Logistic regression analysis goals and approaches

HL Chapter 4 – part 3

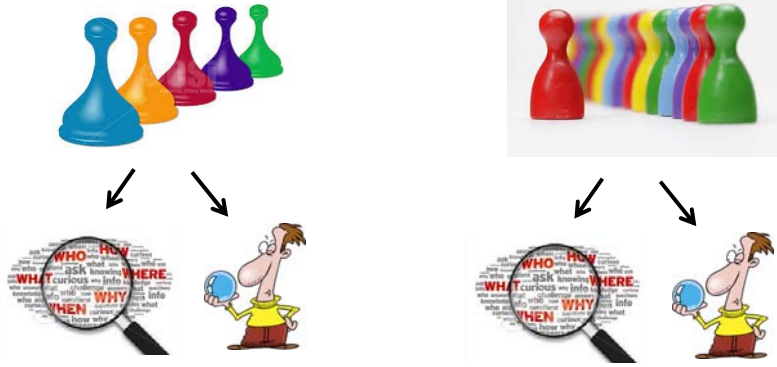
Determine the goal of the analysis

- Are all risk factors of equal interest?

- Is there a risk factor of primary interest?

- Is the model explanatory or predictive?

Potential goals



For all analyses…

Get to know the study variables

- Cross-tabulate categorical variables
- Calculate descriptive statistics for continuous variables (separately for each outcome level)
- Locate unusual or incorrect values

Clean the data

- Delete or correct unusual or incorrect values
- For categorical variables with sparse cells,
  - Collapse categories if possible and meaningful
  - Remove categories otherwise
- As a last resort, remove variables

Select the model covariates

- Model selection is an art as much as it is a science
- Different models may highlight different aspects of the data set






Options

- Purposeful selection
- Automated selection

Important question

- How many variables can be included in a logistic regression model?
- Rough guide:
  No more variables than the “least frequent outcome” divided by 10
- Example:
  n=200, 50 died and 150 lived
  ➔ number of model covariates ≤ 50/10=5 (roughly)

Purposeful model selection…

Potential goal 1: To get the most complete "picture" of the risk factors for the outcome


- Statistically significant variables, confounders and effect modifiers should be included in the model


Univariate analysis

- Check scale of continuous variables
- Determine univariate significance ($p<0.25$) of each study variable

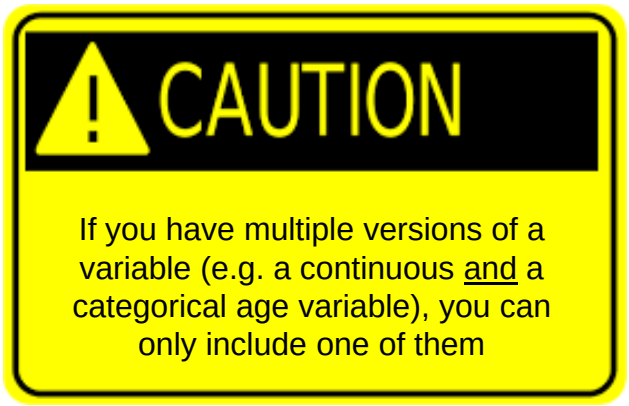
Multivariate analysis

Step 1: Initial main effects model

Include in the model

- Variables that were univariately significant at the 0.25 level
- Biologically important variables
- Known or suspected confounders
- Known or suspected effect modifiers

Multivariate analysis

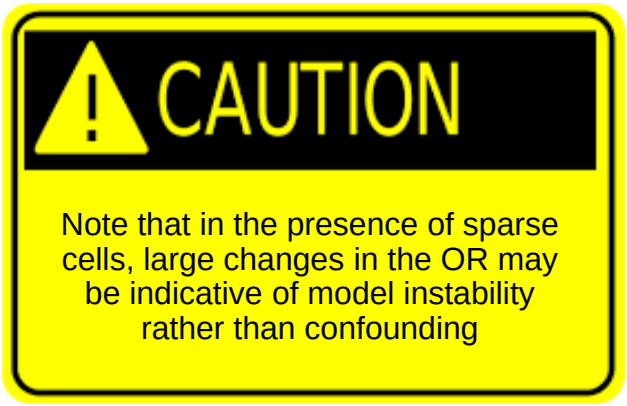


Multivariate analysis

Step 2: Statistical significance and confounding

- Remove the least significant variable and check for confounding
- Return the variable to the model if it appears to be a confounder based on the 10%-change-in-OR rule
- Continue until all model covariates are biologically important, statistically significant ($p<0.05$) or confounders
- You have created your provisional main effects model

Multivariate analysis



Multivariate analysis

Step 3: Additional variables

- Consider variables that have not yet been tested multivariately
- Add these variables to the provisional main effects model
- Determine whether these variables act as confounders or are statistically significant in the multivariate model ($p<0.05$)

Multivariate analysis

Step 4: Scale

- Recheck the scale of all continuous model covariates
- Include correctly scaled variables in the model
- You have created the final main effects model

Multivariate analysis

Step 5: Interactions

- Test significance of interactions between model covariates ($p<0.1$)
- Test significance of interactions between model covariates and "left out" variables ($p<0.1$)

Multivariate analysis

Step 6: Model stability

- Look for unreasonably large ORs or standard errors
- If everything looks OK, you have created your final model

Potential goal 2: To get the most complete "picture" about one specific risk factor

- The risk factor and confounders and effect modifiers of the risk factor should be included in the model

Bi-/tri-variate analysis

- Include the risk factor (RF) in the model
- Check scale of continuous variables
- Determine which variables are confounders or effect modifiers of RF

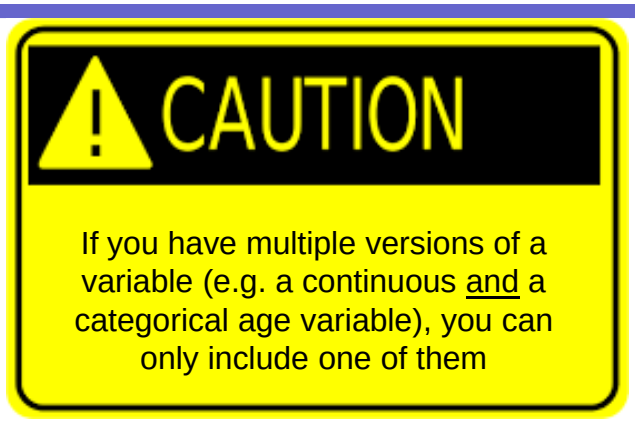
Multivariate analysis

Step 1: Initial model

Include in the model

- RF
- Confounders of RF detected in the bivariate analyses
- Effect modifiers of RF detected in the trivariate analyses (include each effect modifier as a main effect and as part of the interaction term)

Multivariate analysis

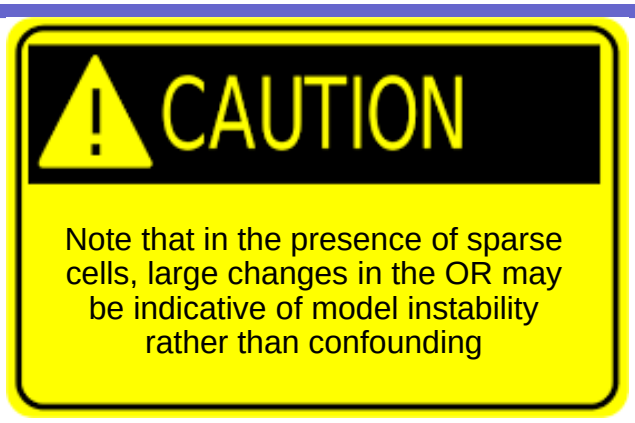


Multivariate analysis

Step 2: Recheck confounders and effect modifiers

- Recheck confounders from bivariate analyses to see if they are still confounders of RF (10%-change-in-OR rule)
- Retest effect modifiers from trivariate analyses to see if they are still effect modifiers of RF ($p<0.1$)
- You have created your provisional model

Multivariate analysis



Multivariate analysis

Step 3: Additional variables

- Consider variables that have not yet been tested multivariately
- Add these variables to the provisional model
- Determine if these variables now act as confounders of RF
- Determine if these variables now act as effect modifiers of RF

Multivariate analysis

Step 4: Scale

- Recheck the scale of all continuous model covariates
- Include correctly scaled variables in the model

Multivariate analysis

Step 5: Model stability

- Look for unreasonably large ORs or standard errors
- If everything looks OK, you have created your final model

Potential goal 3: To best predict the outcome

- Confounders and effect modifiers are only important if they improve the predictive ability of the model
- Note that the same univariate and multivariate analyses can be used as for potential goal 1
- However, you must check the predictive ability of your model and make changes if necessary

Univariate analysis

- Check scale of continuous variables
- Determine univariate significance ($p<0.25$) of each study variable

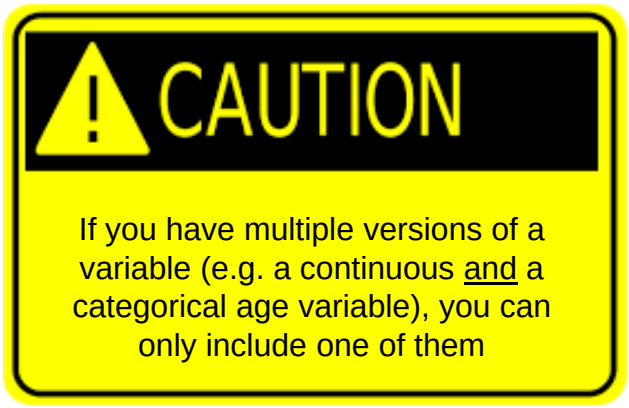
Multivariate analysis

Step 1: Initial main effects model

Include in the model

- Variables that were univariately significant at the 0.25 level
- Biologically important variables
- Known or suspected confounders
- Known or suspected effect modifiers

Multivariate analysis

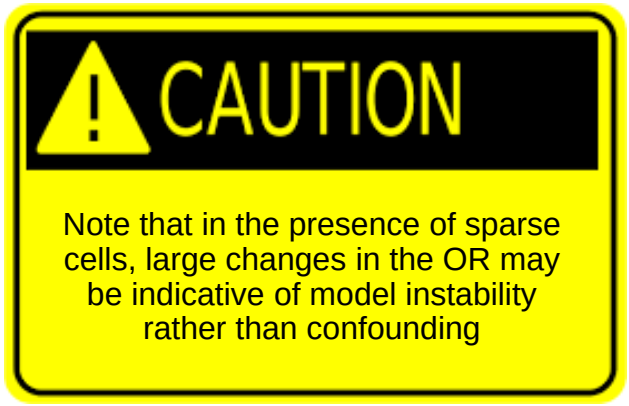


Multivariate analysis

Step 2: Statistical significance and confounding

- Remove the least significant variable and check for confounding
- Return the variable to the model if it appears to be a confounder based on the 10%-change-in-OR rule
- Continue until all model covariates are biologically important, statistically significant ($p<0.05$) or confounders
- You have created your provisional main effects model

Multivariate analysis



Multivariate analysis

Step 3: Additional variables

- Consider variables that have not yet been tested multivariately
- Add these variables to the provisional main effects model
- Determine whether these variables act as confounders or are statistically significant in the multivariate model ($p<0.05$)

Multivariate analysis

Step 4: Scale

- Recheck the scale of all continuous model covariates
- Include correctly scaled variables in the model
- You have created the final main effects model

Multivariate analysis

Step 5:Interactions

- Test significance of interactions between model covariates ($p<0.1$)
- Test significance of interactions between model covariates and “left out” variables ($p<0.1$)

Multivariate analysis

Step 6: Model stability

- Look for unreasonably large ORs or standard errors

Predictive ability of the model

- Evaluate predictive ability of the full model (see chapter 5)
- Remove least significant variable and check if the model still predicts the outcome well
- Return variable to the model if it appears to have an effect on the model's predictive ability
- Continue until all model covariates contribute to the predictive ability of the model

Predictive ability of the model

- Evaluate the effect of interactions on the predictive ability of the model
- Note that models containing different sets of variables may have similar predictive abilities