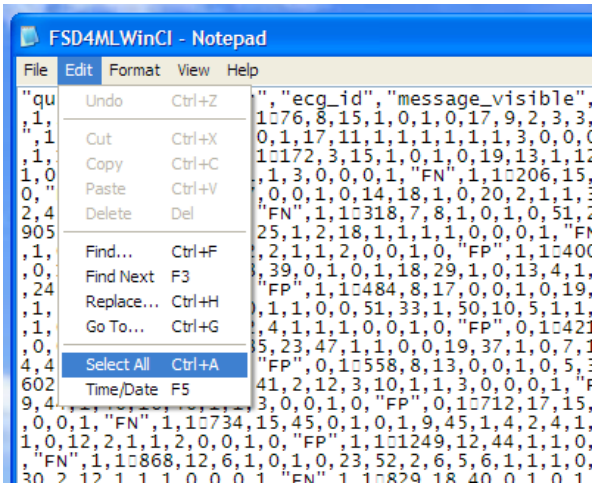
MLWin user guide

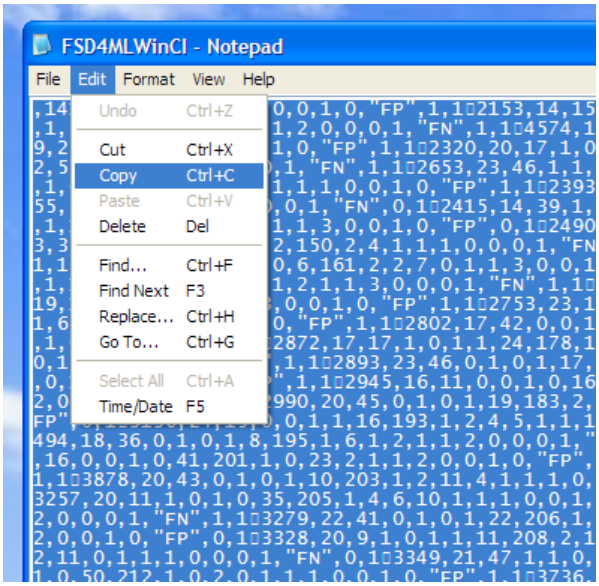
This guide will explain how to take exported CSV files from the R script ProcessRawData.R, import the data into MLWin and run frequentist and Bayesian analysis to compare against the R output.

1. Open FSD4MLWin.csv (or equivalent files for subsets of correct and incorrect computer interpretation) in Notepad

2. Edit > Select All

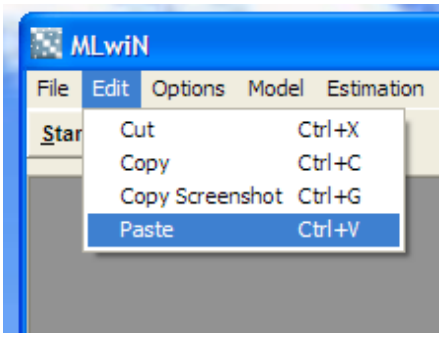


3. Edit > Copy

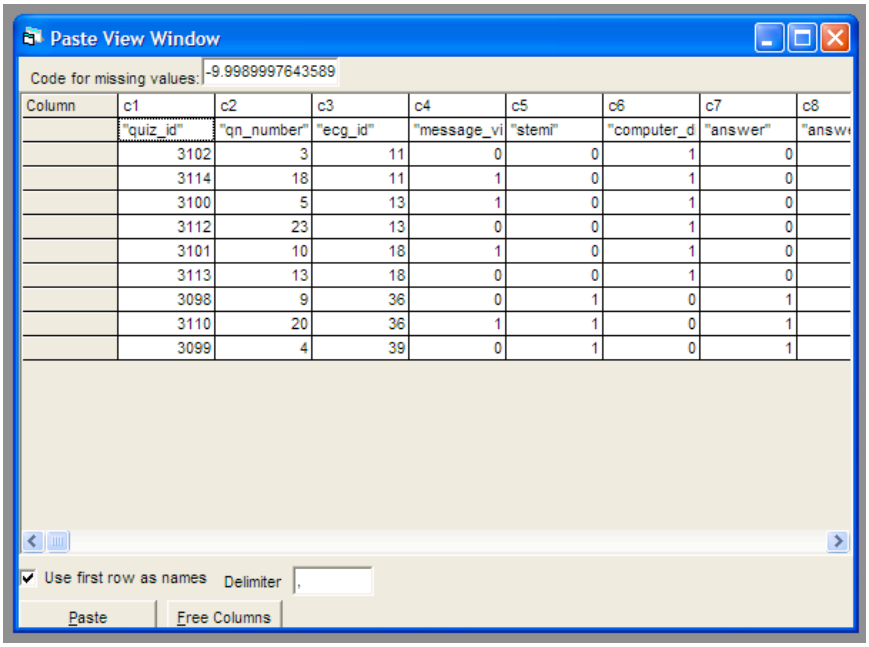


4. Open MLWin and create a new worksheet

5. Edit > Paste

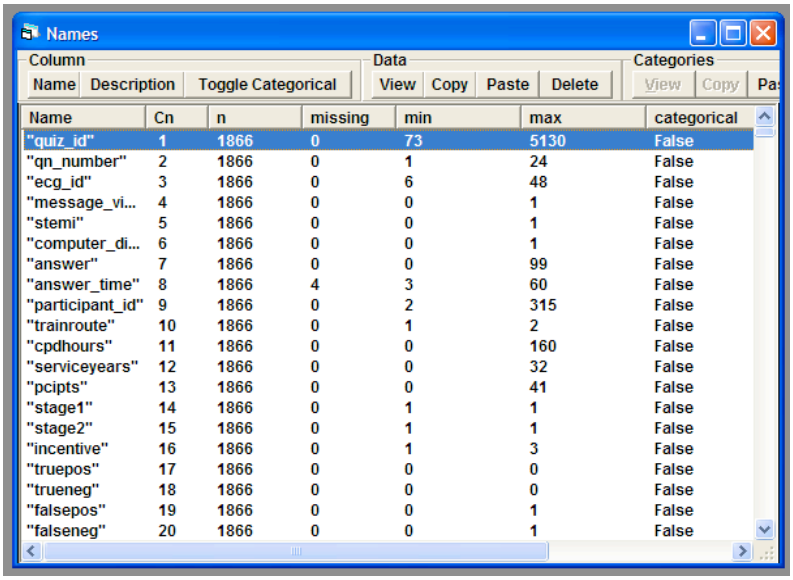


6. The Paste View Window will open. Check that the columns have been filled correctly. Make sure the 'Use first row as names' checkbox is ticked and the click the Past button



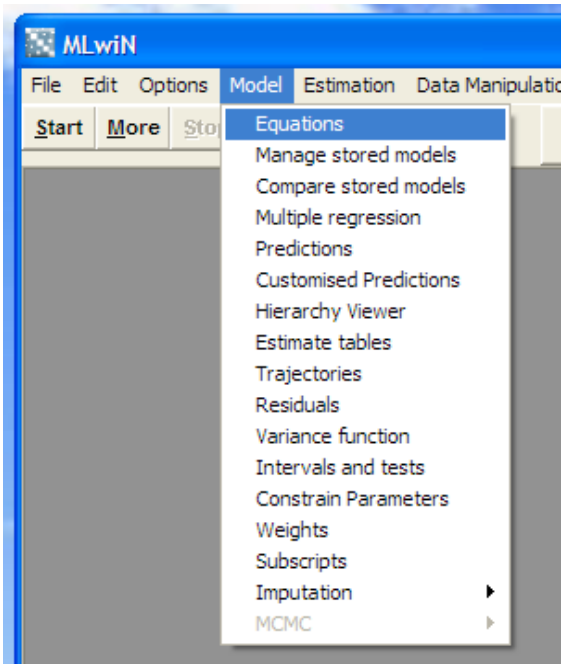
Column	c1	c2	c3	c4	c5	c6	c7	c8
	"quiz_id"	"qn_number"	"ecg_id"	"message_vi	"stemi"	"computer_d	"answer"	"answ
	3102	3	11	0	0	1	0	
	3114	18	11	1	0	1	0	
	3100	5	13	1	0	1	0	
	3112	23	13	0	0	1	0	
	3101	10	18	1	0	1	0	
	3113	13	18	0	0	1	0	
	3098	9	36	0	1	0	1	
	3110	20	36	1	1	0	1	
	3099	4	39	0	1	0	1	

7. The Names window should open showing how the data has been populated into the table. Annoyingly, the double quotes are inserted around the names, which you will have to remember to include in later steps.

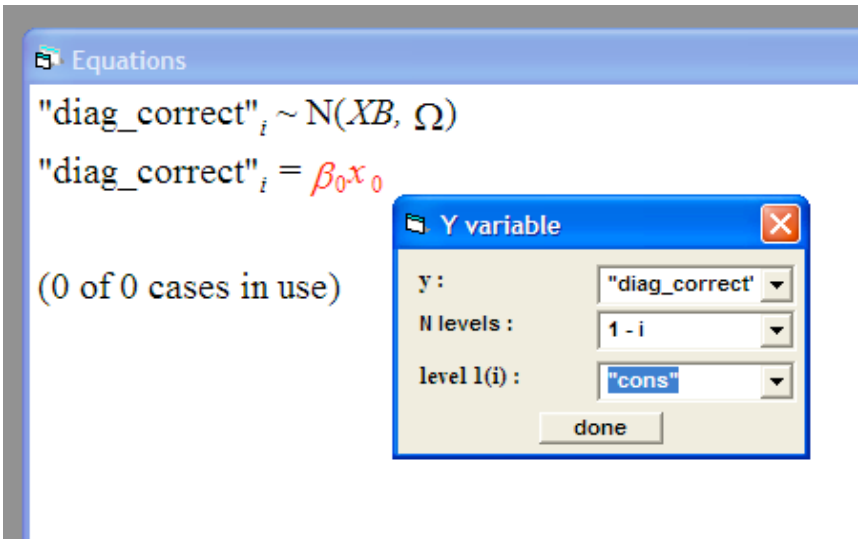


Name	Cn	n	missing	min	max	categorical
"quiz_id"	1	1866	0	73	5130	False
"qn_number"	2	1866	0	1	24	False
"ecg_id"	3	1866	0	6	48	False
"message_vi...	4	1866	0	0	1	False
"stemi"	5	1866	0	0	1	False
"computer_di...	6	1866	0	0	1	False
"answer"	7	1866	0	0	99	False
"answer_time"	8	1866	4	3	60	False
"participant_id"	9	1866	0	2	315	False
"trainroute"	10	1866	0	1	2	False
"cpdhours"	11	1866	0	0	160	False
"serviceyears"	12	1866	0	0	32	False
"pcipts"	13	1866	0	0	41	False
"stage1"	14	1866	0	1	1	False
"stage2"	15	1866	0	1	1	False
"incentive"	16	1866	0	1	3	False
"truepos"	17	1866	0	0	0	False
"trueneg"	18	1866	0	0	0	False
"falsepos"	19	1866	0	0	1	False
"falseneg"	20	1866	0	0	1	False

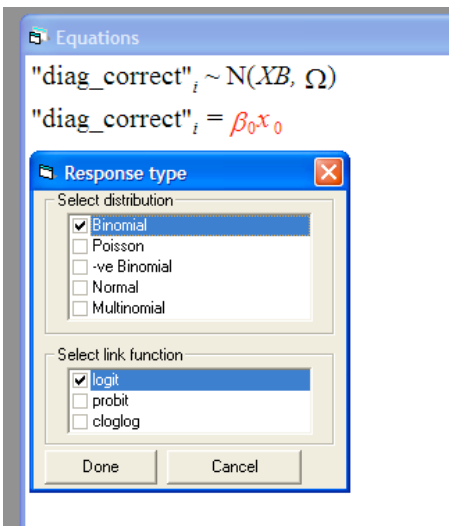
8. Model > Equations This will open the equations window



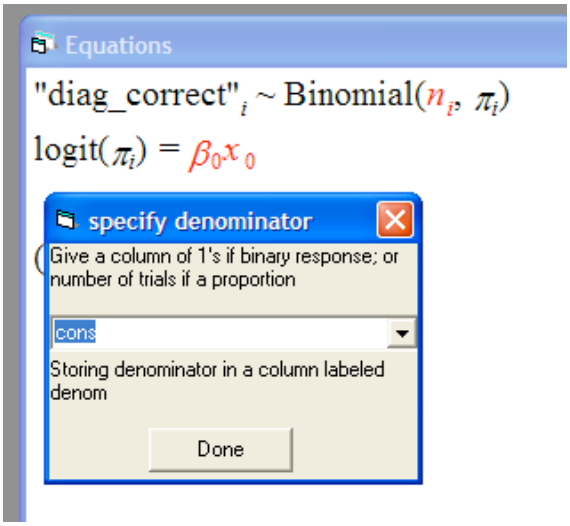
9. Click on the red 'y' in the Equations window and fill out the details as in the image below:



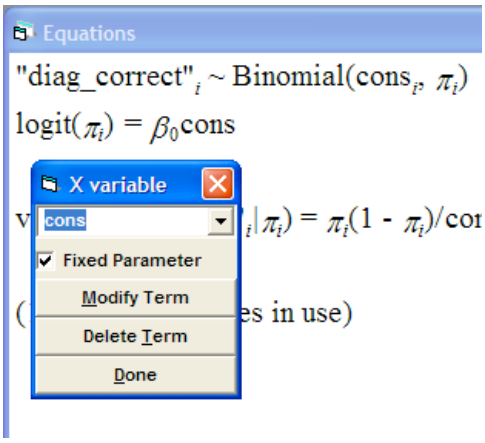
10. Click on N(XB, Omega) and choose the Binomial distribution and logit link function



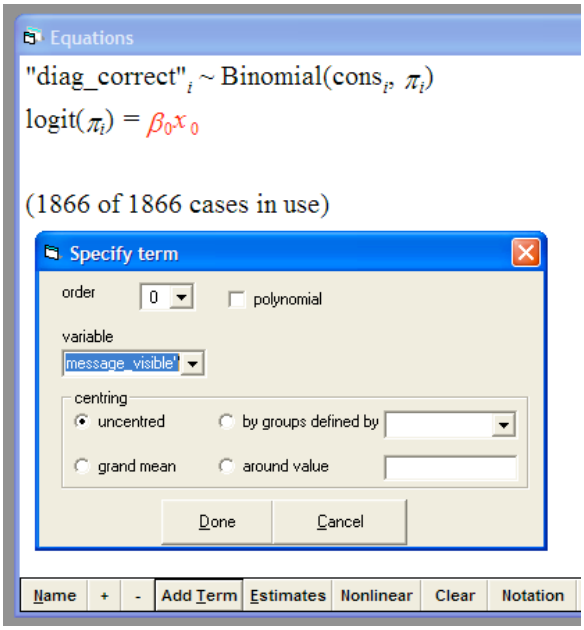
11. Click on the 'ni' just after the word Binomial and select either "cons" or cons as the denominator. They both contain the value of 1 for all rows.



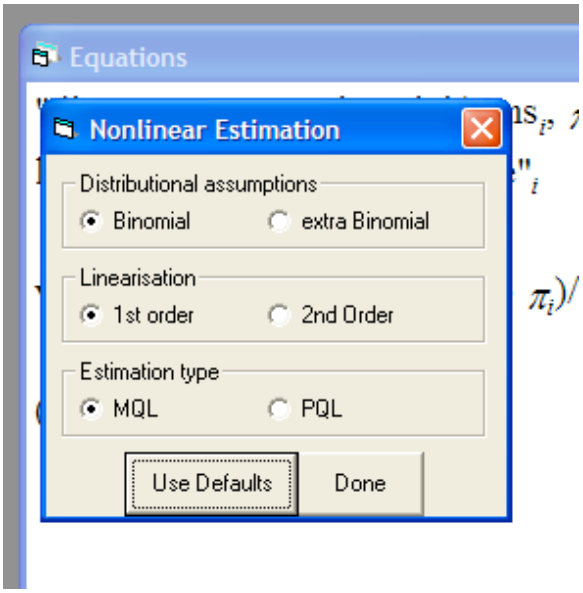
12. Click on the red beta0x0 text. Select 'cons' from the drop-down menu and make sure the 'Fixed Parameter' box is ticked.



13. Click on the 'Add Term' button at the bottom of the Equations window. Select "message_visible" from the variable drop-down menu and click 'Done'



14. Click on the Nonlinear button and click on the 'Use Defaults' button and then 'Done'



15. Clicking on the 'Estimates' button at the bottom of the Equations menu toggles through a range of views, including the notation and the beta value and its associated standard error.

16. In the main MLWin window, click on the 'Start' button. Click on the 'Estimates' button at the bottom of 'Equations' window until you can see the numerical values. MLWin uses iterative generalised least squares (IGLS) to calculate the values, whereas R used LaPlace approximation, but the results should be similar. Check the results from MLWin with the same R results for GLM no Random Effects.

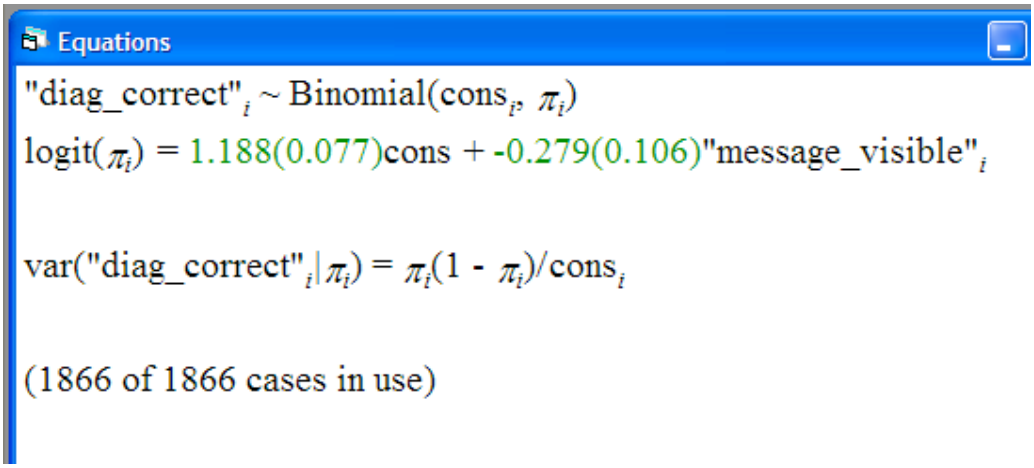
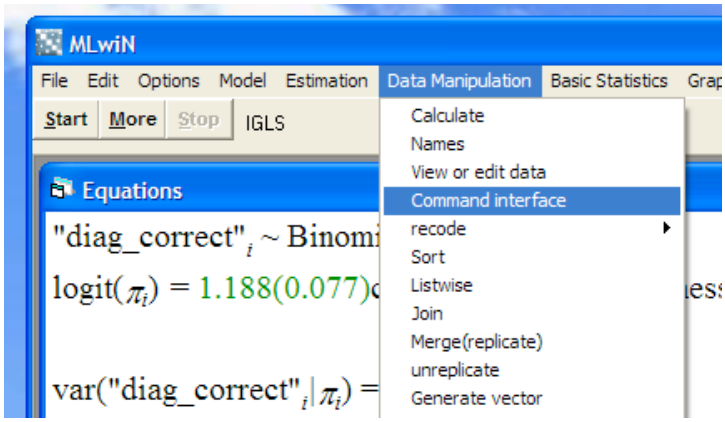


TABLE 4.11: Log odds ratio of correct answer and incorrect computer messages

Parameters	Log OR	Standard error	z	P>|z|	95% CI
GLM no Random Effects					
Constant	1.19	0.08	15.35	0.00	1.04 to 1.34
Message	-0.28	0.11	-2.63	0.01	-0.49 to -0.07

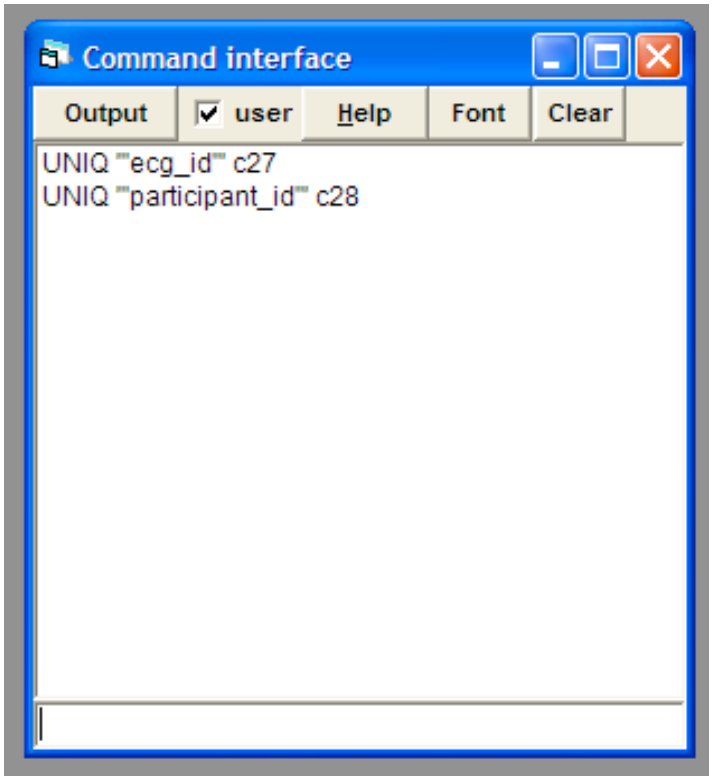
17. In order to conduct the cross-classified random effects analysis, some preparatory work is required in MLWin. Open the command interface, using the menu options: Data Manipulation > Command interface



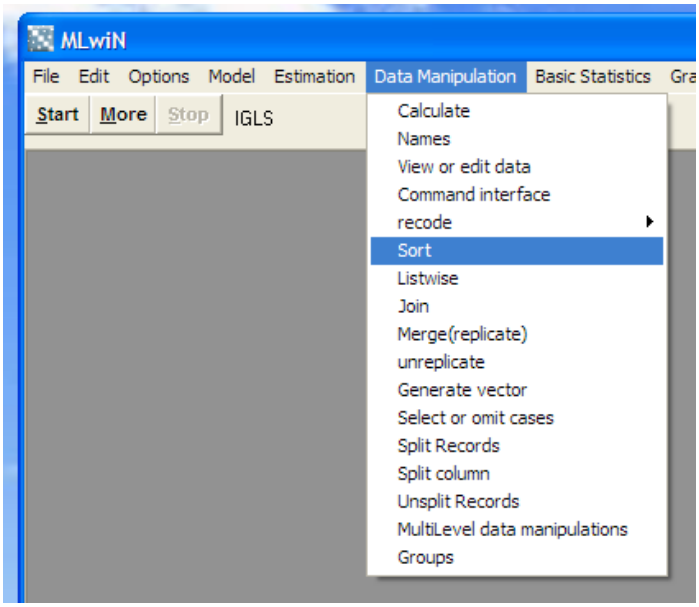
18. Type the following commands in the small data entry window at the bottom of the Command Interface window (Press return after each line):

UNIQ "'ecg_id'" c27
UNIQ "'participant_id'" c28

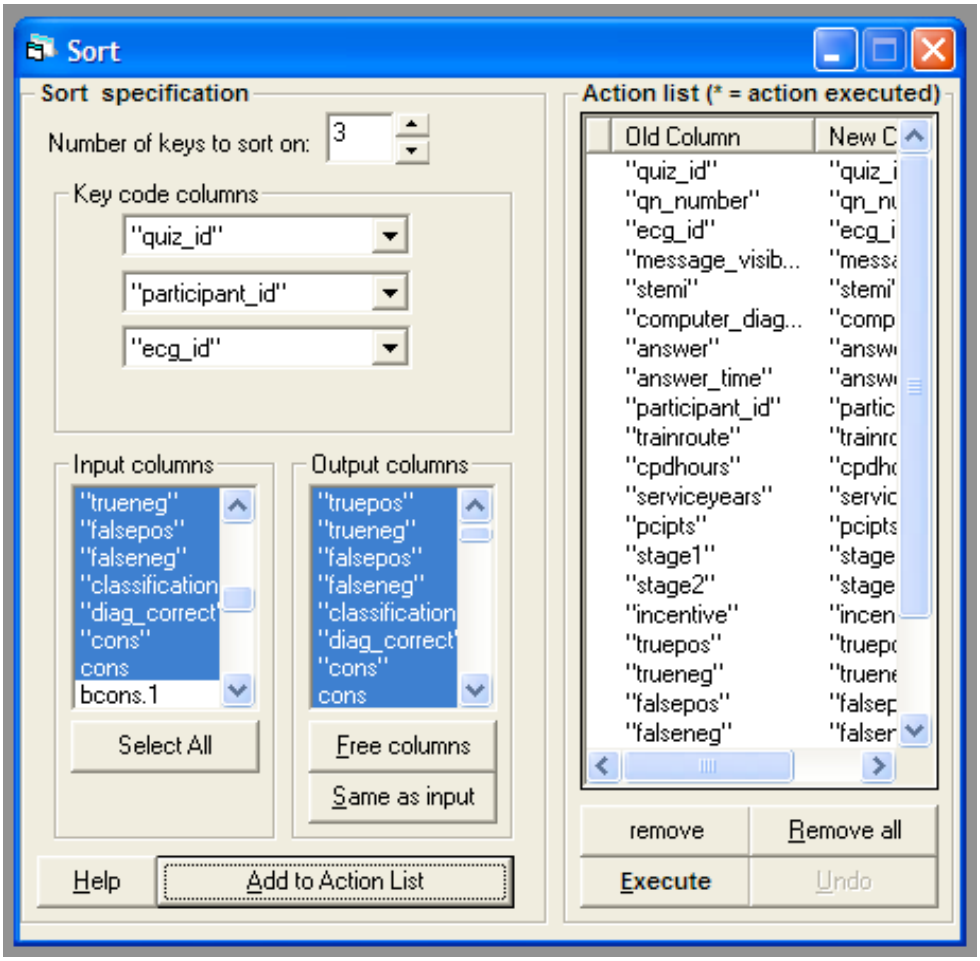
Note that there is a single quote ' and then double quote " required before the variable name and a double quote ", then single ', after.



19. Next, the data needs sorting. Choose the menu option: Data Manipulation > Sort

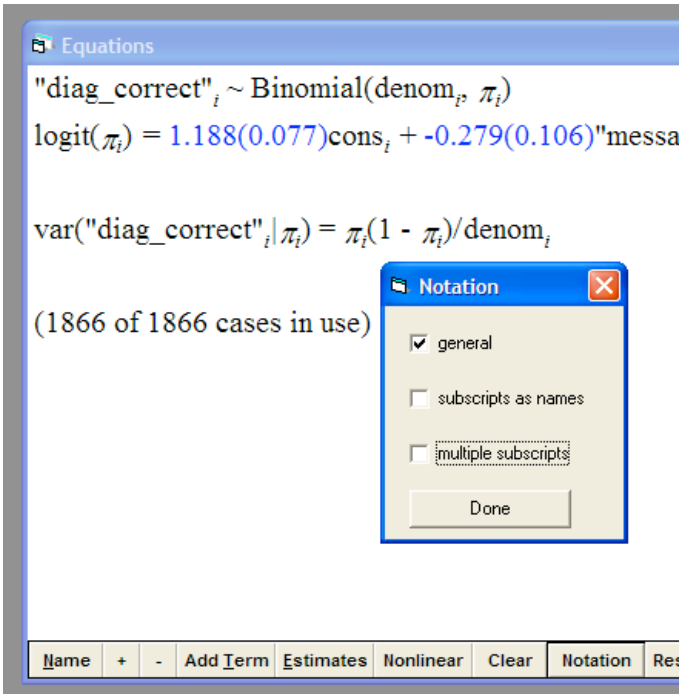


20. Set the number of keys to sort on value to 3 and choose the options from the Key code columns, drop-down menus so that they look like the image below.

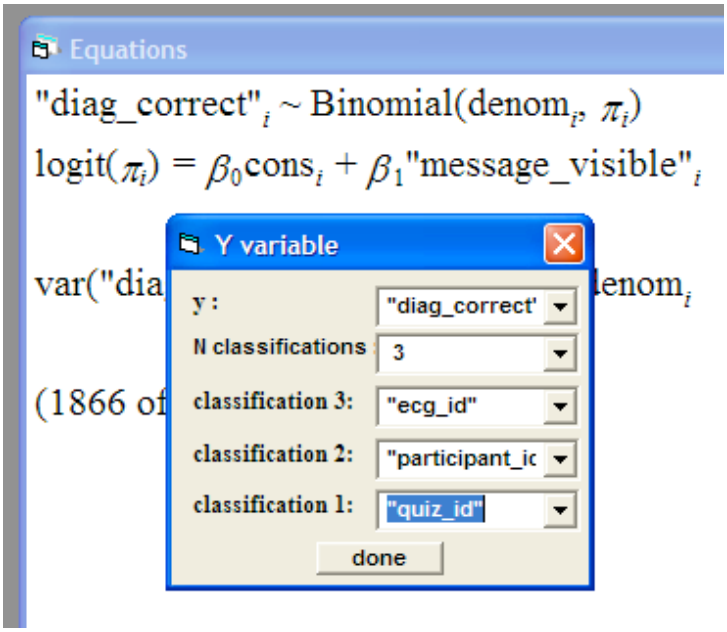


21. In the 'Sort' window, choose the input columns from "quiz_id" to cons. Under the Output columns, click on the 'Same as input' button and then the large 'Add to Action List' button at the bottom of the 'Sort' window. Finally, click on the 'Execute' button on the bottom right-hand side of the 'Sort' window. Small x's should appear next to the Old Column, column if it succeeds.

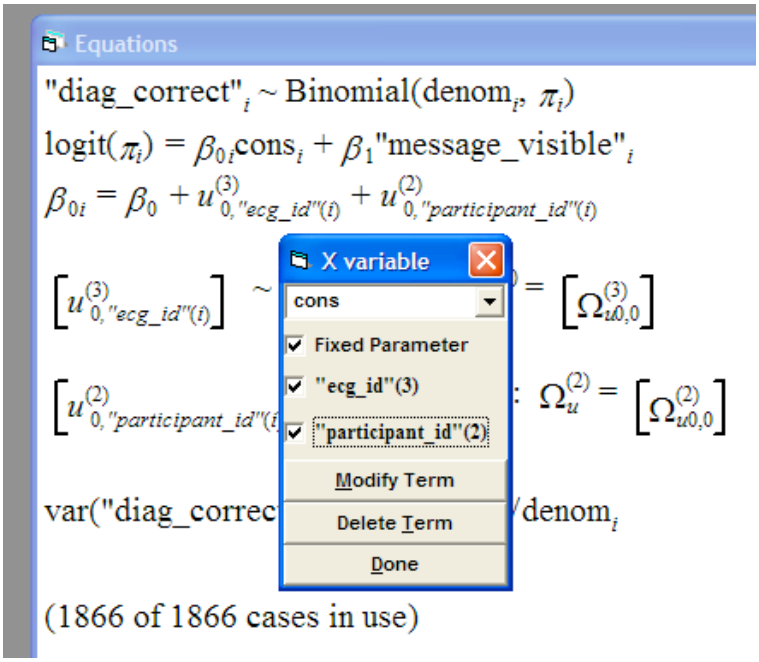
22. Choose the menu option: Model > Equations to bring the Equations window back. Your equation should still be present. Click on the 'Notation' button at the bottom of the Equations window and uncheck the 'multiple subscripts' box.



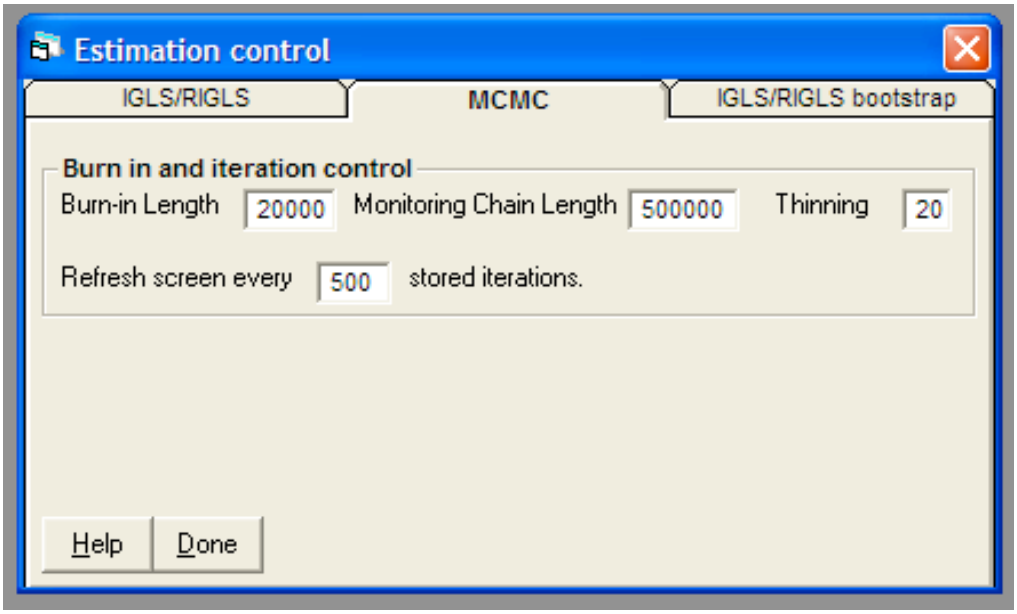
23. In the Equations window, click on the "diag_correct"i text and select options from the 'Y variable' window so that it looks the same as below:



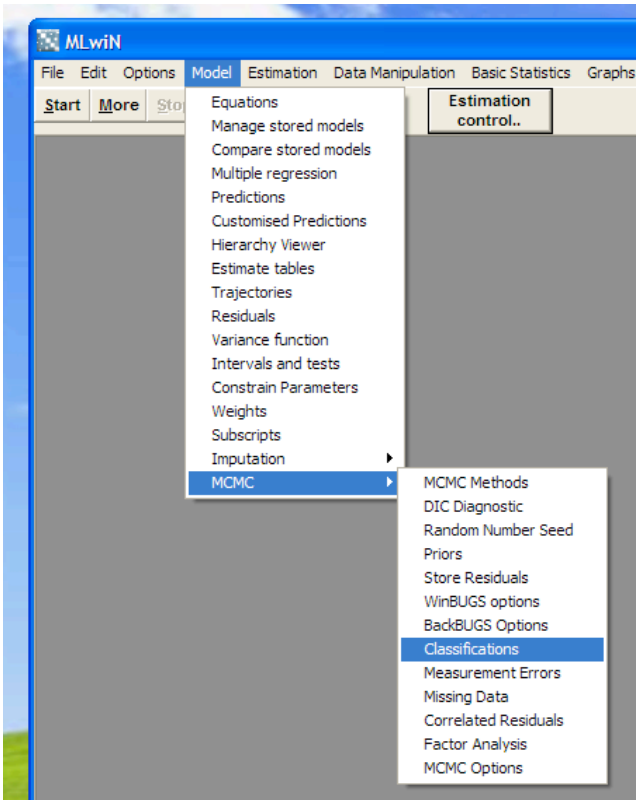
24. In the Equations window, click on the beta0consi text and check all the boxes before clicking on the 'Done' button.



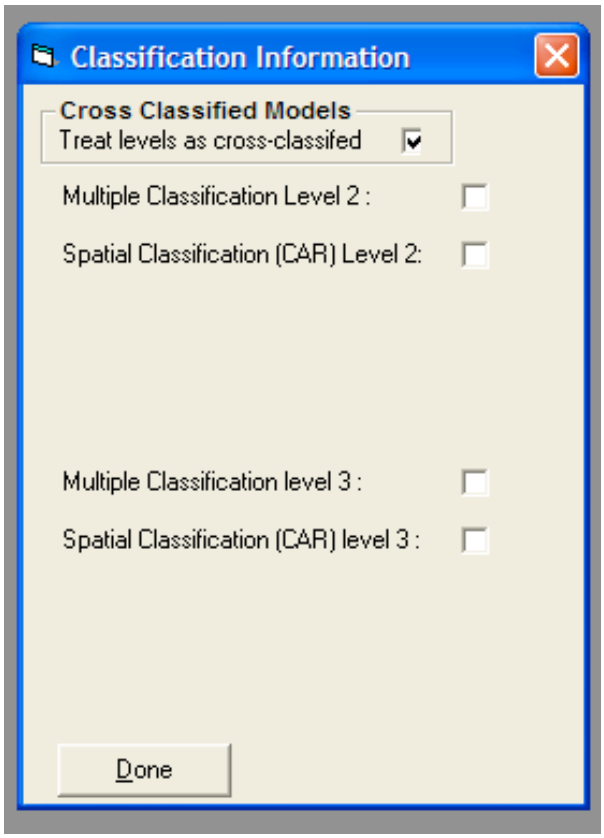
25. In the main MLWin window, click on the 'Start' button to generate some estimates. Once done, click on the 'Estimation control' button and then, in the 'Estimation control' window, click on the MCMC tab. Enter the appropriate numbers into each box, so that it looks the same as the image below and then click 'Done':



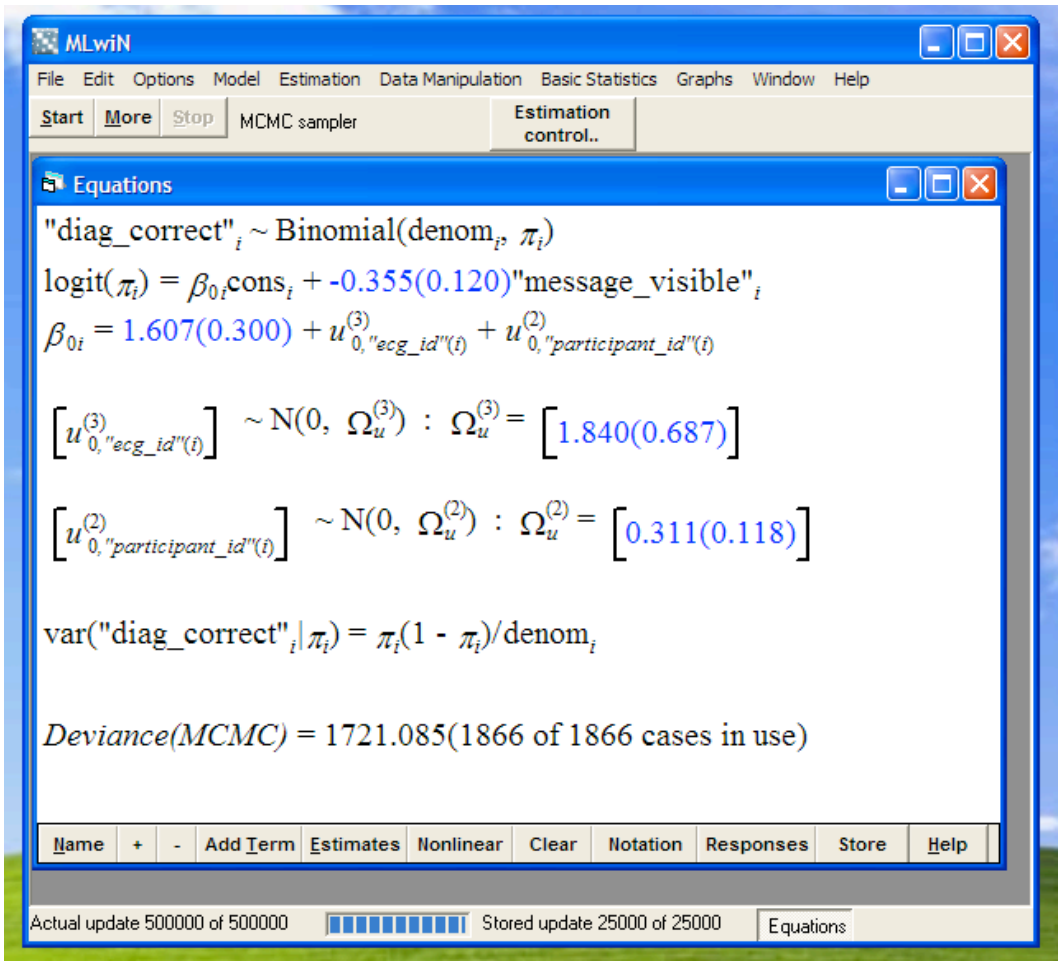
26. From the main MLWin menu, choose the following option: Model > MCMC > Classifications



27. In the 'Classification Information' window, check the box after the 'Treat levels as cross-classified' text.



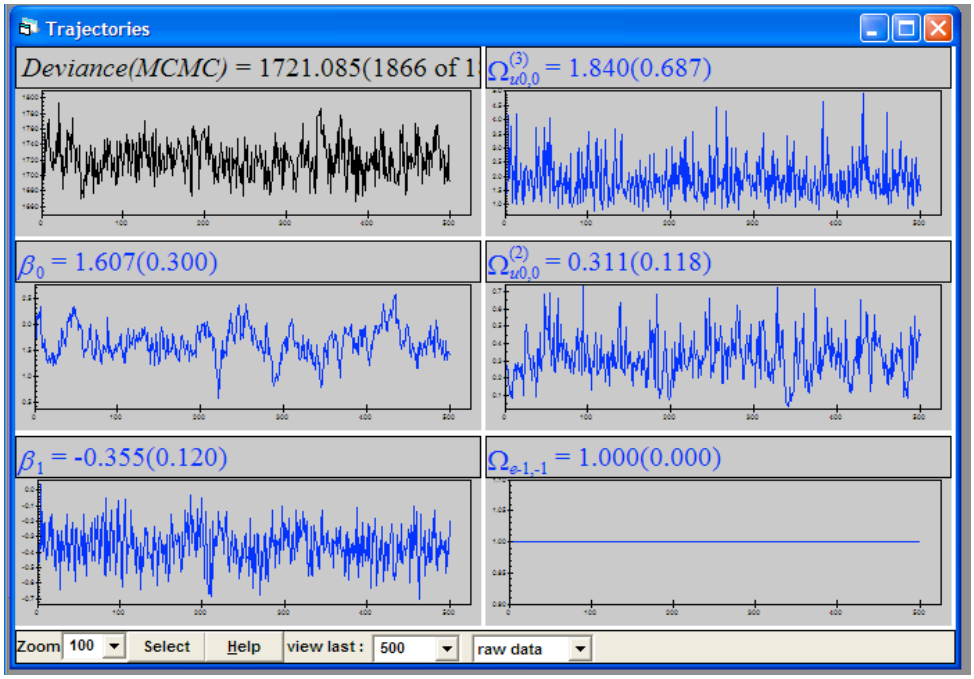
28. In the main MLWin window, click the 'Start' button, then go and have a cup of tea as 500,000 iterations might take 20-30 minutes to complete, depending on the speed of your computer. Unlike with the IGLS estimation, the variables and standard errors do not turn green, but stay blue. The clue that the iterations have completed are provided by the update figures at the bottom of the MLWin window.



These results can then be compared with the R models, by looking at the GLM with Random Effects rows:

GLM with Random Effects		
Constant	1.60	0.28
Message	-0.35	0.12
σ^2_{ecg}	1.57	
$\sigma^2_{participant}$	0.21	

29. To check how well the data has been modeled using the Monte-Carlo Markov chain method, click on the following menu option: Model > Trajectories.



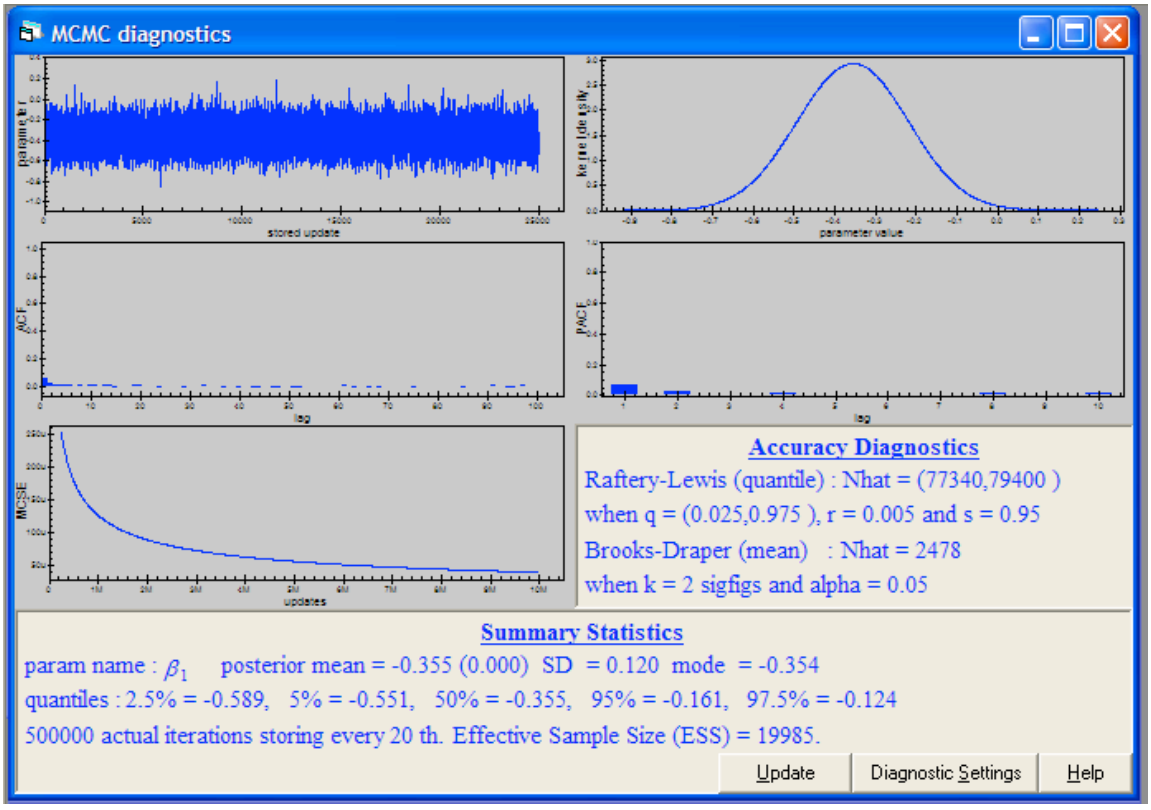
To explore each variable more closely, click on the graph. This will cause a pop-up window to appear asking if you wish to calculate the MCMC diagnostics. Click on the 'Yes' button.

30. Analysis of the convergence diagnostics is rather outside the remit of the dissertation, but the following are useful indicators:

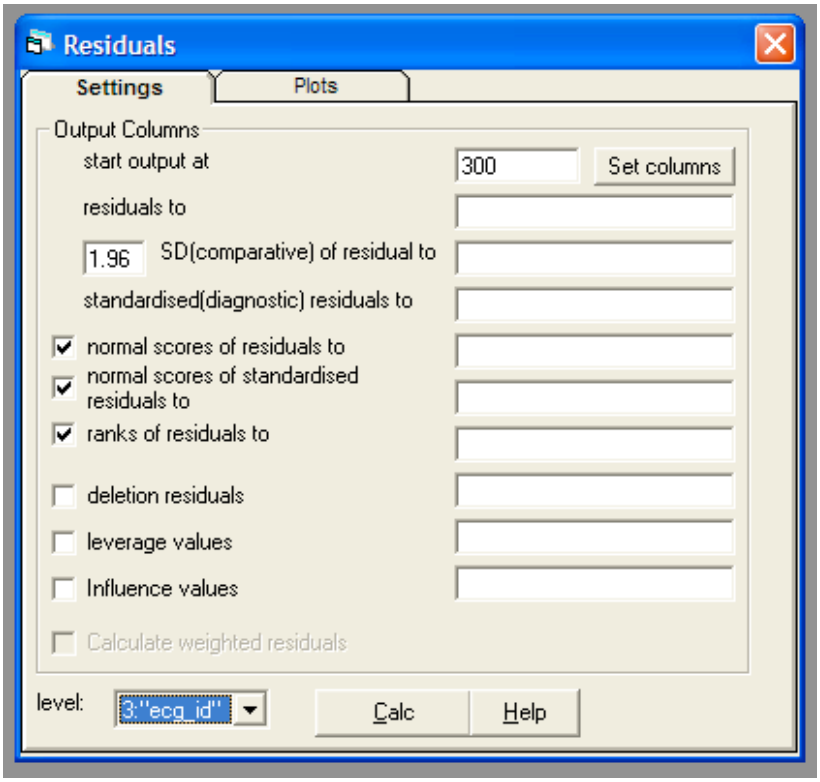
The graph in the top-right hand corner should look like white noise

The top-right hand corner graph should look like a Normal distribution

The number of iterations should be more than both the Raftery-Lewis and Brooks-Draper Nhat values.

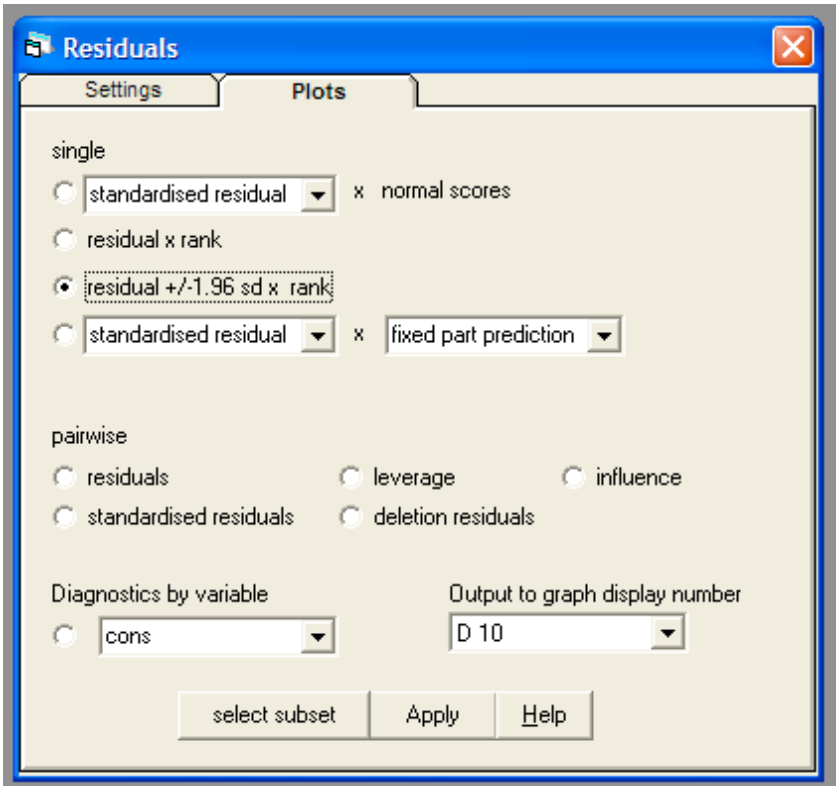


31. Another tool that is quite useful, is the ability to plot the residuals of the model. Select the following menu option: Model > Residuals. To explore the ECGs, change the settings to match the image below:



Once you have changed the settings, click on the 'Calc' button at the bottom of the window.

32. Click on the plots tab and check the 'residual +/- 1.96 s.d. x rank' radio button and then click the Apply button at the bottom of the window.



33. This will produce an interactive graph, plotting the magnitudes of the ECG effects and how they differ from the 'average' ECG. This graph is interactive, making it possible to identify the ECG most associated with the least correct answers (ECG ID 19) and the most (ECG ID 44), in the incorrect computer interpretation subset.