more than two levels. For example, instead of predicting only dead or alive, we may have three groups, namely: dead, lost to follow-up, and alive. In the analysis to follow, a reference group has to be chosen for comparison, the appropriate group would be the alive, i.e. dead compared to alive and lost to follow-up compared to alive. The predictors used are two categorical (gender and race) and four quantitative variables (x1 – x4).

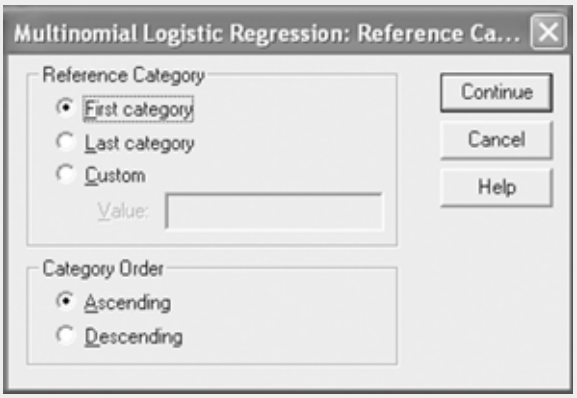
In SPSS, go to Analyse, Regression, Multinomial Logistic to get Template I.

Template I. Multinomial logistic regression.



For the initial analysis, let us just use the two categorical independent variables (gender and race), put them in the Factor(s) option. Put the dependent variable Group (1 = alive, 2 = lost to follow-up, 3 = dead) into the Dependent box. The default Reference-Category is Last. Click on the Reference Category button to get Template II.

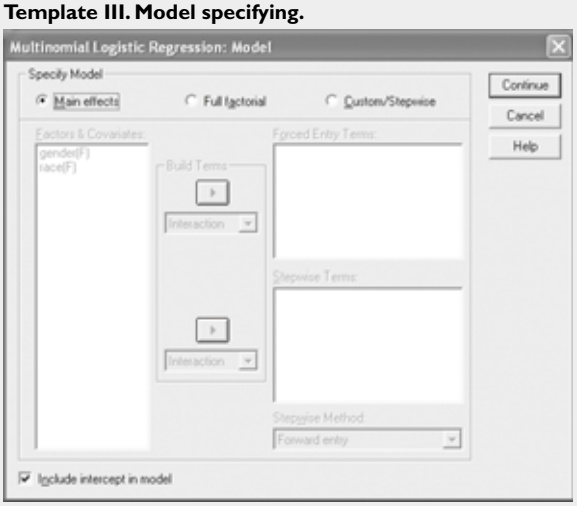
Template II. Reference category definition.



Change the Reference Category to "First category". Leave the Category Order to "Ascending", this means that the smallest value is the first category. The "Descending" option means that the highest category is the first category (a very misleading and redundant option – need to be cautious!).

Click the Model folder in Template I to define the variables to be included in the model, see Template III. The Main effects option will include all the variables specified with no interaction terms whereas the Full factorial option will provide the main effects with all possible interactions. For the Custom/Stepwise option, we have a choice to set up the relevant main effects and interaction terms using the Forced Entry option or to perform a Stepwise analysis. Let us use the Main effects option.

Template III. Model specifying.



Faculty of Medicine
National University of Singapore
Block MD11
Clinical Research Centre #02-02
10 Medical Drive
Singapore 117597

Y H Chan, PhD
Head
Biostatistics Unit

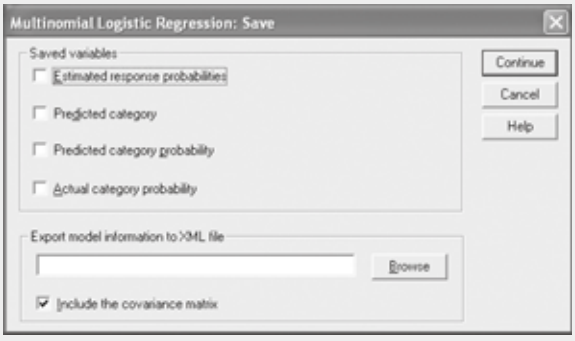
Correspondence to:
Dr Y H Chan
Tel: (65) 6874 3698
Fax: (65) 6778 5743
Email: medcyh@nus.edu.sg

Click on the Statistics folder in Template I.

Template IV. Multinomial statistics folder.

In Template IV, besides the default checked items, tick on Classification table and Goodness-of-fit options. The available saved options (see Template V) could be obtained from the Saved folder in Template I.

Template V. Saved options.



SPSS multinomial output (Gender + Race model)

The heading of the output is "Nominal regression", this assumes that there is no "ranking ordering" in the categorical outcome. Observe that we have six subpopulations (given by 2 [gender] X 3 [race]), see Table Ia. If there are no zero frequencies in each of the subpopulation, no warning-message will be displayed.

Table Ia. Case processing summary: Gender + Race.

Case processing summary			
		N	Marginal percentage
Group	Alive	99	32.0%
	Lost to follow-up	108	35.0%
	Dead	102	33.0%
Gender	Male	151	48.9%
	Female	158	51.1%
Race	Chinese	155	50.2%
	Malay	90	29.1%
	Indian	64	20.7%
Valid		309	100.0%
Missing		0	
Total		309	
Subpopulation		6	

Table Ib. Model fitting information: Gender + Race.

Model fitting information				
Model	-2 log likelihood	Chi-square	df	Sig.
Intercept only	63.979			
Final	50.506	13.473	6	.036

Table Ib shows whether this Gender + Race model gives adequate predictions compared to the Intercept Only (Null model). The Null model uses the modal class (lost to follow-up), see Table Ia, as the model's prediction accuracy – 35%. We want the p-value (sig) of Final to be <0.05. Table Ic shows that this Gender + Race model compared to the Null model gives better accuracies for the "alive" and "lost to follow-up" groups but not for the "dead" group. Though the Model fitting information shows that the current model is outperforming the null, we see that it is not a "good" model if our interest is to predict the "dead" group.

Table Ic. Predictions of the Gender + Race model.

Classification				
	Predicted			
Observed	Alive	Lost to follow-up	Dead	Percent correct
Alive	49	37	13	49.5%
Lost to follow-up	33	54	21	50.0%
Dead	33	51	18	17.6%
Overall percentage	37.2%	46.0%	16.8%	39.2%

Table Id. Goodness-of-fit: Gender + Race.

Goodness-of-fit			
	Chi-square	df	Sig.
Pearson	2.230	4	.694
Deviance	2.216	4	.696

Table Id shows whether the model adequately fits the data. We want the p-values (sig) >0.05. If no warning message is given or the number of subpopulations (cells) with zero frequencies is small, with p>0.05, we could conclude that this model adequately fits the data.

Table Ie. Pseudo R-square: Gender + Race.

Pseudo R-square	
Cox and Snell	.043
Nagelkerke	.048
McFadden	.020

Table Ie indicates the proportion of variation being explained by the model. Only about 5% (maximum 100%) is being explained by the Gender + Race model!

Table If. Likelihood Ratio test: Gender + Race.

Likelihood ratio tests				
Effect	-2 log likelihood of reduced model	Chi-square	df	Sig.
Intercept	50.506[a]	.000	0	.
Gender	60.405	9.899	2	.007
Race	54.405	3.899	4	.420

The chi-square statistics is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

[a.] This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

The Likelihood ratio test (Table If) shows the contribution of each variable to the model – Gender had a significant (p<0.05) contribution but not Race.

Table Ig. Parameter estimates: Gender + Race.

Parameter estimates									
								95% confidence interval for Exp(B)	
Group[a]		B	Std. error	Wald	df	Sig.	Exp(B)	Lower bound	Upper bound
Lost to follow-up	Intercept	.912	.371	6.052	1	.014			
	[gender=1]	-.759	.286	7.012	1	.008	.468	.267	.821
	[gender=2]	0[b]	.	.	0	.	.	.	.
	[race=1.00]	-.624	.379	2.716	1	.099	.536	.255	1.126
	[race=2.00]	-.352	.422	.695	1	.404	.703	.308	1.608
	[race=3.00]	0[b]	.	.	0	.	.	.	.
Dead	Intercept	.854	.376	5.162	1	.023			
	[gender=1]	-.808	.291	7.718	1	.005	.446	.252	.788
	[gender=2]	0[b]	.	.	0	.	.	.	.
	[race=1.00]	-.627	.386	2.637	1	.104	.534	.251	1.139
	[race=2.00]	-.276	.427	.418	1	.518	.759	.329	1.751
	[race=3.00]	0[b]	.	.	0	.	.	.	.

[a.] The reference category is: alive.

[b.] This parameter is set to zero because it is redundant.

How do we interpret Table Ig? The nominal order of Gender and Race are given in Table Ia. For Gender, Male = 1 and Female = 2, the comparison will be male compared to female. The first half of Table Ig has the outcome of "lost to follow-up" compared to "alive" – males compared to females were less likely to be "lost to follow-up", Odds Ratio (OR) = 0.468 (95% CI 0.267 to 0.821), p=0.008. Conversely, we can say that females were more prone to be "lost to follow-up", OR = 2.14 (given by the reciprocal of 0.468). Similarly, females were also more likely to be "dead" – OR = 2.24 (95% CI 1.27 to 3.97), p=0.005.

For Race, the reference group is Indian (from Table Ia, Indian = 3); Race = 1 compares Chinese with Indians, and Race = 2 compares Malays with Indians.

This Gender + Race model is not very adequate – poor prediction for the "dead" group and very low Pseudo R-square, though with adequate goodness-of-fit. In our next model, we shall include the 4 quantitative variables (put x1 – x4 into the Covariate option in Template I).

SPSS multinomial outputs (Gender + Race + x1 to x4 Model)

The first table we get is a warning message (Table IIa).

Table IIa. Warning message.

Warning
There are 618 (66.7%) cells (i.e., dependent variable levels by subpopulations) with zero frequencies.

The reason this warning comes up is that the model includes the continuous covariates (x1 – x4) which results in many subpopulations, 618 + 309 = 927 of them of which 618 are empty and 309 with data (see Table IIb).

Table IIb. Case processing summary: Gender + Race + x1 to x4.

Case processing summary			
		N	Marginal percentage
Group	Alive	99	32.0%
	Lost to follow-up	108	35.0%
	Dead	102	33.0%
Gender	Male	151	48.9%
	Female	158	51.1%
Race	Chinese	155	50.2%
	Malay	90	29.1%
	Indian	64	20.7%
Valid		309	100.0%
Missing		0	
Total		309	
Subpopulation		309[a]	

[a] The dependent variable has only one value observed in 309 (100.0%) subpopulations.

Table IIc. Model fitting information: Gender + Race + x1 to x4.

Model fitting information				
Model	-2 log likelihood	Chi-square	df	Sig.
Intercept only	678.536			
Final	170.343	508.193	14	.000

This model with the addition of x1 – x4 also outperforms the null model (Table IIc) with much improved accuracies for all three groups (Table IId)

Table IId. Prediction accuracies: Gender + Race + x1 to x4.

Classification				
	Predicted			
Observed	Alive	Lost to follow-up	Dead	Percent correct
Alive	84	13	2	84.8%
Lost to follow-up	5	103	0	95.4%
Dead	3	0	99	97.1%
Overall percentage	29.8%	37.5%	32.7%	92.6%

Table IIe. Goodness-of-fit: Gender + Race + x1 to x4.

Goodness-of-fit			
	Chi-square	df	Sig.
Pearson	186196.512	602	.000
Deviance	170.343	602	1.000

Because of the many cells with zero frequencies, this goodness-of-fit test is not relevant now (Table IIe) – ignore this table.

Table IIf. Pseudo R-square: Gender + Race + x1 to x4.

Pseudo R-square	
Cox and Snell	.807
Nagelkerke	.908
McFadden	.749

The pseudo R-square has also increased tremendously, explaining about 75% of the variance (Table IIf).

Table IIg. Likelihood ratio tests: Gender + Race + x1 to x4.

Likelihood ratio tests				
Effect	-2 log likelihood of reduced model	Chi-square	df	Sig.
Intercept	170.343[a]	.000	0	.
x1	357.036	186.693	2	.000
x2	446.851	276.508	2	.000
x3	173.332	2.989	2	.224
x4	175.218	4.875	2	.087
Gender	174.002	3.659	2	.160
Race	172.272	1.929	4	.749

The chi-square statistics is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

[a] This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Significant contributors to the model are x1 and x2 (Table IIg).

For quantitative variables, parameters with significant positive (negative) coefficients increase (decrease) the likelihood of that response category with respect to the reference category. Subjects with increased x1 and decreased x2 were more likely to default whereas those with decreased x1 and increased x2 were more likely to be "dead".

ORDINAL REGRESSION

When the categorical outcomes have an ordinal nature (for example: alive, half-dead, dead – if we consider half-dead is "better" than being dead), the Ordinal regression procedure (also referred to as PLUM) could be used. Here the interest is to determine the direction of the relationship between each predictor and the ordinal nature of the categorical outcome.

Table IIh. Parameter estimates: Gender + Race + x1 to x4.

Parameter estimates									
								95% confidence interval for Exp(B)	
Group[a]		B	Std. error	Wald	df	Sig.	Exp(B)	Lower bound	Upper bound
Lost to follow-up	Intercept	-3.097	6.512	.226	1	.634			
	x1	1.364	.213	41.164	1	.000	3.912	2.579	5.935
	x2	-1.423	.225	39.924	1	.000	.241	.155	.375
	x3	-.051	.053	.898	1	.343	.951	.856	1.056
	x4	.111	.088	1.605	1	.205	1.118	.941	1.328
	[gender=1]	-.697	.486	2.055	1	.152	.498	.192	1.292
	[gender=2]	0[b]	.	.	0	.	.	.	.
	[race=1.00]	-.667	.636	1.100	1	.294	.513	.148	1.785
	[race=2.00]	-.750	.706	1.130	1	.288	.472	.118	1.883
	[race=3.00]	0[b]	.	.	0	.	.	.	.
Dead	Intercept	-14.419	9.189	2.463	1	.117			
	x1	-1.240	.281	19.434	1	.000	.289	.167	.502
	x2	1.448	.278	27.046	1	.000	4.255	2.466	7.345
	x3	.177	.126	1.980	1	.159	1.194	.933	1.528
	x4	-.321	.193	2.767	1	.096	.725	.497	1.059
	[gender=1]	-.997	.806	1.530	1	.216	.369	.076	1.791
	[gender=2]	0[b]	.	.	0	.	.	.	.
	[race=1.00]	-.351	1.022	.118	1	.732	.704	.095	5.222
	[race=2.00]	.287	1.128	.065	1	.799	1.333	.146	12.149
	[race=3.00]	0[b]	.	.	0	.	.	.	.

[a] The reference category is: alive.

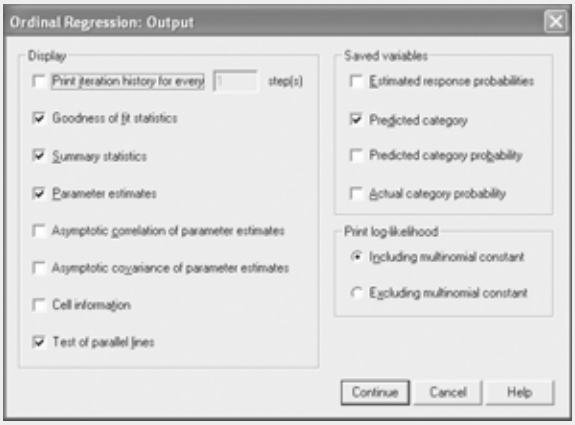
[b] This parameter is set to zero because it is redundant.

In SPSS, go to Analyze, Regression, Ordinal to get Template VI.

Template VI. Ordinal regression.

The setting up of the variables is similar to that of Multinomial except that we do not need to define the reference category as the outcome is ordinal. Click on the Output folder in Template VI to get Template VII.

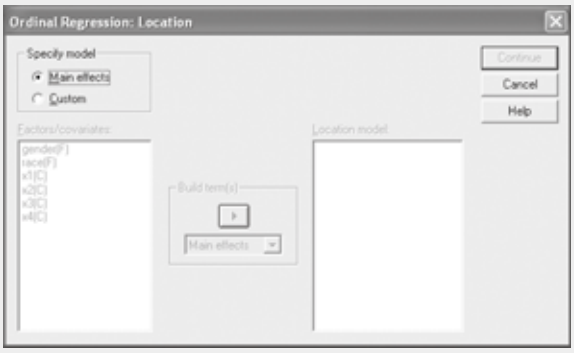
Template VII. Ordinal regression: output.



Besides the default checks, tick on Test of parallel lines, the options of saving the predicted results are available here too. Tick on the Predicted category (this will produce a new variable Pre_1 – Ordinal regression does not have the Classification table option, we have to cross-tabulate Pre_1 with Group to determine the model's accuracies).

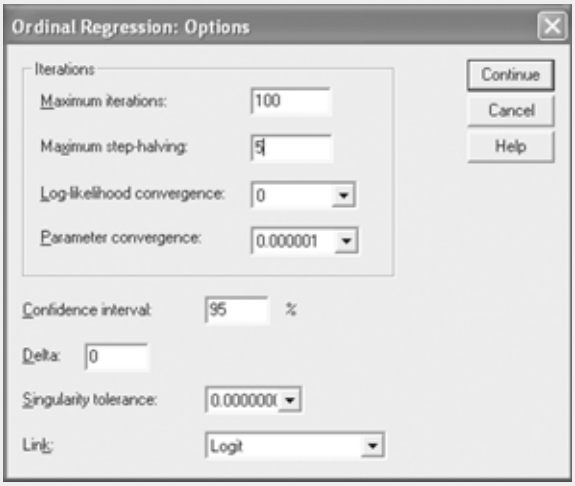
Click on the Location folder in Template VI to define the model. In Template VIII, click on Cancel if we want the Main effects model only, otherwise set-up the Custom model.

Template VIII. Ordinal regression: location.



Click on the Options folder in Template VI, to get Template IX.

Template IX. Choosing the Link function.



The link function is a transformation of the cumulative probabilities of the ordinal outcome to be used in the estimation of the model. Five link functions are available, see Table III. To check the distribution of the ordinal outcome, a bar chart would be most appropriate (Fig. 1). The three groups are quite evenly distributed, thus the Logit link function would be used.

Table III. Link functions.

Link function	Typical application
Logit	Evenly distributed categories
Complementary log-log	Higher categories more probable
Negative log-log	Lower categories more probable
Probit	Latent variable is normally distributed
Cauchit (inverse Cauchy)	Latent variable has many extreme values

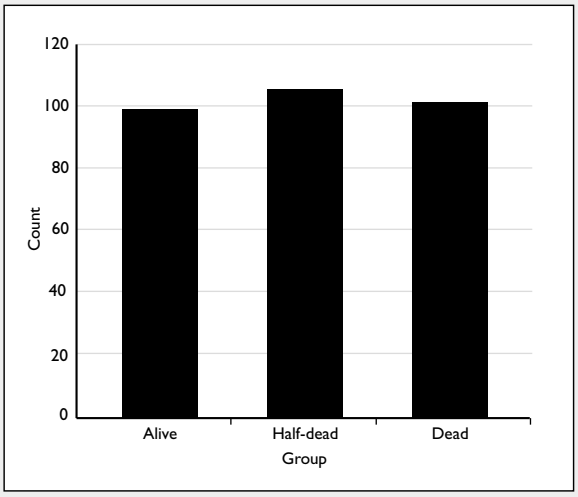
Table IVa. Parameter estimates: Ordinal regression.

Parameter estimates								
							95% confidence interval	
		Estimate	Std. error	Wald	df	Sig.	Lower bound	Upper bound
Threshold	[group = 1]	4.699	2.911	2.605	1	.107	-1.007	10.404
	[group = 2]	6.756	2.924	5.338	1	.021	1.025	12.487
Location	x1	-.196	.046	18.267	1	.000	-.286	-.106
	x2	.296	.041	51.397	1	.000	.215	.377
	x3	-.005	.027	.039	1	.844	-.058	.047
	x4	-.010	.043	.049	1	.825	-.094	.075
	[gender=1]	-.715	.234	9.377	1	.002	-1.173	-.257
	[gender=2]	0[a]	.	.	0	.	.	.
	[race=1.00]	-.541	.303	3.192	1	.074	-1.134	.052
	[race=2.00]	-.218	.336	.423	1	.515	-.876	.439
	[race=3.00]	0[a]	.	.	0	.	.	.

Link function: Logit.

[a.] This parameter is set to zero because it is redundant.

Fig. 1 Bar chart shows distribution of ordinal outcomes.



The SPSS outputs for Ordinal regression are similar to those of Multinomial. We will only discuss on the interpretation of the parameter estimates (Table IVa) and the parallel line testing (Table IVb).

The Threshold portion shows the constants/ intercepts of the model. Significant predictors (for Location) are x1, x2 and gender. A positive relationship exists between x2 and the ordinal outcome. This means that as x2 increases, so does the probability of being in one of the higher categories. On the other hand, x1 has a negative relationship. For Gender, males compared to females had a lower probability to be in a higher category. For Logit link, taking the exponential of the estimates gives us the Odds ratios. For example, a unit increase in x2 will result in an OR of exp (0.296) = 1.34 increase in odds of being in a higher category of the ordinal outcome. For the other link functions, there is no direct interpretation of the estimates due to the complicated nature of the link.

Table IVb. Test of Parallel lines.

Test of Parallel lines[c]				
Model	-2 log likelihood	Chi-square	df	Sig.
Null hypothesis	544.429			
General	231.326[a]	313.103b	7	.000

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

[a.] The log-likelihood value cannot be further increased after maximum number of step-halving.

[b.] The chi-square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.

[c.] Link function: logit.

The test of Parallel lines assesses whether the assumption of all categories having the same parameters is reasonable or not, i.e. whether one set of coefficients for all the categories is appropriate. We want the p-value (sig) for the General in Table IVb to be >0.05. Here p<0.001 means that separate parameters for each category would be more appropriate and thus this current model may not be suitable. This unsuitability could be due to the use of wrong link function or wrong ordering of the categories (perhaps it is better to be dead rather than

half-dead!). We could remodel by using a different link function, the next appropriate one is the Cauchit since the other three link functions would not be "correct" (because of the evenly distributed categories and some of the variables would not satisfy the normal assumptions). If the Cauchit link function is still not appropriate, try re-ordering using alive, dead, half-dead. If all else fails then we have to resort to multinomial regression - ignoring that there is an ordinal nature in the categories.

A word of caution, the p-value of this parallel line test is sensitive to the sample size and the number of independent variables included into a model. Most of the time it has $p<0.05$, we could assess a model via its Pseudo R-square and Classification table of accuracies.

In Template VI, there is the Scale folder which allows us to add in the scale component. This is an optional modification to the basic model to account for differences in variability for different values of the predictor variables. For example, if men have more variability than women in their outcome values, using a scale component to account for this may improve the model. Interested readers could refer to any standard text on Ordinal regression for further information.

CONDITIONAL LOGISTIC REGRESSION FOR MATCHED CASE-CONTROL STUDY

The multivariate extension for McNemar Test for matched case-control study is the Conditional logistic regression. The Multinomial logistic regression can be used to analyse the 1-1 matching (say, by age and gender) in which one case has only one matching control.

Table Va shows the 1st five cases of a matched case-control study. The outcome is death, each death case is matched with an alive person by age and gender. Table Vb shows the variables needed to be computed before we can perform a 1 to 1 conditional logistic which is based on the difference between the case and control. A column of Outcome = 1 is required and the differences for x1 to x3 between dead and alive needs to be computed. For diabetes (1 = yes, 0 = no), to compute diabetes_diff, simply use diabetes_dead - diabetes_alive. For race (1 = Chinese, 2 = Malay and 3 = Indian), a reference category is required, let us say Chinese. Then we need to create dummy variables for the Malays and Indians for both dead and alive groups. For instance, Malay_dead = 1 if the race of the dead person is a Malay otherwise 0; likewise create for the rest: Malay_alive, Indian_dead and Indian_alive. Lastly, compute the Malay_diff using Malay_dead - Malay_alive (similarly for Indian_diff).

Table Va. First five cases of a Matched case-control study: outcomes.

Dead					Alive				
x1	x2	x3	diabetes	race	x1	x2	x3	diabetes	race
84.00	82.10	45.00	1	1	84.0	73.2	47.0	0	1
83.10	86.40	52.00	1	3	86.1	81.1	51.0	1	1
84.60	87.00	53.00	1	2	87.3	76.8	48.0	0	3
84.00	78.80	50.50	0	2	84.2	71.4	48.5	1	1
83.50	88.20	46.00	1	2	83.2	73.7	47.0	1	1

Table Vb. First five cases of a Matched case-control study: variables.

Outcome	x1_diff	x2_diff	x3_diff	Diabetes diff	Malay dead	Indian dead	Malay alive	Indian alive	Malay diff	Indian diff
1	0.00	8.90	-2.00	1.00	0.00	0.00	0.00	0.00	0.00	0
1	-3.00	5.30	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1
1	-2.70	10.20	5.00	1.00	1.00	0.00	0.00	1.00	1.00	-1
1	-0.20	7.40	2.00	-1.00	1.00	0.00	0.00	0.00	1.00	0
1	0.30	14.50	-1.00	0.00	1.00	0.00	0.00	0.00	1.00	0

To perform the analysis in SPSS, go to Analyse, Regression, Multinomial logistic – put the Outcome variable into the Dependent option and all the difference variables computed earlier into the Covariates option (see Template X). Note that the matched variables, age and gender, are not included in the definition of the analysis but could be used for interaction terms in the modelling. The same difference procedure must be followed for the interaction terms - the interaction variables must be created first and then differenced.

Click on the Model folder, Template III is obtained – ** IMPORTANT ** – have to uncheck the "Include Intercept in model" option. Let us use the main effects model.

The goodness-of-fit statistics and the classification table are not valid for matched case-control studies (do not need them), the Model fitting information, the likelihood ratio and R-square statistics are valid and interpreted as usual. Table Vc shows the message that a conditional logistic regression is being performed and Table Vd shows the parameter estimates.

Template X. Conditional logistic definition.

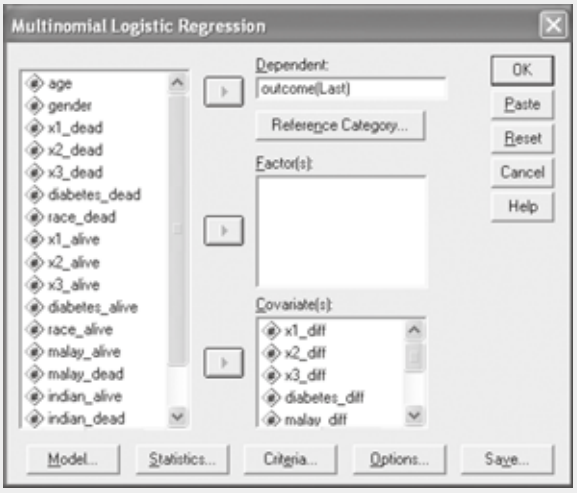


Table Vc. Conditional logistic regression message.

Warning
The dependent variable has only one valid value. A conditional logistic regression model will be fitted.

The significance value of the test for the difference in x2, x3 and diabetes are less than 0.05 – subjects with higher values of x2, lower values of x3 and diabetic are at a higher risk to mortality. The Exp (B) shows the change in the odds of mortality for a one-unit change in the predictor.

For n:m matching case-control study, we will have to use Cox regression[2] to do the analysis. Let us discuss using the above 1:1 matching first. Table VIa shows the data structure for the first three matched subjects (by age and gender).

Table Vd. Parameter estimates: conditional logistic.

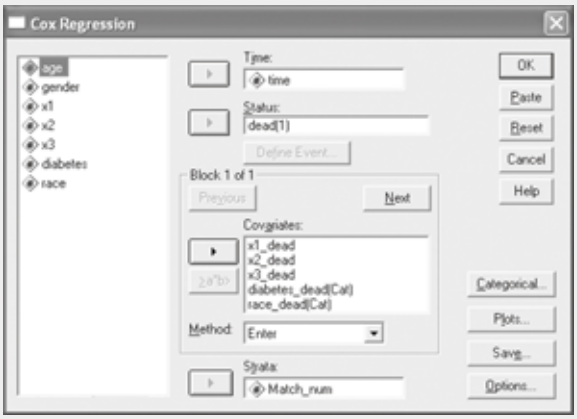
Parameter estimates									
								95% confidence interval for Exp(B)	
Outcome		B	Std. error	Wald	df	Sig.	Exp(B)	Lower bound	Upper bound
Dead	x1_diff	.021	.078	.071	1	.789	1.021	.877	1.188
	x2_diff	.400	.086	21.496	1	.000	1.492	1.260	1.767
	x3_diff	-.499	.143	12.088	1	.001	.607	.458	0.804
	diabetes_diff	1.065	.521	4.188	1	.041	2.902	1.046	8.049
	Malay_diff	-.178	.629	.080	1	.777	.837	.244	2.871
	Indian_diff	-.112	.647	.030	1	.862	.894	.252	3.173

Table VIa. Conditional logistic regression (1:1 matching) using Cox regression option.

Matching number	x1	x2	x3	Dead	Diabetes	Race	Time
1	84.0	82.1	45.0	1	1	1	1
1	84.0	73.2	47.0	0	0	1	2
2	83.1	86.4	52.0	1	1	3	1
2	86.1	81.1	51.0	0	1	1	2
3	84.6	87.0	53.0	1	1	2	1
3	87.3	76.8	48.0	0	0	3	2

We need a matching number to "link" the case and control. The Dead variable is the outcome status of the subject (dead = 1 and alive = 0). Need a variable Time as the response variable where the dead has a Time = 1 and the alive (censored) has Time = 2. To perform the analysis, in SPSS, go to Analyse, Survival, Cox Regression to get Template XI.

Template XI. Conditional logistic using Cox regression.



Put Time in the Time option, dead in the Status option (define Event = 1). Put the variables of interest into the Covariates option and lastly include the Match number (Match_num) in the Strata option. This will produce exactly the same results in Table Vd.

Table VIb shows the data structure for a n:m matching. The n and m do not need to be "fixed" in the same study, i.e. we can have 1:3, 2:3, etc. Age is the matching variable which will be used in the Strata option (see Template XI).

Table VIb. Conditional logistic regression (n:m matching) using Cox regression option.

Matching by age	Outcome Case =1 Control = 0		Time
16	Case		1
16	Control		2
16	Control		2
16	Control	Relevant Variables	2
17	Case		1
17	Case		1
17	Control		2
17	Control		2
17	Control		2

For our next article, we shall discuss the analysis of count data: Biostatistics 306. Loglinear models – poisson regression.

REFERENCES

1. Chan YH. Biostatistics 202. Logistic regression analysis. Singapore Med J 2004; 45:149-53.
2. Chan YH. Biostatistics 203. Survival analysis. Singapore Med J 2004; 45:249-56.

SINGAPORE MEDICAL COUNCIL CATEGORY 3B CME PROGRAMME

Multiple Choice Questions (Code SMJ 200506A)

	True	False
Question 1. In the Multinomial regression, which test gives the contribution of each independent variable to a model?		
(a) Model fitting information.	❑	❑
(b) Goodness-of-fit.	❑	❑
(c) Likelihood ratio.	❑	❑
(d) Pseudo R-square.	❑	❑
Question 2. Which of the following tests are not valid for a matched case-control study?		
(a) Model fitting information.	❑	❑
(b) Goodness-of-fit.	❑	❑
(c) Likelihood ratio.	❑	❑
(d) Pseudo R-square.	❑	❑
Question 3. Which link function should be used if the distribution of the categorical outcome for an Ordinal regression is left-skewed?		
(a) Logit.	❑	❑
(b) Complementary log-log.	❑	❑
(c) Negative log-log.	❑	❑
(d) All of the above.	❑	❑
Question 4. Which technique could be used for a n:m matched case-control study?		
(a) Multinomial logistic.	❑	❑
(b) Cox regression.	❑	❑
(c) Ordinal regression.	❑	❑
(d) All of the above.	❑	❑
Question 5. In which technique is the Parallel line test needed?		
(a) Multinomial logistic.	❑	❑
(b) Cox regression.	❑	❑
(c) Ordinal regression.	❑	❑
(d) All of the above.	❑	❑

Doctor's particulars:

Name in full: _____

MCR number: _____ Specialty: _____

Email address: _____

Submission instructions:

A. Using this answer form

1. Photocopy this answer form.
2. Indicate your responses by marking the "True" or "False" box ☑
3. Fill in your professional particulars.
4. Either post the answer form to the SMJ at 2 College Road, Singapore 169850 OR fax to SMJ at (65) 6224 7827.

B. Electronic submission

1. Log on at the SMJ website: URL http://www.sma.org.sg/cme/smj
2. Either download the answer form and submit to smj.cme@sma.org.sg OR download and print out the answer form for this article and follow steps A. 2-4 (above) OR complete and submit the answer form online.

Deadline for submission: (June 2005 SMJ 3B CME programme): 12 noon, 25 July 2005

Results:

1. Answers will be published in the SMJ August 2005 issue.
2. The MCR numbers of successful candidates will be posted online at http://www.sma.org.sg/cme/smj by 20 August 2005.
3. Passing mark is 60%. No mark will be deducted for incorrect answers.
4. The SMJ editorial office will submit the list of successful candidates to the Singapore Medical Council.